J-STEM: Journal of Science, technology, engineering, and mathematics (2024), Vol.1 Issue.1 pp. 1-16.
doi: XXX

IMPLEMENTATION OF DATA MINING USING THE K-MEANS CLUSTERING METHOD TO DETERMINE TYPES OF EYE DISEASE AT WULUHAN HEALTH CENTER

Mike Yurike Dita Deviyati[1], Hardian Oktavianto[2], Guruh Wijaya[3], Nur Qodariyah Fitriyah[4] Dudi Irawan[5]

[1,2,3,4,5] Teknik Informatika, Jember, Universitas Muhammadiyah Jember, Kabupaten Jember – Indonesia,

Corresponding author: Mike Yurike Dita Deviyati (ditadevi24@gmail.com)

Article Information: submission received XXX; revision: XXX; accepted XXX; first published online XXX

Abstrak

The eyes are one of the most important human senses and have received world attention. All human activities which are basically based on receiving visual information require special attention. Data regarding visual impairment throughout the world is based on WHO estimates. Eye health is an important aspect in human health. Eye vision problems such as: glaucoma, cataracts, pseudochaphia, conjunctivitis, macular degeneration, diabetic retinopathy etc. are serious problems that will affect an individual's quality of life. Located at the Wuluhan Community Health Center with patient complaint attributes (anamneses) and eye examinations carried out at the Wuluhan Community Health Center eye clinic. From the results of this assessment, types of eye diseases were grouped, namely glaucoma, conjunctivitis and cataracts. The algorithm used is the K-Means Clustering algorithm and there are 6 attributes for each type of disease. The eye disease data used was 110 patient data, the data testing results were obtained using a 2 cluster to 4 cluster scenario and calculating the Devies-Bouldin Index (DBI) value. The K-Means algorithm calculation for data grouping was carried out for each cluster. For the type of glaucoma disease, the best cluster was in cluster 3 with a DBI value of 1.3232. In terms of conjunctivitis, the best cluster is in cluster 4 with a DBI value of 1.3901. For this type of cataract disease, the best cluster is in cluster 3 with a DBI value of 0.51249.

Keywords: data mining, *K-Means*, eye diseases, *cluster*

1. INTRODUCTION

The eye is one of the most important human senses and has received world attention. All human activities which are basically based on receiving visual information require special attention. Eye health is an important aspect in human health with eye vision problems such as: glaucoma, cataracts, pseudochaphia, conjunctivitis, macular degeneration, diabetic retinopathy, etc. (Kurnia et al, 2019). The cause of glaucoma is explained because damage to the optic nerve results in permanent blindness and surgery is not possible. Cataracts occur due to clouding of the lens in the eye which results in blurred vision. Conjunctivitis is caused by inflammation, red eyes or infection of the lining of the eye which results in reddish and watery eyes due to viruses or allergies. The method applied in this research is k-means clustering because it is suitable for the research being tested by grouping eye diseases based on complaints or symptoms felt by the patient. By processing medical record data and eye examinations. At the Community Health Center, we can obtain more important and useful information regarding the grouping of eye disease patients, which can be used to plan strategic steps and make further decisions.

This research uses DBI (Davies Bouldin Index) calculations to determine the optimal value for the number of clusters, DBI itself is the average distance between vectors or features in a cluster. The smaller the number obtained from the DBI calculation, the more optimal it is (Ahmad Fauzi & Danar Dana, 2023). DBI is a validation method for testing optimal clusters in the calculation process. Lower index values indicate better clustering results. The index increases as the separation between clusters increases and the variation within clusters decreases. K-means and DBI applied in this research based on several previous studies, Rian Ordila et. Al. research's in 2000 Application of Data Mining for Grouping Patient Medical Record Data Based on Type of Disease Using the Clustering Algorithm, Yandiko Saputra et. Al. research's in 2022 Clustering Inpatients of BPJS Participants Based on Type of Disease Using the K-Means Algorithm, and research from Abdul Rohman in 2020 Implementation of the K-Means Algorithm for Clustering Student Satisfaction with Academic Services.

2. LITERATURE REVIEW

Research conducted by Rian Ordila, Refni Wahyuni, Yuda Irawan, Maulita Yulia Sari in 2020 entitled "Application of Data Mining for Grouping Patient Medical Record Data Based on Type of Disease Using the Clustering Algorithm (Special Study of PT Inecda Poly Clinic)" through data mining techniques, the K-means clustering algorithm helps in analyzing PT medical record data. To group patients at the Inecda Poly Clinic, the first patient age group is adults (4912 patients), the second group is children (1262 patients), and the third group is infants (144 patient) (Ordila dkk., 2020). Research conducted by Yandiko Saputra Sy in 2022

entitled "Clustering Inpatients of BPJS Participants Based on Types of Disease Using the K-Means Algorithm." hospitalization into three clusters (Saputra Sy, 2022).

Data mining is the process of extracting or extracting previously unknown but understandable and useful data from large databases that are used to make critical business decisions. Data mining is also commonly referred to as "data or knowledge discovery" or finding hidden patterns in data. Data mining, in the sense of knowledge discovery or pattern recognition, is the extraction of hidden knowledge from large amounts of data (Setiawan, 2016).

The three previous studies that were the main reference for this research were that these three studies had not used cluster validation, so in this study DBI was added as a method for assessing cluster validation.

Eye polyclinic, or better known as eye polyclinic, eye polyclinic is an institution that provides comprehensive eye health services to the community in a comfortable and reliable manner, including preventive, curative, promotive and rehabilitative surgical and non-surgical aspects with the aim of reducing the rate of blindness in In Indonesia, however, the meaning of eye polythene can differ between the views of experts or medical personnel and the general view (Sundari dkk., 2022).

Clustering or classification is a method of dividing a dataset into groups based on previously identified similarities. A cluster is a group or collection of data objects that are similar to each other in the same cluster and different from objects in other clusters. Objects are grouped into one or more clusters, and the objects in the clusters are very similar to each other (Benri dkk., 2015).

The K-Means algorithm is a data mining grouping algorithm that allows creating groups that have similarities regarding the same attributes. This K-Means grouping algorithm produces groups of k records, in K-means grouping, for example K is a constant that represents the desired number of clusters (Yunita, 2018).

The calculation steps are as follows:

1. Determine the number of k clusters to be formed
2. Initialization of k cluster centers can be done in various ways. However, this often happens randomly. Cluster centers are assigned initial values using random numbers
3. Map all data/objects to the nearest cluster. The closeness of two objects is determined by the distance between them. Likewise, the proximity of data to a particular cluster is determined by the distance between the data and the cluster center. At this stage, we need to calculate the distance of all the data to each cluster center. The maximum distance between data and a particular

cluster determines which data belongs to which cluster. To calculate the distance of all data to each cluster center, you can use Euclidean distance theory which is formulated as follows:

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i) + (X2j)^2 + \cdots + (Xki - Xkj)^2}$$

D (I,j) = Is the distance of data i to the center of cluster j

X ki = In the i data, the k data attribute

X kj = At the jth center point for the kth attribute

4. Recalculate the cluster membership cluster center. The cluster center is the average of all data/objects in a particular cluster. Can use cluster median.
5. Each object uses a cluster center, the cluster center does not change again then the clustering process is complete. Alternatively, return to step 3 until the cluster center no longer changes. ..

DBI (Davies Bouldin Index) was introduced by David L, Davies and Donald W, Bouldin in 1979. The Davies-Bouldin Index (DBI) itself is a method. Thus, this is a validation method for testing optimal clusters in the calculation process. Lower index values indicate better clustering results. The index increases (decreases) as separation between clusters increases and variation within clusters decreases.

$$DB = \frac{1}{n}\sum_{i}^{n} = 1, i \neq j \max(\frac{\sigma i + \sigma j}{d(ci, cj)})$$

σj: the average distance from the data to the jth cluster data center point

ci: cluster data center point i

cj : cluster data center point j

d(ci , cj) : the distance between the centroids of ci and cj ...

Rapid Miner is a solution for carrying out analysis related to data mining, text mining and predictive analysis. And also related to data mining, text mining and predictive analysis. It can also be used as a data mining machine that can be integrated into a product itself. Therefore, there is no need to use this software anymore as open source software for data mining as is known throughout the world. Rapid Miner offers the ability to simplify the GUI (Graphic User Language) analysis pipeline by providing a graphical user interface that generates an XML (Extensible Markup Language) file that identifies the analysis process (Ahmad Fauzi & Danar Dana, 2023).

3. METHOD

The methodology used in this research goes through several stages. This research method includes Problem Identification, Literature Study, Data Collection, Pre-processing, and Data Processing.

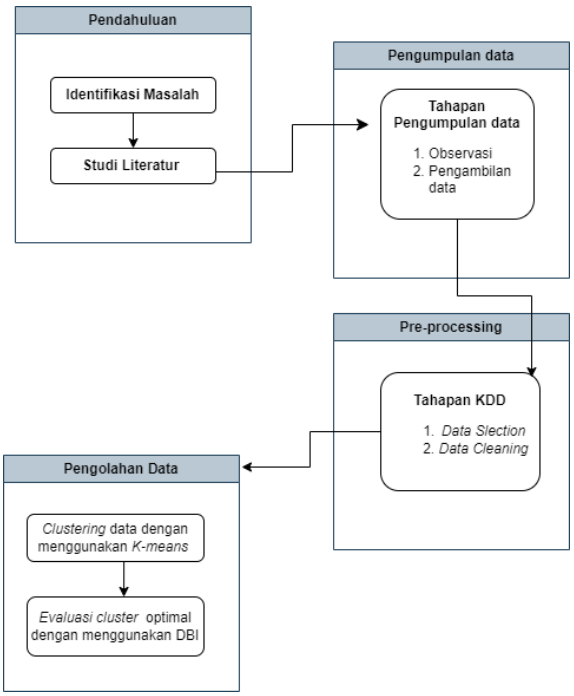

Figure 1. Methodology

Problem Identification

In identifying these problems, the research began by identifying problems within the scope of the Wuluhan Community Health Center. This special study was obtained from the Wuluhan Community Health Center. This research was carried out by identifying what was the problem, especially regarding types of eye disease, starting with the symptoms felt by the patient, with several types of eye disease, namely cataracts, conjunctivitis, glaucoma in the Wuluhan Community Health Center itself.

Literature Study

At the stage of literature study itself, we look for theoretical foundations obtained from sources, journals, books, documentation and also the internet according to the topic that the author is researching and also similar to being able to strengthen arguments and also to search for information.

Data Collection

Data collection is a method that plays an important role in the success and smooth running of research. This research uses a data scale or level of severity of symptoms and characteristics used to analyze the data. Data scale refers to the measurement system or levels used to organize and represent data in research or statistical analysis. Data scale includes characteristics such as the type of measurement, level of accuracy, and variable

structure used in a study. At this stage there are several ways to collect data by making direct observations at the Wuluhan Community Health Center of the patient's condition, eye examination at this stage involves direct examination by medical staff and ophthalmologists, then the patient's medical record at the Wuluhan Community Health Center in this data includes a history of eye disease results. previous examinations and other medical information with observation stages in the Wuluhan Community Health Center environment, especially at the eye clinic itself and also taking patient medical record data to obtain relevant information, regarding the type of eye disease with the patient's symptoms, examination results, there is also treatment that has been given by the party Wuluhan Health Center.

The dataset consist of some attributes, if Pain in the Eyes (NPM), White Dots Appear (MTWP), Watery and Gritty Sensation (SBDB) and Narrowed Visibility (JPM) in the high category, the more serious the case of glaucoma will be. If pain in the eyes (NPM), white dots appear (MTWP), watery and gritty sensation (SBDB) and narrowed visual distance (JPM) are in the moderate category, then glaucoma is not too serious. If pain in the eyes (NPM), white dots appear (MTWP), watery and gritty sensation (SBDB) and narrowed visual distance (JPM) are in the low category, then you only need treatment and regular check-ups.

Pre-processing

This step is the first step in a series of operations before data processing is carried out. These steps are divided into important tasks to be carried out for the purpose of identifying time and work. With the Knowledge Discovery stage for processes that must be carried out on data so that the data mining process can be used, preprocessing in data mining is a process for preparing data, from raw data to into data that can be used, there are stages for processing the data consisting of Data Slection and Data Cleaning. As for how to do data selection and data cleaning

1. Data Slection

Slection data itself is a collection of raw data obtained by observation and having medical record data of eye disease patients. With the data from this selection, it will be used for the data mining process with variables to determine the type of eye disease with the patient's symptoms, the patient's health history will be able to determine the diagnosis.

Table 1. Dataset

NO	NAMA	UMUR	NPAM	MTWP	MM	PB	SB	JPM
1	PNT	49	1,33	1,67	1,33	1,67	1,67	1,67
2	NRN	33	1,67	1,33	1,33	1,33	1,33	1,67
3	SDYH	60	1,33	1,00	1,33	1,33	1,00	1,33

NO	NAMA	UMUR	NPAM	MTWP	MM	PB	SB	JPM
4	AD. MK	62	1,67	1,33	1,33	1,67	1,67	1,33
5	SR	69	1,67	1,67	1,67	1,67	1,33	1,67
6	LK	60	1,67	1,33	1,67	1,33	1,67	1,33
7	ANK W	42	1,67	1,33	1,33	1,33	1,00	1,33
8	IDYL	71	1,67	1,67	1,00	1,67	1,33	1,67
9	HR	43	1,33	1,33	1,33	1,33	1,00	1,33
10	RY HY	39	1,67	1,67	1,00	1,67	1,67	1,67

..

2. Data Cleaning

Data Cleaning This process is carried out to remove duplicate data by checking the data and also correcting errors in the data held.

NO	NAMA	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16	K17	K18
1	PNT	1	2	1	1	1	3	1	2	1	1	1	3	1	1	3	1	1	3
2	NRN	1	1	3	1	2	1	1	2	1	1	2	1	1	2	1	1	1	3
3	SDYH	1	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	2	1
4	AD. MK	1	1	3	1	2	1	1	2	1	1	1	3	1	1	3	1	2	1
5	SR	1	1	3	1	1	3	1	1	3	1	1	3	1	2	1	1	1	3
6	LK	1	1	3	1	2	1	1	1	3	1	2	1	1	1	3	1	2	1
7	ANK W	1	1	3	1	2	1	1	2	1	1	2	1	1	1	1	1	2	1
8	IDYL	1	1	3	1	1	3	1	1	1	1	1	3	1	2	1	1	1	3
9	HR	1	2	1	1	2	1	1	2	1	1	2	1	1	1	1	1	2	1
10	RY HY	1	1	3	1	1	3	1	1	1	1	1	3	1	1	3	1	1	3

.

Data processing

This stage is carried out after preprocessing has been completed and continues, processing data mining in determining the type of eye disease at the Wuluhan Community Health Center using:

1. K-Means Clustering method. Research uses the K-Means method with clustering data obtained through data selection and data cleaning processes using a predetermined number of clusters.

Table 2. Initial Centroid

Centroid Awal	MM	RGM	MBR	KMB	MBKL	PSM
C1	1,67	1,33	1,67	1,33	1,00	1,33
C2	1,67	1,33	1,33	1,33	1,00	1,33

Table 3. Initial Centroid of 2nd Iteration

Centroid Awal	MM	RGM	MBR	KMB	MBKL	PSM
C1	1,43	1,49	1,67	1,46	1,33	1,46
C2	1,53	1,47	1,23	1,57	1,43	1,43

Table 4. 2nd Iteration Calculation for Cluster 2

C1	C2	C1	C2
0,265135	0,19289		OK
0,394881	0,301354		OK
0,096704	0,28209	OK	
0,218058	0,44209	OK	
0,265135	0,09929		OK
0,376573	0,275354		OK
0,564943	0,394578		OK
0,209689	0,42209	OK	
0,590066	0,124131		OK
0,71142	0,284131		OK
0,260973	0,456954	OK	
0,351966	0,19289		OK
0,225927	0,484554	OK	
0,105073	0,39569	OK	
0,564958	0,302143		OK
0,425296	0,408166		OK
0,209689	0,42209	OK	
0,625112	0,190131		OK
0,467127	0,508978	OK	
0,070027	0,32969	OK	
0,538281	0,454566		OK
0,189835	0,525754	OK	
0,82712	0,415784		OK
0,433666	0,323343		OK
0,729727	0,216531		OK
0,216512	0,478154	OK	
0,341527	0,302954		OK
0,28765	0,502954	OK	
0,191381	0,39609	OK	
0,564958	0,302143		OK
		13	17

Table 5. Conjunctivitis Clustering Results

NO	NAMA	C1	C2
N1	RSTI		OK
N2	GRI		OK
N3	LM	OK	
N4	SNO	OK	
N5	AB RN		OK
N6	HH		OK
N7	FDI F		OK
N8	HTL MA	OK	
N9	SRA		OK
N10	SMN		OK
N11	WIJI	OK	
N12	SLO PI		OK
N13	YDS	OK	
N14	FLAH	OK	
N15	LN		OK
N16	WNT		OK
N17	LND	OK	
N18	MA		OK
N19	EK KTO	OK	
N20	SR TI	OK	
N21	SRL		OK
N22	IM MD	OK	
N23	SO		OK
N24	AH		OK
N25	RY		OK
N26	MJITI	OK	
N27	IM MD		OK
N28	AI	OK	
N29	STK HI	OK	
N30	SO		OK
		13	17

...

2. Evaluate optimal clusters using DBI (Davies-Boulding Index) calculations for the optimal number of clusters.

Table 6. Cluster Centroid for 2 Cluster DBI Calculations

Centoid Akhir	NPAM	WTWH	MT	PB	SB	JPM
C1	1,60	1,62	1,33	1,65	1,53	1,62

C2	1,42	1,24	1,42	1,40	1,33	1,44

Calculate the distance value from the i-th attribute data to the ith centroid value for cluster 1. In table 7 the data and centroid values for cluster 1 are as follows:

Table 7. Cluster 1 Data Calculation

N	NPAM	MTWH	MT	PB	SB	JPM	Jarak rata-rata
N1	1,33	1,67	1,33	1,67	1,67	1,67	0,311622
N4	1,67	1,33	1,33	1,67	1,67	1,33	0,444179
N5	1,67	1,67	1,67	1,67	1,33	1,67	0,406335
N8	1,67	1,67	1,00	1,67	1,33	1,67	0,401358
N10	1,67	1,67	1,00	1,67	1,67	1,67	0,371440
N12	1,67	1,67	1,33	1,67	1,67	1,33	0,335442
N14	1,67	1,67	1,00	1,67	1,33	1,67	0,401358
N15	1,67	1,67	1,33	1,67	1,33	1,67	0,225539
N17	1,33	1,67	1,33	1,67	1,67	1,67	0,311622
N18	1,67	1,67	1,33	1,67	1,33	1,67	0,225539
N22	1,33	1,67	1,33	1,67	1,67	1,67	0,311622
N24	1,67	1,33	1,33	1,67	1,67	1,67	0,335442
N26	1,67	1,67	1,67	1,33	1,33	1,67	0,515068
N27	1,67	1,67	1,33	1,67	1,67	1,67	0,166577
N30	1,67	1,67	1,67	1,67	1,67	1,67	0,376813

Results of calculating attribute and centroid distances:

$$C1\ (xi\ , cj) = \sqrt{\begin{matrix}(1{,}33-1{,}60)^2+(1{,}67-1{,}62)^2+(1{,}33-1{,}33^2+\\(1{,}67-1{,}65)^2+(1{,}67-1{,}53)^2+(1{,}67-1{,}62)^2\end{matrix}}$$

$$= 0{,}3116$$

$$C1\ (xn4, cj) = \sqrt{\begin{matrix}(1{,}67-1{,}60)^2+(1{,}33-1{,}62)^2+(1{,}33-1{,}33)^2+\\(1{,}67-1{,}65)^2+(1{,}67-1{,}53)^2+(1{,}33-1{,}62)^2\end{matrix}}$$

$$= 0{,}4441$$

$$C1\ (xn5, cj) = \sqrt{\begin{matrix}(1{,}67-1{,}60)^2+(1{,}67-1{,}62)^2+(1{,}67-1{,}33)^2+\\(1{,}67-1{,}65)^2+(1{,}33-1{,}53)^2+(1{,}67-1{,}62)^2\end{matrix}}$$

$$= 0{,}4063$$

...calculated to the last data attribute

Calculate the distance value from the data for cluster 2. In table 8 are the data and centroid values for cluster 2 as follows:

Table 8. Cluster 2 Data Calculation

N	NPAM	MTWH	MT	PB	SB	JPM	Jarak rata-rata
N2	1,67	1,33	1,33	1,33	1,33	1,67	0,365547
N3	1,33	1,00	1,33	1,33	1,00	1,33	0,450657
N6	1,67	1,33	1,67	1,33	1,67	1,33	0,513106
N7	1,67	1,33	1,33	1,33	1,00	1,33	0,453613
N9	1,33	1,33	1,33	1,33	1,00	1,33	0,390220
N11	1,33	1,33	1,33	1,33	1,33	1,00	0,475596
N13	1,33	1,33	1,00	1,33	1,33	1,33	0,459599
N16	1,33	1,00	1,33	1,33	1,33	1,67	0,361872
N19	1,33	1,33	1,33	1,67	1,33	1,67	0,386641
N20	1,67	1,00	1,33	1,33	1,33	1,67	0,429470
N21	1,33	1,33	1,67	1,67	1,33	1,33	0,406085
N23	1,33	1,33	1,67	1,33	1,33	1,67	0,365547
N25	1,33	1,33	1,67	1,67	1,67	1,33	0,528342
N28	1,67	1,33	1,67	1,33	1,67	1,33	0,513106
N29	1,00	1,00	1,33	1,33	1,33	1,67	0,547843

Results of calculating attribute and centroid distances:

$$C2\ (xi\ , cj) = \sqrt{\begin{matrix}(1{,}67-1{,}42)^2+(1{,}33-1{,}24)^2+(1{,}33-1{,}42)^2+\\(1{,}33-1{,}40)^2+(1{,}33-1{,}33)^2+(1{,}67-1{,}44)^2\end{matrix}}$$

$$= 0{,}3655$$

$$C2\ (xn3, cj) = \sqrt{\begin{matrix}(1{,}33-1{,}42)^2+(1{,}00-1{,}24)^2+(1{,}33-1{,}42)^2+\\(1{,}33-1{,}40)^2+(1{,}00-1{,}33)^2+(1{,}33-1{,}44)^2\end{matrix}}$$

$$= 0{,}4506$$

$$C2\ (xn6, cj) = \sqrt{\begin{matrix}(1{,}67-1{,}42)^2+(1{,}33-1{,}24)^2+(1{,}67-1{,}42)^2+\\(1{,}33-1{,}40)^2+(1{,}67-1{,}33)^2+(1{,}33-1{,}44)^2\end{matrix}}$$

$$= 0{,}5131$$

.... calculated until the last data attribute

The next stage of calculation is to find SSW from the results of data calculations and centroids using the SSW formula as follows:

SSW Calculation 1:

$$SSW^1\frac{1}{15} = \frac{\begin{matrix}0{,}3116+0{,}4441+0{,}4063+0{,}4013+0{,}3714+0{,}3354+0{,}4013+\\0{,}225+\cdots 0{,}1665+0{,}3768+0{,}3655\end{matrix}}{15}$$

$$= 0{,}3426$$

SSW Calculation 2:

$$SSW^2 \frac{1}{15} = 0{,}3655 + 0{,}4506 + 0{,}5131 + 0{,}4536 + 0{,}3902 + 0{,}4755 + \cdots + 0{,}5283 + 0{,}5131 + 0{,}5478 \over 15 = 0{,}4431$$

After calculating the Sum of Squares Within Cluster (SSW), then calculate the Sum of Squares Between Clusters (SSB), which is calculating the distance between centroids and other centroids as below:

$$SSB\ (c1, c2) = \sqrt{(1{,}60 - 1{,}42)^2 + (1{,}62 - 1{,}24)^2 + (1{,}33 - 1{,}42)^2 + (1{,}65 - 1{,}40)^2 + (1{,}53 - 1{,}33)^2 + (1{,}62 - 1{,}44)^2} = 0{,}5680$$

After calculating SSW and SSB, the next step is to calculate the ratio by adding up the SSW results and dividing by the SSB value as follows:

$$R1{,}2 = \frac{0{,}3426 + O, 4431}{0{,}5680} = 1{,}3833$$

In the next stage of DBI calculation:

Table 9. DBI results

Rasio	C1	C2	R MAX
C1	0	1,383384	1,383384
C2	1,383384	0	1,383384

After carrying out testing to find the most optimum DBI value, there was a DBI result with a value of 1.3883 for testing in cluster 2.

Table 10. Results from DBI 2 Clusters

N Cluster	DBI
2 Cluster	1,3833
3 Cluster	1,3232
4 Cluster	1,4549

For the DBI value in each cluster, the DBI value in cluster 2 was found to be 1.3833, while the DBI value in cluster 3 was 1.3232 and the DBI value in cluster 4 was 1.4549. After comparing the DBI values from cluster 2, cluster 3, cluster 4, there is the most optimum DBI value in cluster 3 with a total value of 1.3232, so there is the most optimum DBI value for glaucoma in cluster 3.

Data Elimination

The patient's medical record data will be selected first to obtain data that suits research needs. There were 200 data obtained after carrying out the data cleaning stage, there were 110 data that were in accordance with the research, apart from that the rest were not included because there were multiple diagnoses or there were no genuine symptoms of cataract, conjunctivitis and glaucoma.

4. RESULT AND DISCUSSION

This chapter explains the results of testing the k-meas clustering algorithm, using the Rapidminer tool, which has been carried out on all data consisting of 3 types of diseases, namely cataracts, glaucoma, conjuntivitis. The total amount of data obtained was 110 records in 2022 to 2023, for data on glaucoma there were 30 patient data, conjuntivitis there were 30 patient data while for cataract there were 50 patient data. The main data used for this research are the results of examinations and symptoms from patient complaints (anamnese) and eye examinations (objective), which were taken from the Wuluhan Community Health Center. This research uses the Devies-Bouldin Index (DBI) calculation, by calculating the Devies-Bouldin Index value for each number of clusters and the smallest DBI value will be selected to be the optimum cluster.

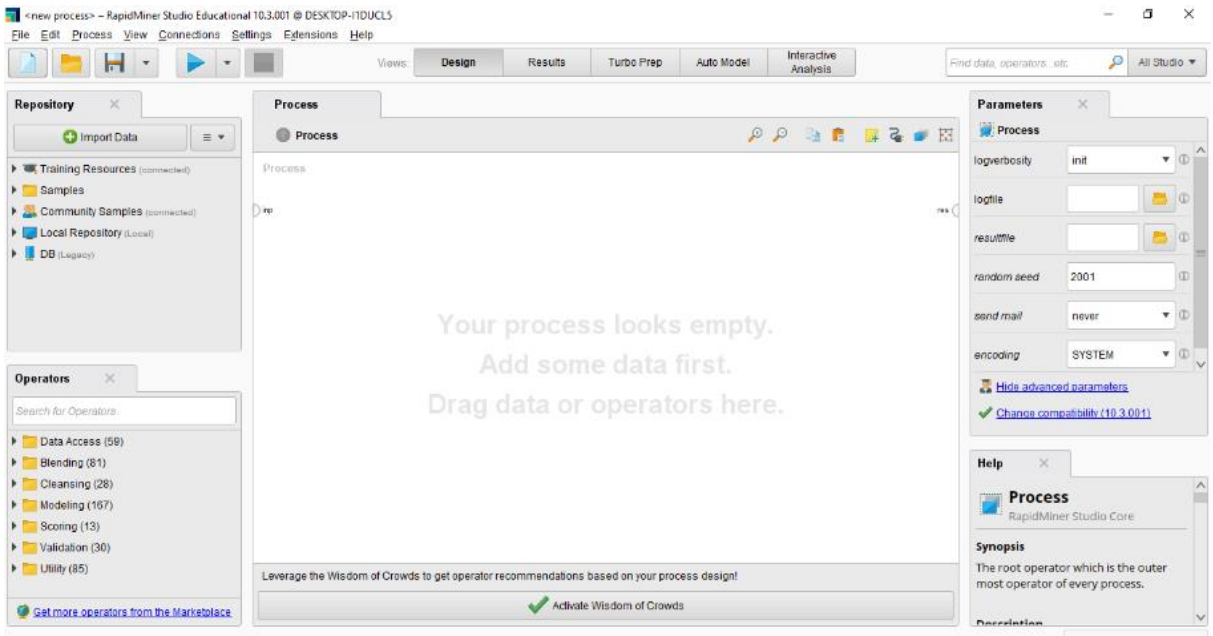

Figure 2. Rapidminer Studio

The total amount of data obtained was 110 records in 2022 to 2023, for data on glaucoma there were 30 patient data, conjuntivitis there were 30 patient data while for cataract there were 50 patient data. The main data used for this research are the results of examinations and symptoms from patient complaints (anamnese) and eye examinations (objective), which

were taken from the Wuluhan Community Health Center. In this research, grouping was also carried out based on the type of disease. This research also grouped each cluster into for glaucoma, conjunctivitis and cataracts there are 2 clusters, 3 clusters, 4 clusters.

The number of clusters with the smallest DBI value is considered the optimal cluster, because the smaller the DBI value obtained, the better the cluster obtained from the K-Means grouping used. The most optimal DBI value is found in cluster 3 with a value of 1.3232. The DBI value in conjunctivitis data is in cluster 4 with a value of 1.390102. Meanwhile, the DBI value in cataract data is in cluster 3 with a value of 0.12495.

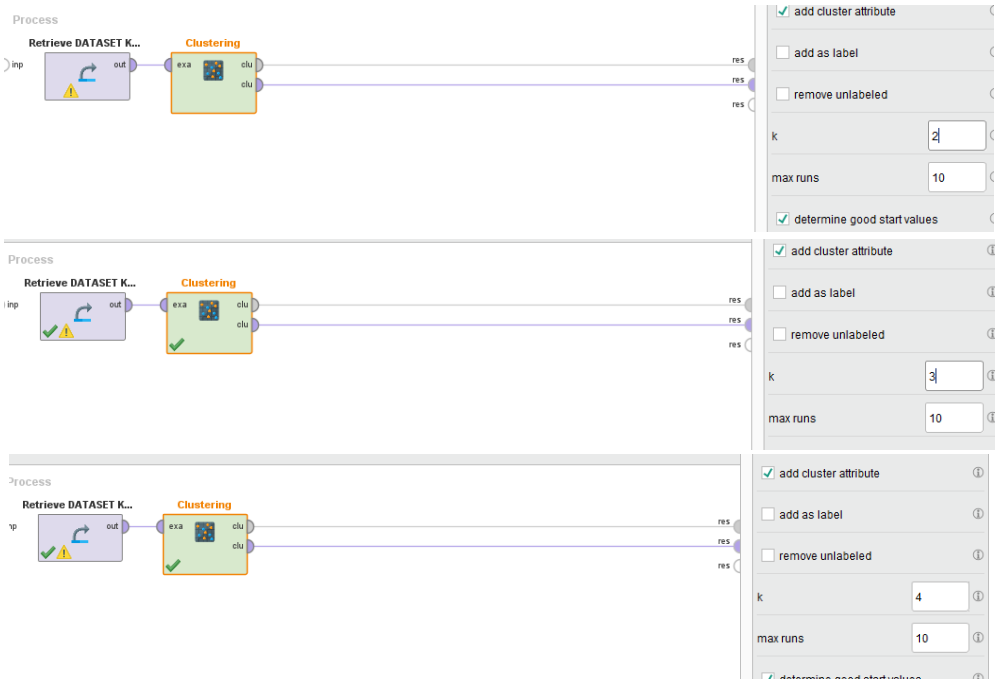

Figure 3. Rapidminer Configuration..

After carrying out calculations and obtaining clustering test results using the K-Means algorithm and optimal clusters using DBI values, it is known that the optimum cluster is in cluster 2 for glaucoma data. Profiling cluster data is carried out to understand the characteristics of each cluster which has been tested in several stages.

Table 11. DBI Result

	Glukoma	Konjungtivitis	Katarak
Cluster	DBI	DBI	DBI
2 *Cluster*	1,38	2,05	1,50
3 *Cluster*	1,32	1,46	0,51
4 *Cluster*	1,45	1,39	1,55

Table 11 shows DBI results from each Glukoma, Konjungtivitis, and Katarak. The yellow colored cells show smallest DBI values that means it is the choosen ones which indicates as optimum clusters, so we get Glukoma optimum in 3 clusters with DBI value 1,32, Konjungtivitis optimum in 4 clusters with DBI value 1,39, and Katarak optimum in clusters with DBI value 0,51..

5. CONCLUSION

The DBI value in each cluster has an optimum value for glaucoma. The optimum cluster is found in cluster 3 with a value of 1.3232. The DBI value in each cluster has an optimum value for conjunctivitis. The optimum cluster is found in cluster 4 with a value of 1.3901. In each cluster there is an optimum value for cataract disease. The optimum cluster is found in 3 clusters with a value of 0.512495.

Future Work

An information system or application can be developed that can help in the process of grouping eye diseases, through this research because this research only applies the K-Means algorithm. Can use different methods to determine the optimal number of clusters.

References

Abdurrauf, M. (2016). Konjungtivitis Bakteri Akut. *Jurnal Kedokteran Syiah Kuala*, 16, 181–184.

Ahmad Fauzi, I., & Danar Dana, R. (2023). Implementasi Data Mining Clustering Dalam Mengelompokan Kasus Perceraian Yang Terjadi Di Provinsi Jawa Barat Menggunakan Algoritma K-Means. *Jurnal Riset Ilmu Manajemen Dan Kewirausahaan*, 1(4).

Anindya Khrisna Wardhani. (2016). Implementasi Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14, 30–37.

Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. https://doi.org/10.46576/djtechno.v1i1.964

Benri, M., Metisen, H., & Latipa, S. (2015). Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118. https://core.ac.uk/download/pdf/287160954.pdf

Kurnia, F., Fahmi, I., Wahyudi, E., & Mige, G. E. S. (2019). Penerapan Algoritma K-Means Untuk Pengelompokan Diagnosa Penyakit Mata Berdasarkan Rentang Usia. *Jurnal Spektro*, 2(1), 10–17.

Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. https://doi.org/10.22202/ei.2016.v2i2.1465

Nairda, D. P., Informatika, F., Telkom, U., Informatika, F., Telkom, U., Tujuan, B., & Hipotesis, C. (2023). *Klasifikasi Keparahan Penyakit Glaukoma pada Citra Fundus Retina dengan Deep k-Nearest Neighbor*. 10(6), 5404–5409.

Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. https://doi.org/10.33060/jik/2020/vol9.iss2.181

Saputra Sy, Y. (2022). Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means. *Jurnal Sistim Informasi Dan Teknologi*, 5, 33–37. https://doi.org/10.37034/jsisfotek.v5i2.162

Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, 3(1), 76–92.

Sundari, S. S., Agustin, Y. H., & Rihadisha, A. (2022). Rancang Bangun Aplikasi Sistem Pakar Diagnosa Penyakit Mata Berbasis Web Dengan Metode Forward Chaining Dan Case Based Reasoning (Studi Kasus : Poli Mata RSIA Widaningsih Tasikmalaya). *E-Jurnal JUSITI (Jurnal Sistem Informasi Dan Teknologi Informasi)*, 11(01), 91–100. https://doi.org/10.36774/jusiti.v11i1.914

Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru. *Sistemasi*, 7(3), 238. https://doi.org/10.32520/stmsi.v7i3.388