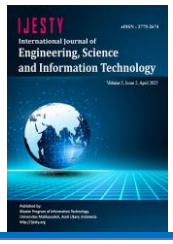
International Journal of Engineering, Science and Information Technology
Volume 5 No. 2 (2025) pp. 342-351
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i2.847
Research Paper, Short Communication, Review, Technical Paper



Data Mining Analysis for Clustering the Number of Tb Patients in North Aceh Health Centers Using the Spectral Method Clustering

Khainesya[1]*, Eva Darnila, Risawandi[3]

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: khainesya.200170229@mhs.unimal.ac.id*

The manuscript was received on 22 June 2024, revised on 25 October 2024, and accepted on 23 February 2025, date of publication 8 April 2025

Abstract

Tuberculosis (TB) is one of the infectious diseases that is a significant concern in the world of health, especially in the North Aceh region. Grouping the number of TB patients based on severity and region is very important to support decision-making in further prevention and treatment efforts. This study applies the Spectral Clustering method to cluster the number of TB patients at Baktiya Health Center, Bayu Health Center, and Lhoksukon Health Center to identify patient distribution patterns based on severity categories. The system built is a web-based data mining analysis system using PHP and MySQL as a database. Clustering is done by dividing patients into three categories, low, medium, and high, based on five main criteria, namely age, gender, month of treatment, diagnosis results, and patient address. The results showed that Lhoksukon Health Center had the highest number of TB patients, with 136 patients (37.06%), an average age of 48.6 years, and the most cases occurred in December 2022. Bayu Health Center was at a moderate level with 130 patients (35.42%), most of whom were 45.5 years old, and most cases occurred in November 2023. Meanwhile, Baktiya Health Center had the lowest number of patients, 101 (27.52%), with the most cases occurring in November. From the clustering results, it can be concluded that the Spectral Clustering method can group TB patients well to help medical personnel and related parties develop more effective intervention strategies based on the region and severity of the patient.

Keywords: *Data Mining, Clustering, Spectral Clustering, Tuberculosis*.

1. Introduction

Increasing air pollution due to industrial growth and motor vehicles has reached alarming levels. This has led to various respiratory diseases in the local population, including Tuberculosis (TB), which is a dangerous lung disease. TB is a worldwide public health problem and has been declared a global emergency by the World Health Organization (WHO) [1].

The clustering process in analyzing TB patients takes a long time because the data has enormous dimensions. One technique that can be used is Spectral Clustering, which has proven effective in identifying TB patients [2] [3]. This method was developed to improve regression accuracy and group data based on similarities between data[4]. Spectral Clustering can better partition data with complex structures than traditional clustering methods such as k-means, fuzzy c-means, and SOM. This technique uses the eigenvalues and eigenvectors of the adjacency matrix to form more accurate clusters.

Medical classification of TB disease data is essential in predicting the disease; it can even help doctors make decisions on the disease's diagnosis. Thus, it is necessary to make an early diagnosis to reduce the transmission of TB to the broader community. TB disease is one of the most serious public health problems to be considered because it is a disease that is categorized as spreadable and contagious. This disease is transmitted through air polluted by Mycobacterium Tuberculosis mixed with air pollution due to pollution [5] [6]. TB continues to be a major scourge in the world of health, especially in developing and developed countries. TB also still cannot be eradicated optimally.

On this occasion the research to be carried out is Clustering the Number of TB Patients at the North Aceh Puskesmas Using the Spectral Clustering Method by taking a study of 3 Puskesmas, including the Baktiya Puskesmas, the Lhoksukon Puskesmas and the Bayu Puskesmas and clustering the number of TB patients to know the highest, medium and lowest number of TB patients in any area.

2. Literature Review

2.1. Tuberculosis

Tuberculosis (TB) is an infectious disease caused by mycobacteria tuberculosis. This is a powerful bacteria that requires a long time to treat. This bacterium infects the lungs (90%) more often than other organs [7]. Tuberculosis (TB) is an infectious disease that most often affects the bar parenchyma, usually caused by mycobacterium tuberculosis. TB can spread to almost every part of the body, including the meninges, kidneys, bones, and lymph nodes. Initial infection usually occurs within 2 to 10 weeks after exposure. Patients may then develop active disease as the immune system response declines or is not robust [8].TB is transmitted when a person with active lung disease expels the organism. Susceptible individuals inhale the droplets and become infected. Bacteria are transmitted to the alveoli and multiply. The inflammatory reaction produces exudates in the alveoli and bronchopneumonia, granulomas, and fidrosa tissue [9]. Tuberculosis (TB) is a chronic infectious disease that is still a serious problem found in the world population, including Indonesia. This lung disease caused by mycobacterium tuberculosis is found to have infected nearly one-third of the world's population, which has become a major health problem globally, according to the World Health Organization [10] [11] [12].

2.2. Causes and Risks of Tuberculosis Incidence

TB is caused by the rod-shaped Mycobacterium Tuberculosis germ measuring ± 0.3-0.6 and ± 1-4 µ long. Has unique properties resistant to acid in coloring. TB germs die quickly in direct sunlight but can survive for several hours in a dark and humid place. These germs can lie dormant in body tissues for several years [13]. There are several types of Mycobacterium, such as Mycobacterium africanus, Mycobacterium bovis, Mycobacterium kansasii, Mycobacterium avium and Mycobacterium nenopi. However, the important one is Mycobacterium tuberculosis, which causes tuberculosis and mainly attacks the lungs [14]. According to the WHO survey (2003), 90% of TB sufferers in the world attack weak or poor socioeconomic groups and according to Enarson, TB is the most common disease that attacks countries with low-income populations.

2.3. Data Mining

Data mining is a series of processes that extract added value in the form of information that has not been known manually from a database. Data mining began to exist in the 1990s as a correct and appropriate way to retrieve patterns and information used to find relationships between data to group into one or more clusters so that objects that are in one cluster will have a high similarity between one another, data mining is part of the process of discovering knowledge from databases Knowledge Discovery in Databases [15].

2.4. Clustering

Clustering is one of the data exploration methods used to find patterns in a dataset. Generally, the pattern can be seen from the similarity of the records' properties, characteristics, or characteristics in the dataset [16]. The clustering process will group data items into small groups so that each group has an essential similarity, making it easier to search for data based on existing similarities. In other words, a cluster has a set of objects grouped because of their similarities or closeness. Clustering is one of the unsupervised methods [17].

Clustering does not contain target variables, which distinguishes it from classification. Experts have developed many methods of Clustering. Each method has its characteristics, advantages, and disadvantages. According to their structure, clustering methods are divided into two parts, namely hierarchical Clustering and partitioning. Hierarchical Clustering has a rule where one single piece of data can be considered a group, two or more small groups can be merged into one large group, and so on until all data can be incorporated into one group. The partitioning clustering method groups data into several groups that do not overlap between one group and another, meaning that each data is only a member of one group [18].

2.5. Spectral Clustering

Spectral Clustering is an algorithm for finding clusters using the eigenvectors of a matrix. This matrix is derived from the similarity matrix between each data group in pairs[19]. Several methods based on the different Laplacian graph formulas are used in the Spectral Clustering algorithm.

1. Forming Similarity Matrix

   The similarity matrix helps show whether or not there is a relationship between one data set and another. If there is a relationship, there is a value between one data and another, while if there is no relationship, it will be zero. The diagonal value in the similarity matrix will be zero because there is no relationship to the data. Similarity measurement can be done by calculating the Euclidean Distance, a method for finding the distance between two data points. The smaller the distance, the closer and more similar the data will be[20]. The calculation of the similarity value is calculated by Euclidean Distance using the formula on the next page:

   ……..……..…….(1)

   $$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_{training}^{i} - y_{testing}^{i})^2}$$

   Description:
   d(x, y) = Distance
   x_training^i = Training Data
   y_testing^i = Testing Data
   i = Variable Data
   n = Data Dimension

2. Forming Degree Matrix
Degree Matrix is a diagonal matrix that describes the degree of vertices (the number of edges connected to each vertex) in a graph. Similarity matrix and degree matrix are used to form a laplacian matrix. The formula for calculating the diagonal value is:

…… (2)

$$d_{ii} = \sum^{n} W$$

Description:
D = diagonal matrix
n = number of data
i and j = data row and column numbers
W = similarity matrix

3. Forming Un-Normalized Laplacian Matrix and Normalized Laplacian Matrix
Laplacian Matrix is a matrix that represents a graph, and it is helpful to find many valuable insights from the graph. Laplacian Matrix is divided into several methods, including Un- Normalized Laplacian and Normalized Laplacian. The difference between the two methods lies in the formula applied; if Un-Normalized Laplacian uses an unnormalized formula, then Normalized Laplacian uses a normalized formula. Normalized Laplacian is also known as Symmetric Laplacian.

……(3)

$$L = D_i$$

……..(4)

$$L_{svm} = 1 - D^{-1/2} A D^{-1/2}$$

Description:
L is the value of Un-Normalized Laplacian
$Lsym$ is the value Normalized Laplacian

3. Research Method

3.1 Type of Research Data

This research uses quantitative research methods, which are divided into three sub-categories: inferential, experimental, and simulation. The research design used in this research is experimental. The experimental research method can be interpreted as seeking the effect of specific treatments on others under controlled conditions [21].
The types of experimental research are divided into two, namely absolute experiments and comparative experiments.

3.2. Population and Sample Selection Method

Population is a generalization area consisting of objects/subjects with specific qualities and characteristics set by researchers to study and then draw conclusions[22]. The population in this study consisted of TB patients, who came from internal data from the Lhoksukon Health Center, Baktiya Health Center, and Bayu Health Center, North Aceh Regency. The sample is part of the number and characteristics of the population. Research can use samples taken from the population. The conclusion will apply to the population based on what is learned from the sample. The sample taken from the population must be genuinely representative (representative). The sample of this study is medical recap data for patients with positive tuberculosis disease.

4. Result and Discussion

4.1. Data Analysis

The dataset used in this study is patient data from health centres in the North Aceh district. Data was obtained from Baktiya Health Center, Lhoksukon Health Center and Bayu Health Center. The data is in the form of subsections of patient data at the puskesmas. The following is the patient data of Puskemas Baktiya, which was recapitulated between 2020 and 2024. Here is one example of the dataset used in this study.

Table 1. Patient Data of Baktiya Health Center in 2020

No	Age	Gender	Month of Treatment	Diagnosis Result	Diagnosis Result
1	77	L	February	Pulmonary TB	Alue Bili Rayeuk
2	70	L	January	Pulmonary TB	Matang Raya Timu
3	35	P	January	Pulmonary TB	Cinta Makmur
4	32	P	January	Pulmonary TB	Arongan Lise
5	25	P	June	Pulmonary TB	Alue Buya
6	60	P	June	Pulmonary TB	Geulumpang Samlakoe
7	41	L	January	Pulmonary TB	Cinta Makmur

The data above is patient data obtained from the Baktiya Health Center based on the 2020 period from January to December. More complete data can be seen in the attachment list. Then below is the patient data of the Baktiya Health Center in 2021.

Table 2. Patient Data of Bayu Health Center in 2024

No	Age	Gender	Month of Treatment	Diagnosis Result	Diagnosis Result
1	59	P	October	Pulmonary TB	Rheng
2	27	L	September	Pulmonary TB	Rheng
3	70	L	September	Pulmonary TB	Blang Awe
4	19	P	August	Pulmonary TB	Blang Awe
5	70	L	August	Pulmonary TB	Pulo Blang Trieng
…	…	…	…	Pulmonary TB	…
13	43	P	February	Pulmonary TB	Alue Manjron
14	25	P	February	Pulmonary TB	Baroh Blang Rimueng
15	53	L	January	Pulmonary TB	Blang Awe
16	49	P	January	Pulmonary TB	Meudang Ara
17	69	L	December	Pulmonary TB	Blang Awe

The above is patient data obtained from the Bayu puskesmas for the period of 2024, starting from January to December.

4.2. Pre-processing Data

The pre-processing stage is carried out to prepare the data so that it is ready for use. In this research, the pre-processing carried out is data normalization. Normalization helps eliminate these scale differences by converting the data into a uniform scale. There are various normalization methods; the one used in this research is data normalization using min-max normalization. The normalized data is as follows: The table below shows the dataset after normalization for the Bayu Puskesmas area in 2024.

Table 3. Dataset after normalization of Bayu sub-district in 2024

No	Name	Age (Normalized)	Gender(1=L, 2=P)	month (1-12)	Location (Encoded)
1	-	0.7843	2	10	0
2	-	0.1569	1	9	0
3	-	1.0000	1	9	0
…	…	…	…	…	…
15	-	0.6667	1	1	0
16	-	0.5882	2	1	0
17	-	0.9804	1	12	0

Table 4. Dataset after normalization of Baktiya sub-district in 2020

No	Name	Age (Normalized)	Gender(1=L, 2=P)	month (1-12)	Location (Encoded)
1	-	1.0000	1	2	0
2	-	0.8654	1	1	0
3	-	0.1923	2	1	0
4	-	0.1346	2	1	0
5	-	0.0000	2	6	0
6	-	0.6731	2	6	0
7	-	0.3077	1	1	0

4.3. Forming the Similarity Matrix

The similarity matrix is formed using the k-nearest neighbours method, where the working principle is to find the closest distance between the data to be evaluated and its k-closest neighbours, which is calculated using the Euclidean distance formula [18]. In this research, the initialization of the k value uses a value of 2, considering that the optimum k value is not too small or too large and considering the amount of data. The similarity matrix results are represented with a number 1, where the number 1 means that the data are close to each other. For example, below is a similarity matrix of the Lhoksukon District dataset in 2024, which can be seen in the following table:

Table 5. Similarity Matrix of Lhoksukon Sub-district in 2024

	Data-1	Data-2	Data-3	…	Data-24	Data-25	Data-26
Data-1	1.0000	0.5005	0.9384	…	0.3229	0.9125	0.4980
Data-2	0.5005	1.0000	0.6628	…	0.9108	0.6450	0.9941
Data-3	0.9384	0.6628	1.0000	…	0.4773	0.9528	0.6741
…	…	…	…	…	…	…	…
Data-24	0.3229	0.9108	0.4773	…	1.0000	0.5005	0.9171
Data-25	0.9125	0.6450	0.9528	…	0.5005	1.0000	0.6489

Data-26	0.4980	0.9941	0.6741	…	0.9171	0.6489	1.0000

4.4.Constructing the Laplacian Matrix

The Laplacian matrix is a representation matrix of a graph. This Laplacian matrix is the main component that will later be used to classify data with spectral Clustering—the following table of Laplacian matrices calculated based on the normalized Laplacian formula.

Table 6. Laplacian matrix on Lhoksukon sub-district data 2024

	Data-1	Data-2	Data-3	…	Data-24	Data-25	Data-26
Data-1	1.0000	-0.0283	-0.0566	…	-0.0191	-0.0557	-0.0280
Data-2	-0.0283	1.0000	-0.0343	…	-0.0463	-0.0338	-0.0481
Data-3	-0.0283	1.0000	-0.0343	…	-0.0259	-0.0533	-0.0348
…	…	…	…	…	…	…	…
Data-24	-0.0191	-0.0463	-0.0259	…	1.0000	-0.0275	-0.0464
Data-25	-0.0557	-0.0338	-0.0533	…	-0.0275	1.0000	-0.0339
Data-26	-0.0280	-0.0481	-0.0348	…	-0.0464	-0.0339	1.0000

4.5. Calculating Eigen Values and Eigen Vectors

Eigenvalue and eigenvector help generate new data that more deeply represent the similarity characteristics between data [23]. First, the eigenvalue is calculated, and after it is known, the eigenvector is formed based on the year, sub-district and number of patients into an eigenvalue. At this stage, eigenvector formation is carried out for each category so that clusters are formed. The following is the eigenvector table that has been formed [24].

Table 7. Eigenvalues and eigenvectors on Lhoksukon Sub-district data 2024

No	Eigenvalue	Eigenvector
1	1.0588	[-0.0283, 0.1086, -0.1895, 0.1418, 0.0951, 0.0948, 0.1103, -0.1748, -0.0904, -0.1468, -0.0994, -0.0557, 0.2721, 0.0675, 0.5115, 0.0948, -0.1881, -0.5281, 0.2501, 0.0886, 0.0872, 0.0951, -0.1866, -0.1877, -0.0705, 0.0495]
2	1.0624	[-0.0391, -0.0024, -0.1457, 0.2252, 0.0555, 0.1203, 0.0524, -0.1669, 0.2044, -0.1774, -0.1747, -0.0210, 0.6119, -0.0123, -0.4567, 0.1203, -0.0249, 0.2659, -0.1548, 0.0980, 0.0675, 0.0555, -0.1719, -0.1846, -0.0186, -0.0102]
3	1.0580	[0.0210, -0.0045, -0.0900, 0.2934, 0.0421, 0.1200, 0.0383, -0.1003, 0.1778, -0.1045, -0.0923, -0.0694, -0.4722, -0.0205, 0.3989, 0.1200, 0.0385, 0.4513, -0.2443, 0.0933, 0.0565, 0.0421, -0.1304, -0.3543, -0.0580, -0.0209]

4.6. Determining Clustering Results

The eigenvector results of each category are then grouped based on the analysis of eigenvalues obtained from the Laplacian matrix. This process produces a more optimal representation in low-dimensional space, so it can be used to determine TB patients' Clustering accurately. From the clustering results that have been carried out, patient data at three health centres, namely Puskesmas Baktiya, Puskesmas Lhoksukon, and Puskesmas Bayu, were successfully grouped into three categories based on the severity of the disease, namely high, medium, and low. This Clustering was done by considering the number of patients in each region as well as the individual characteristics of the patients. The final results of this Clustering are shown in the table below, which illustrates the distribution of patients by severity in each puskesmas.

Table 8. Spectral Clustering Results of Patient Data

No	District	Year	Number of Patients	Average Age	Gender		Most Months	Silhouette Score	Characteristics Cluster
				Year	Male	Female			
1	Baktiya	2020	7	48,6	42,9%	57,1%	January	0,0000	Cluster 1
		2021	22	49,8	81,8%	18,2%	March	0,0000	Cluster 1
		2022	14	48,6	92,9%	7,1%	August	0,1806	Cluster 1
			5	42,2	60%	40%	November		Cluster 2
		2023	4	40,3	25%	75%	May	0,2355	Cluster 1
			10	50,6	40%	60%	February		Cluster 2
		2024	34	45,7	41,2%	58,8%	November	0,2732	Cluster 1
			5	56,6	40%	60%	March		Cluster 2
2	Bayu	2020	3	52	66,7%	33,3%	September	0,3071	Cluster 1
			7	43,6	42,9%	57,1%	August		Cluster 4
			15	48,7	60%	40%	June		Cluster 5
		2021	14	40,4	64,3%	35,7%	November	0,0000	Cluster 1

		2022	16	42,7	43,8%	56,3%	August	Cluster 1
			8	46,9	75%	25%	Desember	0,4968	Cluster 5
			17	44,4	52,9%	47,1%	October	Cluster 6
		2023	27	41,8	44,4%	55,6%	November	0,4200	Cluster 1
			6	46,2	66,7%	33,3%	February	Cluster 2
		2024	14	47,8	57,1%	42,9%	July	0,3764	Cluster 1
			3	45,7	33,3%	66,7%	February	Cluster 2
		2020	8	52,5	62,5%	37,5%	March	0,3965	Cluster 1
			2	50,5	50%	50%	October	Cluster 4
			16	49,2	43,8%	56,3%	January	Cluster 6
		2021	4	54	50,0%	50,0%	March	0,4171	Cluster 1
			19	48,3	68,4%	31,6%	June	Cluster 2
3	Lhoksukon	2022	35	51,4	62,9%	37,1%	Desember	0,0000	Cluster 1
		2023	20	49,2	75%	25%	September	0,3020	Cluster 1
			6	46,5	50%	50%	February	Cluster 2
		2024	12	48,7	66,7%	33,3%	May	0,2263	Cluster 1
			7	46,3	85,7%	14,3%	April	Cluster 5
			7	37,1	85,7%	14,3%	June	Cluster 6

4.4. System implementation

System implementation is implementing and operating a system designed to function as expected. This process includes software and hardware installation, data migration from the old system to the new system if needed, and testing to ensure the system runs without significant bugs or errors. In addition, user training is conducted so that they can use the system properly, followed by

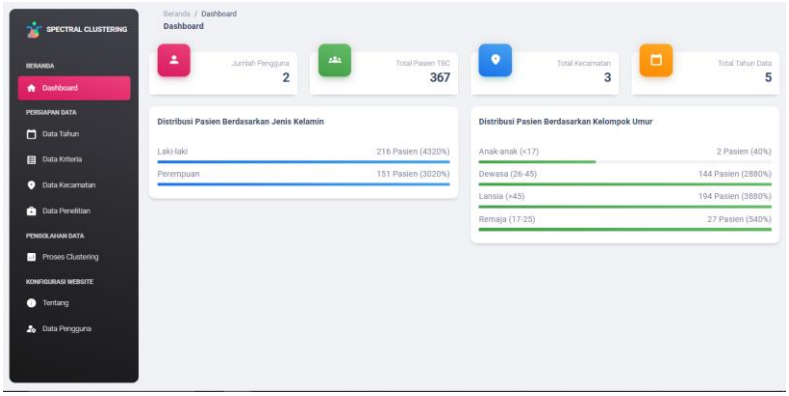


Fig 1. Dashboard Page (Indonesia)

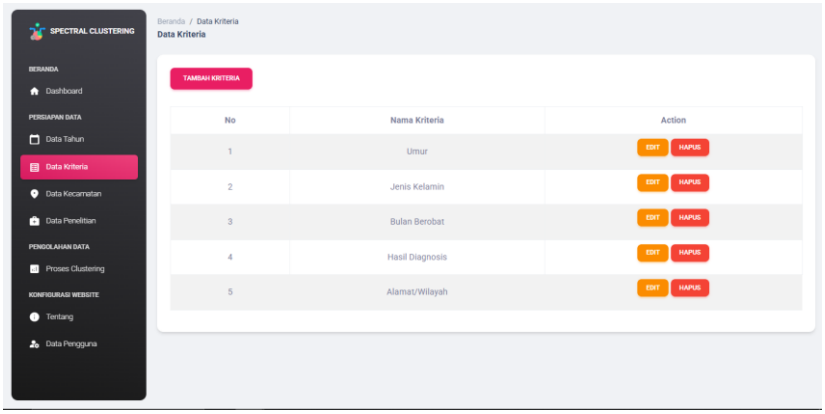


Fig 2. Criteria Data Page (Indonesia)

Figures 1 and 2 are the system dashboard page and the criteria data page on this page; all menus in the system are displayed. The dashboard page shows all the menus in the system, such as year data, criteria data, sub-district data, research data, clustering process, and information about the system and user data. On the criteria, the data page contains variables from patient data at the Baktiya, Bayu and Lhoksukon health centres. The requirements selected are the most influential in the spectral clustering process.

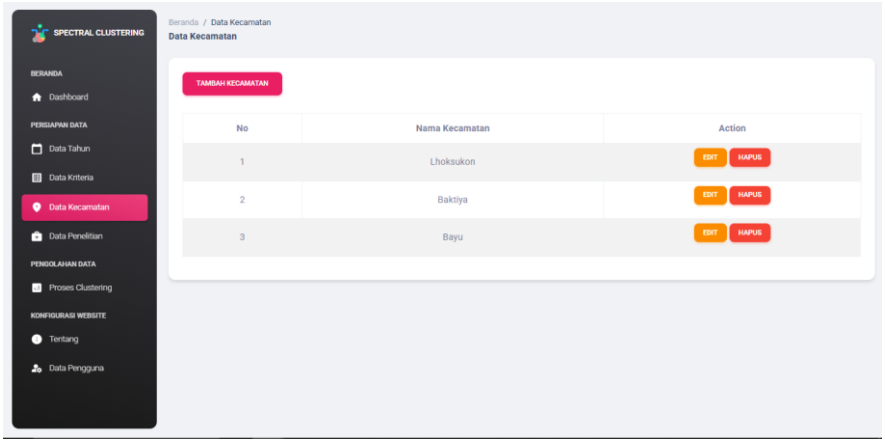


Fig 3. Subdistrict Data Page (Indonesia)

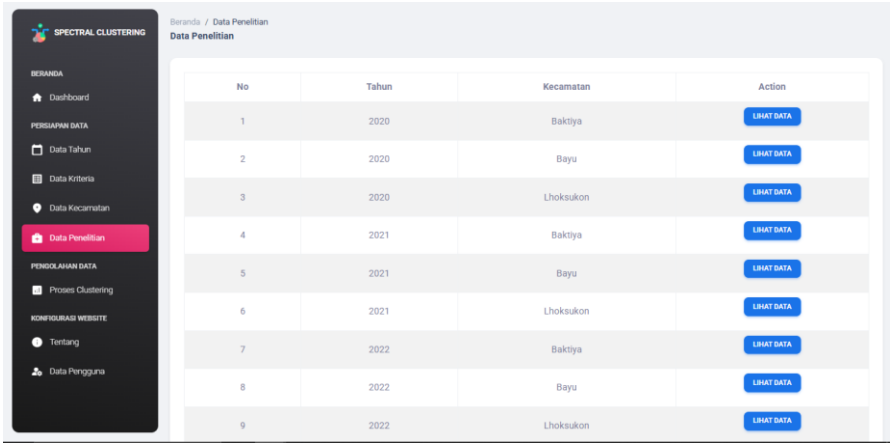


Fig 4. Research Data Page (Indonesia)

The sub-district data page above displays sub-district data on spectral clustering research, and the research data page is taken from a collection of all-year data, criteria data and sub-district data that has been filled in previously. In the research data, for the patient data input process, the admin presses the “view data” button to fill in patient data.

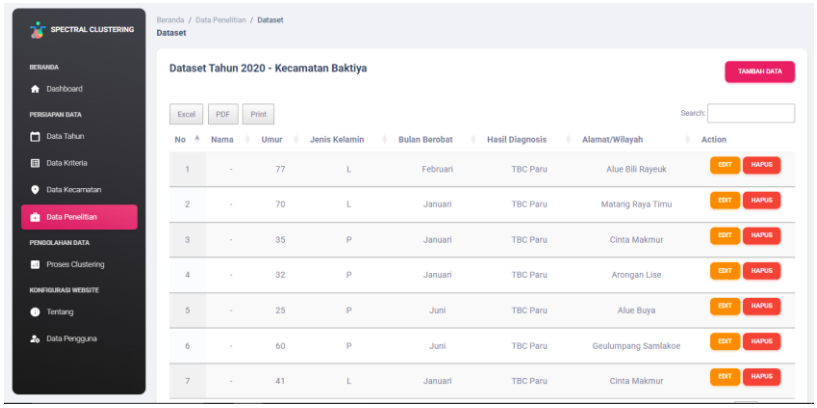


Fig 5. Baktiya sub-district dataset page 2020 (Indonesia)

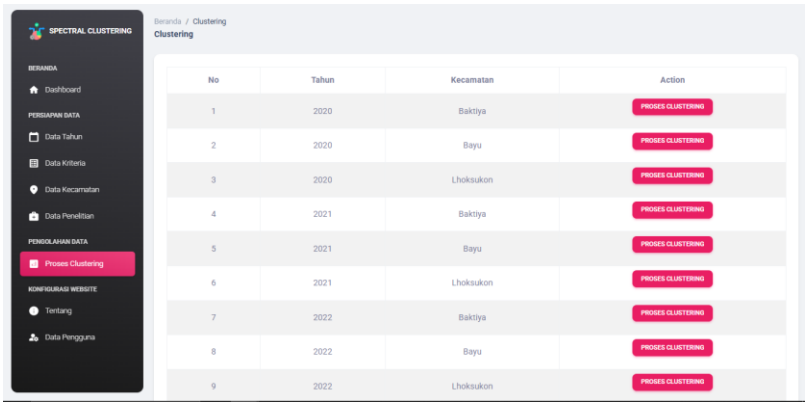


Fig 6. Clustering Process Page (Indonesia)

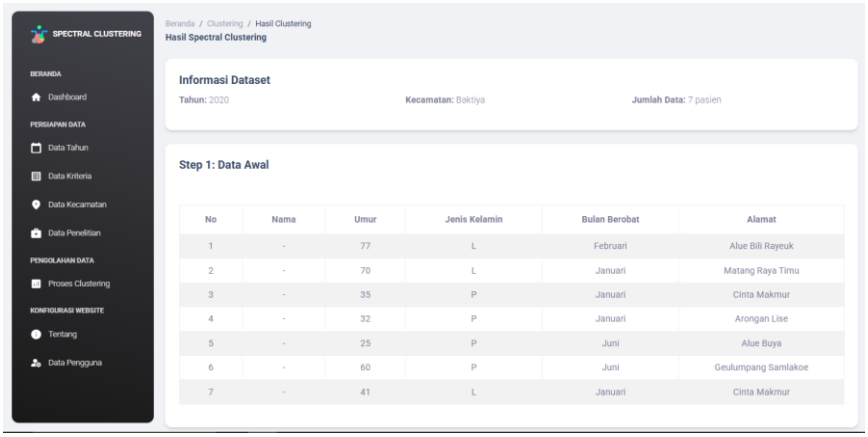


Fig 7. Spectral Clustering Results Page (Indonesia)

The spectral clustering results page displays the calculation results, starting from the initial data table, data after normalization, similarity matrix, normalized Laplacian matrix, eigenvalues & eigenvectors, and clustering results.

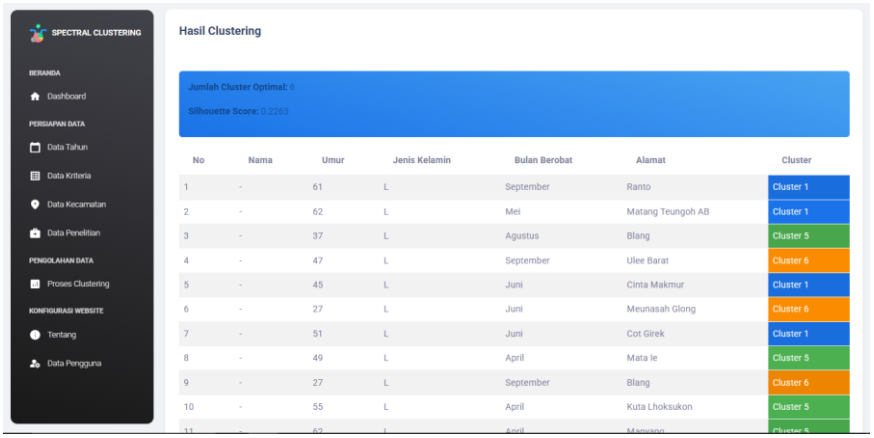


Fig 8. Clustering Result Page (Indonesia)

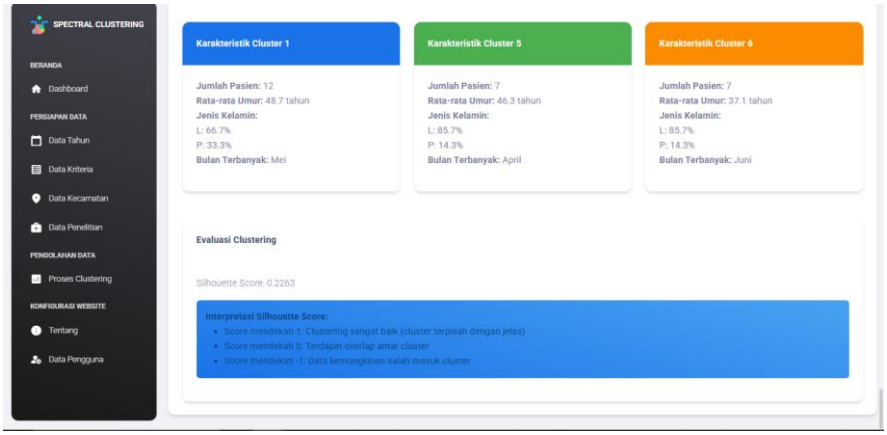


Fig 9. Clustering Result Page (Indonesia)

The final result of the spectral clustering system displays the cluster's final results, which display the number of patients, the average age of patients, the percentage of gender, and the number of patients with TB disease in most months. From all the research data that has been clustered, the admin will determine which area has the highest level of TB patients based on the research data that has been clustered.

5. Conclusion

The conclusion of this study shows that TB patient data from Puskesmas in several sub-districts in the 2020-2024 period is divided into clusters based on the distribution level. Lhoksukon sub-district is included in the cluster with the highest level of patients, with 136 patients or 37.06% of the overall data. The average patient in this sub-district is 48.6 years old, with the majority being male (63.70%) and the peak number of patients occurring in December 2022. Meanwhile, the Bayu sub-district is included in the low-level cluster, with 130 patients or 35.42%. The average age of patients in this area is 45.5 years, with a composition of 55.19% males and 44.82% females, and the highest number of patients was recorded in November 2023. Baktiya sub-district, which is included in the lowest cluster, had the fewest total patients, 101 patients or 27.52%, with an average patient age of 47.8 years and a majority of males (52.8%). The peak number of patients in this sub-district occurred in November. From the results of this Clustering, it can be concluded that the Lhoksukon sub-district has the highest TB case rate, while the Baktiya sub-district has the lowest case rate during the study period.

References

[1] A. J. Melya Findi, “Dampak Polusi Udara Pada Meningkatkan Penderita Tuberkulosis,” *Sobat TB. Yayasan KNCV Indonesia.*, 2022. .

[2] R. D. Cahyaningrum, A. Bustamam, and T. Siswantining, “Implementation of spectral clustering with partitioning around medoids (PAM) algorithm on microarray data of carcinoma,” *AIP Conf. Proc.*, vol. 1825, no. March, 2017, doi: 10.1063/1.4978976.

[3] D. Ismail, “Recognition System Of The Al Qur’an Surah Al-Falaq Verse 1-5 Through Voice Using Ada-Boost,” *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 3, 2021, doi: 10.52088/ijesty.v1i3.85.

[4] D. Trivedi, C. D. Rahn, W. M. Kier, and I. D. Walker, “Soft robotics: Biological inspiration, state of the art, and future research,” *Appl. Bionics Biomech.*, vol. 5, no. 3, pp. 99–117, 2008, doi: 10.1080/11762320802557865.

[5] Kemenkes, “Kesehatan Di Fasyankes Tingkat Pertama (Fktp) Pasien Tuberkulosis.”

[6] F. Matriadi, I. Ikramuddin, M. Adamy, and C. Chalirafi, “Implementation of Digital Leaderships on Pertamina Hulu Energy in Aceh,” *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 125–129, 2021, doi: 10.52088/ijesty.v1i2.132.

[7] A. Muthiah, U. Padjadjaran, J. Prof, E. No, K. Sukajadi, and K. Bandung, “Jurnal Epidemiologi Kesehatan Indonesia Karakteristik Pasien Tuberkulosis Lost to Follow Up dari Empat RS di Kota Bandung Karakteristik Pasien Tuberkulosis Lost to Follow Up dari Empat RS di Kota Bandung Characteristics of Tuberculosis Lost to Follow Up Patients from Four Hospitals in Bandung,” vol. 3, no. 1, 2019, doi: 10.7454/epidkes.v3i1.3208.

[8] U. L. U. Palembang, “Unnes Journal of Public Health,” vol. 6, no. 4, 2017.

[9] E. Mutiara and S. Novita, “Characteristics of Multi-drug Resistant Tuberculosis (MDR-TB) Patients in Medan City in 2015-2016,” pp. 2–7, 2018.

[10] “Mycobacterium tuberculosis ,” pp. 10–21, 2017.

[11] A. T. Sitanggang and Y. Desnelita, “Tingkat Pemahaman Mahasiswa antar Pembelajaran Online dan Offline dalam Masa Pandemi Covid-19 Menggunakan Metode Forward Chaining,” *J. Inf. dan Teknol.*, vol. 4, no. 1, pp. 64–69, 2022, doi: 10.37034/jidt.v4i1.187.

[12] K. H. Manurung and A. E. Syaputra, “Expert System Diagnosis Penyakit Asma Bronkial dengan Certainty Factor pada Klinik Sari Ramadhan Berbasis Web,” *J. Sistim Inf. dan Teknol.*, vol. 5, no. 2, pp. 4–7, 2023, doi: 10.37034/jsisfotek.v5i1.218.

[13] M. N. Bustan, *Pengantar Epidemiologi*. 2006.

[14] *Buku Ajar Ilmu Penyakit Dalam Edisi 6 Jilid 1*. .

[15] F. Zahra, M. A. Ridla, and N. Azise, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap),” *JUSTIFY J. Sist. Inf. Ibrahimy*, vol. 3, no. 1, pp. 55–65, 2024, doi: 10.35316/justify.v3i1.5335.

[16] Y. S. Haryanti, Syarifah Nur , Nugroho, “Perbandingan 3 Metode Dalam Data Mining Untuk Penjurusan Siswa Di SMA N 3 Boyolali,” 2015.

[17] I. M. A. W. Putra, G. Indrawan, and K. Y. E. Aryanto, "Sistem Rekomendasi Berdasarkan Data Transaksi Perpustakaan Daerah Tabanan dengan menggunakan K-Means Clustering," *J. Ilmu Komput. Indones.*, vol. 3, no. 1, pp. 18–22, 2018.
[18] N. Salsabila, "Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang (Studi Kasus : Ukm Mar ' Ah Jilbab Kediri)," 2019.
[19] M. Meilă, "Spectral Clustering: a Tutorial for the 2010's," *Handb. Clust. Anal.*, p. 753, 2015.
[20] A. Yudhana, S. Sunardi, and A. J. S. Hartanta, "Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggergajian Kayu Sengon," *Transmisi*, vol. 22, no. 4, pp. 123–129, 2020, doi: 10.14710/transmisi.22.4.123-129.
[21] Soegiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. 2011.
[22] Prof.Dr.Sugiyono, "Metode_Penelitian_Pendidikan_Sugiyono_20.pdf." pp. 47–281, 2015.
[23] Q. R. Hidayati and S. Surono, "Implementasi Algoritma Spectral Clustering Untuk Analisis Sentimen," *Delta J. Ilm. Pendidik. Mat.*, vol. 9, no. 1, p. 27, 2021, doi: 10.31941/delta.v9i1.1229.
[24] t bayu Kurniawan and Syarifuddin, "Perancangan Sistem Aplikasi Pemesanan Makanan dan Minuman Pada Cafetaria NO Caffe di TAnjung Balai Karimun Menggunakan Bahasa Pemrograman PHP dan MySQL," *J. Tikar*, vol. 1, no. 2, pp. 192–206, 2020.