Terbit *online* pada laman web jurnal : https://jes-tm.org/index.php/jestm/index

Journal of Engineering Science and Technology Management

| ISSN (Online) 2828-7886 |

Article

Mortality Rate Analysis Covid-19 Patients Based on Condition Comorbidities with Approach K-Means Clustering

Adri Yanto[1]✉, Epu Margi Astuti[2], Aai Mustika[3]
Al Insyirah Institute of Health and Technology, Pekanbaru, Indonesia(1, 2)
AMIK Mahaputra Riau, Pekanbaru, Indonesia(3)

E-mail: adriyanto@ikta.ac.id

ARTICLE INFORMATION

Volume 5 Issue 1
Received: 16 February 2025
Accepted: 29 March 2025
Published Online: 31 March 2025
Online: at https:JESTM.org/

Keywords

Covid-19
Comorbidities
K-Means Cluster
Rapidminer
ICD-10

A B S T R A C T

The COVID-19 pandemic, which emerged in China in late 2019, rapidly spread to over 200 countries, including Indonesia. In response, various preventive measures were implemented to mitigate the increasing number of infections. West Sumatra Province ranked 11th out of 34 provinces in terms of confirmed COVID-19 cases. Patients infected with COVID-19 in this region often presented with comorbidities such as hypertension, diabetes mellitus, chronic obstructive pulmonary disease (COPD), cardiovascular disease (CVD), liver disorders, obesity, renal disease, and malignancies conditions known to contribute significantly to COVID-19-related mortality. This study aims to identify the mortality risk associated with comorbidities in COVID-19 patients using clustering analysis. A total of 91 patient records from a hospital in Padang City were analyzed. The data included age, primary and secondary diagnoses, all classified using the International Classification of Diseases (ICD-10). The K-Means Clustering algorithm was employed to categorize comorbidities into high, medium, and low-risk groups. Data processing was conducted using RapidMiner software, and accuracy was evaluated using mathematical calculations. The results indicated that Cluster 1 (high-risk) consisted of 5 comorbid diseases, Cluster 2 (medium-risk) included 9 diseases, and Cluster 3 (low-risk) encompassed 35 comorbid diseases. These findings demonstrate that clustering analysis is effective in classifying mortality risk based on comorbid conditions and can be utilized to support clinical prioritization and resource allocation during pandemic management.

1. INTRODUCTION

Coronavirus disease , also known as as COVID-19, is disease infectious disease caused by a virus called syndrome breathing I severe coronavirus 2 (SARS-CoV-2) from the family of viruses that cause diseases in humans and animals . SARS-CoV-2 is believed to in a way official originate from bats and spread to humans [1][2]. SARS CoV-2, also known as the corona virus is an infectious virus that causes COVID-19 and is included in family coronaviridae . One of the symptom general from this virus like cough , fatigue , fever , and respiratory disorders breathing , which is a number of symptom similar to flu besides lost senses smell and taste , pain head , and pain muscles [2]. Source this virus is important No identified , analysis order the genome of this virus show that this virus including in the genus COV from coronavirus family , animal rodents and bats is host most For this type of virus [2]. Air can become carrier infection besides contact direct with infected people , and infiltrate to cells raspitory with attaches to angiotensin converting enzyme 2 (ACE2) [2][3].

The first Covid-19 infection was confirmed at the end of month December 2019 in China , precisely in the city of Wuhan. Pandemic This has spread to all countries in the world, and on January 30, 2020 the World Health Organization (WHO) stated that pandemic This as world concern about crisis health in a way general , doctors who work same and specialist public others around the world are struggling oppose pandemic This [2][3]. A highly contagious disease . This causes pneumonia and disease breathing severe similar with SARS and MERS with level global mortality is around 6.13% [1]. As of December 31 , 2021, more than from 287 million confirmed cases of COVID - 19 worldwide have reported to WHO including 5.4 million death [4][5][6]. Rapid transmission of the virus from person to person can happen through droplets or contact and if control infection No strict or No available tool protector adequate self , thing This can endanger officer health line first [7][8].

This virus with fast become global pandemic , causing problems global big , including problem health , economy , preferences age , and others [1]. In addition , COVID-2019 is increasingly develop breed all over the world and even has spread in a way No on purpose to 200 countries in the world [3]. The Covid-19 pandemic is things experienced almost all countries in the world, Indonesia is one of them [4], as a result Lots infected patient died [5][9], one of which was in the province of West Sumatra. Based on data from the West Sumatra Provincial Health Service as of March 11, 2021 , 29,985 people were declared positive for COVID-19, 659 of them died and 28,297 people were declared recovered , while in the case of COVID-19 in Padang City 14,820 people were declared positive for COVID-19, 288 of them died and 14,188 were declared cured [10][11][12], and this virus in a way No on purpose has burdensome system service health in Indonesia and especially in the West Sumatra Province [6].

The height level death patient due to COVID-19 in Padang city is based on the disease comorbid like hypertension , diabetes, COPD, CVD, liver diseases, obesity, renal diseases and malignancy [7][13] . comorbid be one of factor risks that can occur increase death patient consequence COVID-19 pandemic [9][12]. Disease comorbid grouped become category high , medium , and low . Data on patients infected with COVID-19 and other diseases accompanying obtained from Dr. M. Djamil General Hospital , Padang City, West Sumatra Province . Patient data the will classified with the International statistical classification of diseases and related health problems (ICD) about diseases and problems health revision to 10 [14].

Death data patient due to COVID-19 which is accompanied by disease accompanying usually saved in record Hospital medical and not yet There is classification disease accompanying deceased patient consequence the COVID-19 pandemic , then required A method For grouping or disease cluster comorbidities in patients infected with COVID-19 [15] , and grouping mortality data patient due to COVID-19 based on disease companion [7][16]. The method used For grouping level death COVID-19 patients based on disease default use K -Means Cluster (KMC) method . KMC is a techniques used in do data grouping [17][18][19]. This KMC algorithm used For share COVID-19 patient data accompanied by disease default become cluster grouping disease comorbid . Algorithm This make cluster documents that have more between-cluster variation (BCV). tall compared to with within cluster variation (WCV) [11][16].

With use K -Means Cluster method can produce cluster death COVID-19 sufferers based on disease accompanying and identifying COVID-19 deaths due to disease the companion who is one of the factor risk COVID-19 deaths , and using application rapidminer For get cluster data and usage K -Means Cluster method to in system rapidminer . K -Means clustering method often used For grouping large data . Grouping large data set manually can eat time and vulnerable to error . With K- Means method clustering , the process of grouping data can walk in a way automatic so that can focus on analysis and interpretation results only . On the basis of that's it so required application rapidminer as data management software based on K- Means cluster algorithm .

2. LITERATURE REVIEW

2.1 Data Mining

The development of information technology has produced large amounts of databases and data in various fields. Research in the field of databases and information

technology has given rise to approaches to storing valuable data for further decision making. Data Mining is a logical process used to search large amounts of data to find useful data and become information. Data mining is currently increasingly sophisticated, reflecting a combination of statistics, data science, database theory, artificial intelligence, and machine learning practices. The goal of this technique is to find previously unknown patterns. Once these patterns are found, these patterns can then be used to make certain decisions [20]. Data mining is also known as knowledge discovery in data (KDC) which is the process of uncovering patterns and other valuable information from large data sets. Data Mining can be referred to as knowledge mining form data , knowledge extraction, data/pattern analysis, data archaeology, and data dredging [21].

data mining process relies on the implementation of effective data collection and processing. Data mining can be used to describe a target data set, predict outcomes, detect fraud or security issues, learn more about a user base, and detect bottlenecks. Data mining is more useful today because of the large growth of data and the need for data storage. Data specialists who use data mining must have experience in coding and programming languages, as well as statistical knowledge to process and interpret data. Various techniques can be used for data mining in various data science applications. Pattern recognition is a common data mining use case, as is anomaly detection, which helps identify outlier values in a data set. Popular data mining techniques include classification, prediction, association rules, neural networks, and clustering [22].

2.2 K-Means Clustering

Data clustering is an important research topic and has applications in various fields such as statistics, data mining, computer science, pattern recognition, image processing, and marketing. Clustering is considered a pure multivariate technique but can also be applied to univariate and bivariate data . Clustering is done on the basis of similarity or distance and it is one of the best multivariate analysis approaches and a common methodology for statistical data analysis [23]. Data mining technology can be used to process piles of data in databases to uncover new, interesting, and useful information. Clustering is an approach to data collection as one of the techniques for grouping data into clusters or groups [24]. The K- Means Clustering algorithm is the most popular clustering formulation whose goal is to maximize the expected similarity between data items and their associated cluster centers [25].

K- Means is an iterative procedure that partitions N objects into K separate cluster.

K- Means is probably the most widely used clustering method, and especially the most well-known partition-based clustering method that uses centroids for cluster presentation. The quality of K-Means clustering is measured through the within- cluster squared error criterion [23]. The K- Means algorithm is used to minimize the K- Means problem , and has many variants that will be discussed later but in order to use any K - Means algorithm, the number of clusters present in the data must be the same. known; several runs or trials will be needed to find the best number of clusters . There is no best K- Means algorithm, because the tendency to produce a global optimum depends on the characteristics of the data set, its size and also the number of variables in the case. The K- Means clustering method has two iteration phases, namely the assignment or initialization phase which involves an iterative process where each data point is assigned to the nearest centroid using the euclidean metric. Next is the centroid update phase , where the cluster centroids are updated based on the partitions obtained in the previous phase. The iteration process stops when there are no cluster changes in data points or the maximum number of iterations is reached [23].

Forgy [25] proposed a batch algorithm called the traditional k-means algorithm, this algorithm is based on minimizing the mean squared euclidean distance between data points and cluster centers known as centroids , where centroids are the centers of geometric objects and are seen as a generalization of the mean . Forgy's algorithm starts by selecting the number of clusters k that represent the cluster centers, then assigns data points from the data set to the clusters that have the closest centroids , updates the new centroid for each cluster by averaging the data points or objects belonging to the cluster, if there is no change in the cluster centers, then the iteration is stopped. Lloyd [26] proposed the standard K- Means algorithm which is also a batch algorithm, the difference between Forgy's algorithm and Lloyd's algorithm is that Forgy's algorithm treats the data distribution as continuous while Lloyd's algorithm treats the data distribution as discrete.

One of the main stages in implementing the K- Means Cluster algorithm is to determine the centroid , number of clusters, and centroid distance . The first stage is to perform clustering using K- Means Cluster by determining how many clusters will be created. Next, determine the centroid of the existing data. This stage is carried out repeatedly until no more data is moved. After determining the centroid, you can calculate each data distance to the centroid , using equation 1 below.

$$d_{euclidean}(x, y) = \sqrt{\sum_{i=1}^{n} (xi - xy)^2} \quad (1)$$

Where d(x,y) is the distance of data x to the center of cluster y, xi is the i-th data on the n-th data attribute, and yi is the j-th data on the n-th data attribute. After getting the distance of data to the centroid, the next step is to group the data based on the shortest distance. Then find a new centroid based on the amount of data from each cluster by calculating the data from each cluster.

2.3 Rapidminer

Rapidminer studio is a visual design application for quickly building complete predictive analytics workflows. It provides a deep library of machine learning algorithms, data preparation and exploration functions, and model validation tools to support all data science projects. Rapidminer provides an integrated environment for machine learning, data mining, text mining, predictive analytics, business analytics , business and industrial applications as well as for research, education, training, prototyping and application development. Rapidminer supports all steps of the data mining process including results visualization, validation and optimization. The application provides data mining and machine learning procedures including data loading and transformation, data preprocessing and visualization, predictive analysis and statistical modeling, evaluation and deployment. Rapidminer is written in the Java programming language , provides a GUI for designing and executing analytical workflows. The workflows are called “Processes” in Rapidminer and consist of several operators. Each operator performs one task in the process and the output of each operator forms the input for the next task. The GUI will generate an XML (extensible markup language) file that defines the analytical process the user wants to apply to the data. This file is then read by rapidMiner to run the analysis automatically [27].

Rapid Miner provides learning schemes and models and algorithms from Weka and R scripts that can be used through extensions. Rapidminer's functionality can be extended with additional plugins . The rapidmine r extensions market provides a platform for developers to create data analysis algorithms. Rapidminer has about 100 learning solutions for clustering, classification and regression analysis. Rapidminer also supports about 22 file formats , such as. xls, csv, and so on [28].

2.4 Comorbid

In general, the term comorbid means two or more medical conditions that occur simultaneously and are interdependent, meaning that one medical condition causes, is caused by, or is related to another condition in the same individual [29]. Comorbid diseases are usually long-term or chronic. They may or may not interact with each other. Doctors use the term comorbidity to understand and explain how these conditions can affect physical and mental health, either simultaneously or separately. Comorbidity refers to accompanying diseases by other names, such as coexisting or co-occurring conditions. Other commonly used terms include multimorbidity or multiple chronic conditions [30].

For example, if someone is diagnosed with both social anxiety disorder and major depressive disorder, they are said to have comorbid (meaning they coexist) anxiety and depression. Other comorbid conditions include physical illnesses such as diabetes, cardiovascular disease, cancer, infectious diseases, and dementia [30]. It is important to note that comorbidities are not the same as complications. Complications are side effects or medical problems that may occur during an illness or after a procedure or treatment. They may be caused by the illness, procedure, treatment, or be unrelated.

Comorbidity is a separate disease that occurs at the same time as the main health problem. As people age, older people are more likely to experience health problems. However, younger people can also have comorbidities. A study looking at data from more than 1 million people found that many things can cause someone to experience multimorbidity, or two or more diseases at the same time, such as the type of disorder (physical or mental), sex, age, and socioeconomic status.

Some examples of comorbid diseases include heart disease, high blood pressure, respiratory diseases, mental health problems such as dementia , cerebrovascular disease, joint disease, diabetes, sensory disorders, and arthritis. Comorbid diseases carry a risk of experiencing health impacts such as functional limitations, treatment complications, or even death [30].

3. METHODOLOGY

3.1 Analysis And Design Stages

For do design and manufacture system need done A analysis in a way structured . The system that will made later is system For identify COVID-19 death data based on disease comorbid as studies the case . Before doing design to system , need done analysis to the data to be used in research This with use K - Means Clustering method . After finished do analysis , next will designed a computerized system with use same method use application rapidminer .

In research This is the main data used is disease data comorbidities in COVID-19 patients in 2020 amounted to 91 COVID-19 patient data . After cleaned up only remaining 49 patient data . The variables used in study This in the form of age , code disease code data companion disease is code data based on international standard classification of diseases (ICD10). the data will processed use K -Means Clustering Method assisted with Microsoft Excel for search manual . Disease data comorbid 2020 used as test data For grouping disease influencing factors COVID-19 sufferers later . Before the data is processed, the data must be made into numeric Because code disease in the form of letters and numbers.

Code data then initialized moreover formerly in form numeric after That new Can done calculation. Calculation done with use K -Means Cluster algorithm .

As for the purpose from algorithm This that is grouping data by maximize similarity of data in One cluster and minimize similarity of data between cluster. Function distance is size similarity used in cluster , so that maximum results from data similarity based on distance shortest between data and point centroid. Stages of K- Means Clustering can seen in the flowchart in Figure 1.

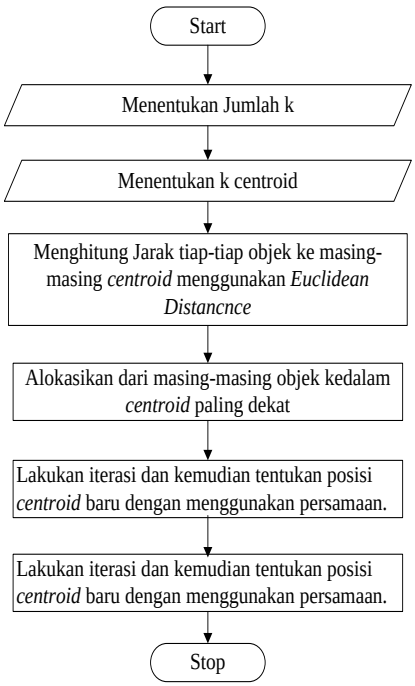

Figure 1. Flowchart of K-Means Clustering Process

The first stage is to determine the K value from COVID-19 patient data at RSUP M.Djamil Padang City to determine the number of clusters formed. Furthermore, after the K value is obtained, the next stage is to determine the K value of the cluster center point (centroid) which is done randomly. After obtaining the K value, and the centroid value, calculate the distance from each object to each centroid of each cluster euclidean distance (equation 1). The next stage is to allocate from each object to the closest centroid . Then perform iteration calculations and determine and then determine the position of the new centroid using equation 2 below.

$$D_{i,j} = \sqrt{(X_{1i} - X_{1j})^2} + (X_{2i} - X_{2j})^2 + ... + (X_{xi} - X_{xj})^2 \quad (2)$$

Based on steps in K -Means Clusters algorithm , repeat count distance every object to each centroid until centroid data obtained same new .

3.2 Determining The Cluster Value Of Each Data

In this case, it is necessary to determine which cluster value is closest to the data, so it is necessary to calculate the distance of each data with the centroid of each cluster . The calculation starts from iteration-1 to iteration-n where when grouping data in iteration-1 (K1) with grouping data in iteration-n (Kn) the results are the same and there is no change, then the calculation process is stopped and the grouping has been determined. Before the comorbid data of COVID-19 patients is processed, the data is initialized first so that the data can be processed using the K- Means Clustering technique as in Table 1 below.

Table 1 Data On Commanding Diseases Of Covid-19 Patients

No	Age	S1	S2	S3	S4	S5
1	54 years old	I10				
2	58 years old	I13.1	N18.5			
3	34 years old	C34.2	Z34.8			
...	...	...	...	...	...	...
49	20 years	C91.0	D63.0	D69.0		

Data on comorbid diseases of COVID-19 patients are grouped into clusters S1 to S5, consisting of 48 patients in cluster S1, 20 patients in cluster S2, 7 patients in cluster S3, 2 patients in cluster 4, and 2 patients in cluster 5. From these data, there are 30 patients with 1 comorbid, 13 patients with 2 comorbid, 4 patients with 3 comorbid, 0 patients with 4 comorbid, and 2 patients with 5 comorbid. From the 49 data, which were previously in the form of disease codes in the form of letters and numbers that were grouped based on the statistical classification of diseases and related health problems 10 (ICD 10).

The next stage is that the data will be initialized into numbers, and then converted into numeric form so that calculations can be carried out based on Table 2 below.

Table 2 Data After Initialization Into Numbers

No	Age	S1	S2	S3	S4	S5
1	54 years old	18	0	0	0	0
2	58 years old	21	32	0	0	0
3	34 years old	3	38	0	0	0
...	...	...	...	...	...	...
49	20 years	4	6	10	0	0

After being initialized into a number, the first stage is to determine the cluster center by taking the value from one of the cluster data 1-5 using the equation below.

$$\mu = \frac{\sum_{i=1}^{n} xi}{n}$$

cluster center can be taken from the data in Table 1.

For C1, the maximum value (Max) of the S1-S5 values is taken as the initial centroid .

C1 = 35 ; 38 ; 39 ; 13 ; 30

For C2, the Middle value of the secondary values 1-5 is taken as the cluster center .

C2 = 16.55 ; 8.87 ; 3.22 ; 0.49 ; 0.89

For C3, the smallest/below (minimum) value of the secondary values 1-5 is taken as the cluster center .

C3 = 0 ; 0 ; 0 ; 0 ; 0 ; 0

After getting the initial centroid value , the calculation is carried out using the equation below.

$$D_{i,j} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + ... + (X_{xi} - X_{xj})^2}$$

centroid data , the calculation can be started from the first iteration to the last iteration. In the calculation of iteration data, the closest distance from each iteration can be taken as a cluster from each iteration 1 to iteration 49 and the closest distance from each cluster is obtained.

3.3 Iteration Calculation

In the research This disease comorbid grouped become category high (C1), medium (C2), and low (C3) clusters , so can done calculation iteration data distance from center cluster . Calculation iteration 1 is done with use equation 3. The value of cluster S1 in Table II is X_{1i} which has a value of 18, and the value of X_{1j} taken from mark The initial centroid C1 is 35, and the value of $X_{is\ 2i}$ taken from S2 value in Table II. While X value $_{2j}$ from C1 value at centroid to 2, and so on until with S5 value in Table II in iteration First . Because there are 49 COVID-19 patient data with comorbid , then calculation iteration done up to data 49. Values from 5 centers cluster used For count COVID-19 patient data iteration with comorbid . From equation 3, then can done calculation mark iteration First until with iteration to nine , where each iteration has 49 COVID-19 patient data. After obtained iteration 1 data , then done calculation with equation 3 to with iteration to 9. In the first data calculation Can taken distance closest as cluster namely is at C2 with value 9.6.

Second data calculation Can taken distance closest as cluster , namely data that is in C2 with value 23.7, 3rd data distance closest as cluster is at C2 with value 32.2, the 4th data is in C2 with value 9.6, the 5th data is in C2 with value 9.61, the 6th data is in C1 with value 21.42. Calculation This done up to the 49th data , where the 49th data was obtained distance closest with cluster is at C3 with value 12.32. After do calculation iteration first , then done same calculation with equation 3 for iteration to 2 because not yet the same data was obtained in accordance with stages of the analysis process K -Means Cluster method , where the process of calculating iteration done until position new centroid same. Calculation iteration 2nd done with determine centroid new C1, C2, and C3 for iteration 2. Determine centroid done For every iteration to 3,4,5 and up to iteration to 9 with data of 49 COVID-19 patients with comorbid.

Calculation iteration done up to iteration to 9 because in the calculation results iteration to 9 no there is data that has changed or The same with calculation on iteration 8 then calculation iteration will discontinued. Based on calculation iteration First up to iteration 9th obtained data C 1 : 5, C 2 : 35 and C 3 : 9 , based on mark cluster the so can taken conclusion, highest cluster namely located in C1, cluster currently located at C3 and lowest located at C35. Based on mark cluster the so can taken conclusion , cluster highest namely located in C1, cluster currently located in C3 and cluster lowest located at C35. Next obtained data results from calculation K -Means Cluster method like on below This .

Table 3 Final Results Of Calculation Using K-Means Clustering Method

No	Age	S 1	S2	S3	S4	S5	Cluster
1	54 years old	18	0	0	0	0	2
2	58 years old	21	32	0	0	0	3
3	34 years old	3	38	0	0	0	3
...	...	...	...	...	...	...	...
49	20 years	4	6	10	0	0	2

4 RESULT AND DISCUSSION

From the calculations that have been made done with use K -Means Cluster method For determine disease cluster the burden of COVID-19 patients , then obtained that For disease cluster the highest participant influence death COVID-19 patients , namely with code disease I10, N17.9, K71.9, E44.1, L21.9, K92.1, I11.9, O13, Z37.0, E87.4, E87.6, D69.6, E87.1, O63.0, O82.1. 2. While disease cluster comorbid level being influencing death COVID-19 patients , namely with code diseases I13.1, N18.5, C34.2, Z34.8, E11.9, I50.1, I10, T82.4, A16.2. And disease clusters comorbid level low that affects death COVID-19 patients , namely with code disease I10, D68.9, I61.5, I20.8, I12.0, K25.4, D63.8, B 24, D64.9, D36.7, O60.0, C91.0, D63.0, D69.0, E11.9, K61.0, N18.5, I50.0, A16.2, G40.9, I11.9. Disease codes This Already standardized by the International statistical classification of diseases and related health problems (ICD) 10. If you look at based on disease data with i10 code is disease that affects death COVID-19 patients. In high cluster Lots attacking adults with disease comorbid more from one , there is also 1 teenager with 2 comorbidities . In the moderate cluster Lots attacks the average adult with 2-to-3 diseases comorbid , and there is also 1 child with 2 comorbidities . While at the level low Lots attack adults , teenagers and children with 1 disease comorbid.

4.1 Analysis With Rapidminer Application

In the sub chapter This will done design from identification level death COVID-19 sufferers based on disease comorbidities (companions) using application rapidminer which will built based on analysis that has been presented in the sub- chapter previously . At the stage implementation This applied to in computerized system
with objective For test results from design and analysis that has been done done previously Already in accordance with the system that has been computerized . For do testing system so prepared moreover first the data that will be processed , namely disease data comorbid COVID-19 sufferers . In testing system This using 2020 data in the form of disease data accompanying from disease secondary 1 to secondary 5. Because of the focus discussing data about K-Means , then writer more focus to K- Means cluster operation with use application rapidminer .

Stages use rapidminer that is with add module read excel , clustering and modules performance. After That all the module is connected , in the section module clusterin g part K is made 3 (Function) For determine number of clusters) and parts max runs are filled with 10 , namely maximum its iteration maximum 10 iterations as in Figure 2. After module added , rapidminer program is running For get cluster data disease comorbid COVID-19 patients who cause death . Cluster results on the application The same with results manual calculations consisting of from cluster 0 : 35 items, cluster 1 : 5 items and clusters 3 : 9 items . Next centroid data and cluster data of COVID-19 patients who died were obtained Because disease default as in Figure 2, and Figure 3.

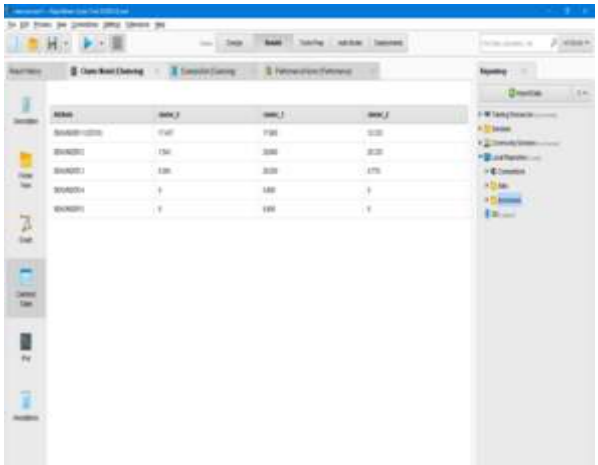

Figure 2. Results of Centroid Calculation with Rapidminer

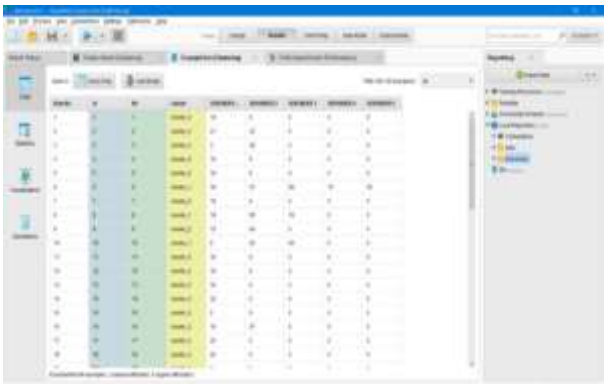

Figure 3. Cluster Grouping Results With Rapidminer

4.2 Discussion With Rapidminer Application

Cluster data with experiments using the rapidminer application with a total of 91 patient data, it was found that the high cluster is located in cluster 2 with 6 items, for medium it is located in cluster 1 with a total of 8 items, and low is located in cluster 0 with a total of 85 items as in Figure 6. The results of cluster calculations manually and using the rapidminer application , it was found that there were similarities between manual calculations and using the rapidminer application. From both calculations, it was found that the calculation in the high cluster attacked many adults with more than two comorbidities, there was also 1 teenager with 2 comorbidities. in the medium cluster , it attacked many adults with an average of 2 comorbidities, there was also 1 child with 2 comorbidities, and at a low level it attacked many adults, teenagers and children with 1 comorbidity.

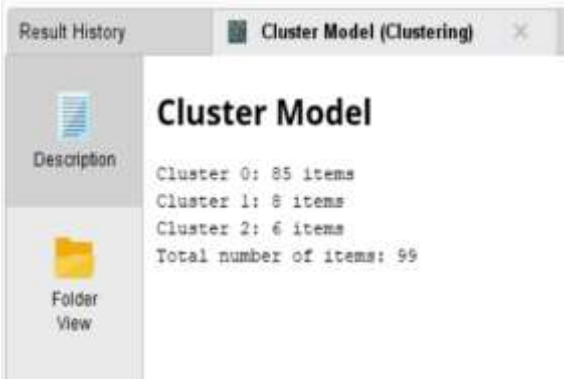

Figure 4. Cluster of Test Results with 91 Patient Data

Based on the data obtained using the K- Means Cluster method , it was found that the deaths of COVID-19 patients based on comorbidities were dominated by diseases with code i10, namely hypertension or high blood pressure.

5 CONCLUSION

Based on testing of data that has been processed above so Can taken conclusion disease comorbidities with cluster high 5 clusters , medium 9 clusters and low 35 clusters . In the cluster tall Lots attacking adults with comorbid more from one , there is also 1 teenager with 2 comorbidities. cluster currently Lots attacks adults on average with 2-3 comorbidities , also occurs in 1 child with 2 comorbidities . While at the level low Lots attack adults , teenagers and children with 1 comorbidity . For study next will more Good done compare mortality data consequence comorbid and data that is still life Because comorbid as well as more data addition Lots .

Conflict of interest

The author declares that this scientific article has no conflict of interest. The results obtained are the author's own thoughts.

Thank you note

The author would like to thank Dr. M. Djamil General Hospital in Padang City, West Sumatra Province, for allowing the 2020 COVID-19 patient data to be obtained, so that this study can be conducted to analyze the mortality rate of COVID-19 patients with comorbidities using the K- Means Cluster method and the rapidminer application .

REFERENCES

James Louis, Jean Magdonald Aime, On the novel coronavirus (COVID -19): a global pandemic, 2020, doi: 10.29238/teknolabjournalv9i1.230.

Janet O. Agbaje, Oluwatosin Babasola, Kabiru Michael Adeyomo, Abraham Baba Zhiri, Modeling COVID-19 Disease with Deterministic and Data-Driven Models Using Daily Empirical Data in the United Kingdom, 2024, doi : 10.3390/covid4020020

Haider Fathi Kadom, Ruaa Majeed Azeez, Analyzing and Predicting the Deaths and Cases of Covid-19 in Different Countries Using Fuzzy K Means Algorithm and Exponential Method, 2021, doi: 10.1109/IT-ELA52201.2021.9773676

Prabhat Jha , Patrick E Brown Rashid Ansumana, Counting the global COVID-19 dead, 2022, doi: 10.1016/S0140-6736(22)00845-5.

William Msemburi, Ariel Karlinsky, Victoria Knutson, Serge Aleshin-Guendel, Somnath Chatterji, Jon Wakefield, The WHO estimates of excess mortality associated with the COVID-19 pandemic, 2023, doi : doi.org/10.1038/s41586-022-05522-2 .

data.who.int/dashboards/covid19/cases?n=c, 2023, Accessed 23 December 2023.

Wu, Yi-Chi, Chen, Ching-Sung, Chan, Yu-Jiun, The outbreak of COVID-19: An overview, 2020, doi: 10.1097/JCMA.0000000000000270.

covid-19.go.id, 2021, access date: December 10, 2023

Mapping the Spread of COVID-19 Virus in India Using K-Means Clustering, 2021, doi: 10.1109/ICCED53389.2021.9664839

Corona.sumbarprov.go.id, 2021, Access Date 24 December 2023.

corona.padang.go.id/, Access Date 24 December 2023.

Nova Indriyani, Yessy Susanty Sabri, Afriani, Association Between Comorbidities and Outcome of COVID-19 Patients at Dr. M. Djamil General Hospital Padang, 2022, doi: https://doi.org/10.36497/respirsci.v3i1.59.

Maria Susan Anggreainy, Zaky Ahmad Hanif Raksakusumah, Tan Jonathan Budiman Tantoro, Said Achmad, Mikhael Theone Airis, The Relationship Between COVID-19 Cases and ArcGIS Technology to Present in Indonesia, 2022, doi: 10.1109/ICCIT55355.2022.10118972

https://satusehat.kemkes.go.id/platform/docs/id/terminology/icd/icd-10/, 2024, Accessed January 10, 2024.

I Made Arya Yogiswara Mahayasa, Moulid Hidayat, Prima Belia Fathana, Indana Eva Ajmala, Relationship of Comorbidity and Age to the Severity of Clinical Symptoms of COVID 19 at Mataram University Hospital, 2022, doi: http://dx.doi.org/10.29303/jbt.v23i1.4581

Rahmad Kurniawan, Siti Norul Huda Sheikh Abdullah, Fitra Lestari , Mohd Zakree Ahmad Nazri, Akhmad Mujahidin, Noridayu Adnan, Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries, 2020, doi: 10.1109/CITSM50537.2020.9268920

Ansari Saleh Ahmar, Darmawan Napitupulu, Robbi Rahim, Rahmat Hidayat, Yance Sonatha, Meri Azmi, Using K-Means Clustering to Cluster Provinces in Indonesia, 2018, doi: 10.1088/1742-6596/1028/1/012006

Try Iryanto Saputra, Rini Arianty, Implementation of K-Means Clustering Algorithm on Sentiment Analysis of Indosat User Complaints, 2019, doi: 10.35760/Ik.2019.V24i3.2361

Miftachur Robani, Achmad Widodo, K-Means Clustering Algorithm for Grouping Quran Verses in Indonesian Translation, 2016, doi: 10.21456/vol6iss2pp164-176

Mrs. Bharati M. Ramageri, Lecturer, Data Mining Techniques and Applications, 2020

2 , Data Mining as a Technique for Healthcare Approach, 2022, doi: 10.4236/ijcns.2022.159011

https://www.techtarget.com/searchbusinessanalytics/definition/data-mining, accessed December 28, 2023

Eric U. Oti, Michael O. Olusola, Francis C. Eze, Samuel U. Enogwe, Comprehensive Review of K-Means Clustering Algorithms, 2021 doi: 10.31695/IJASRE.2021.34050

Mohammad Annas, Siti Norida Wahab, Data Mining Methods, K-Means Clustering Algorithms, 2023, doi: https://doi.org/10.34306/ijcitsm.v3i1.122

EW Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classifications", Biometrics, 1965, vol. 21, 768-769

S. Lloyd, Least squares quantization in PCM. IEEE Transaction on Information Theory, 2013

academy.rapidminer.com, accessed January 4, 2024

Vincentius Riandaru Prasetyo, Hamzah Lazuardi, Aldo Adhi Mulyono, Christian Lau, Implementation of RapidMiner Application to Predict Rupiah Exchange Rate Against US Dollar Using Linear Regression Method, 2021, doi: https://doi.org/10.25077/TEKNOSI.v7i1.2021.8-17

Miro Jakovljević, Ljerka Ostojić, Comorbidity and Multimorbidity in Medicine Today: Challenges and Opportunities for Bringing Separated Branches of Medicine Closer To Each Other, 2013

Carol DerSarkissian, MD, what is comorbidity, November 17, 2021