Original article

Analysis of the Omicron virus cases using data mining methods in rapid miner applications

Johanes Fernandes Andry [1], Hendy Tannady[2], Glisina Dwinoor Rembulan [3], David Freggy Dinata [1]*

1-Department of Information Systems, Universitas Bunda Mulia, Jakarta, Indonesia
2-Department of Management, Universitas Multimedia Nusantara, Banten, Indonesia
3-Department of Industrial Engineering, Universitas Bunda Mulia, Jakarta, Indonesia

A B S T R A C T

Background: Omicron has respiratory problems and pneumonia in general and specific terms. This pandemic was ravaging all countries in the world. This virus outbreak had new types to appear or so-called new variants that are still being studied by experts. Computer-assisted methods (includes smart intelligence systems, algorithms, and data mining) is key solution for detecting variants of virus. **Methods:** In present study, it discussed and analyzed the omicron variant which is one of the variants of the Coronavirus 2019 (COVID-19). It's a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The emergence of this Omicron variant of COVID-19, raised more concern in the world because of its dangerous ability and the high level of spread of omicron cases. Analysis using the k-means algorithm in order to determine the level of distribution of the virus variant. **Result**: From the results and outputs found in this method, it is concluded that this method is used to divide the data into 3 clusters of case distribution of the Omicron variant which has been understood as a level in the distribution of cases where cluster 0 is low level, cluster 1 is high level, and cluster 2 is medium level. **Conclusion**: Therefore, this data mining method with special clustering and data-mining techniques give the highest number of virus distributions in which countries and divide some countries into several clusters.

Introduction

In 2019, the world was shocked by the emergence of a disease outbreak called by experts as Corona virus disease 2019 or abbreviated as COVID-19. Corona virus disease 2019 was a new virus that can cause respiratory problems and pneumonia due to viral infarction for people exposed to this virus. The respiratory disorder is assumed to have come from bats and then was transferred to another organisms [1, 2], and then to humans [2,3]. Once transmitted to humans, the respiratory disorder virus began spreading quickly among individuals [3]. Symptoms that prove a person has been exposed to this virus also vary, ranging from mild symptoms such as the common cold to severe symptoms such as pneumonia or sepsis [4]. After many conducted studies on COVID (COVID-19), the puzzle of this virus is not completed yet.

It can be seen that the virus spreads by means of close contact with sufferers and the entry of droplets into the respiratory tract [2]. Droplets are small particles from the patient's mouth that contain the disease virus when coughing, sneezing or talking [4]. These droplets can stick to clothing or nearby objects but will not last long. Therefore, people are required to use masks or other protective equipment that can cover the nose and mouth to prevent droplets from entering our respiratory tract and contain the spread of the virus [4,5].

Corona virus disease 2019 is becoming a global problem in some regions of the world as the number of cases continues to rise [6]. This virus is growing and mutating day by day, creating new types of variants that are still being researched by experts. Several Variants of Concern (VOC) have been defined by the World Health Organization (WHO). One of them is Omicron (B.1.1.529) which is the newest VOC [7]. Omicron has many mutations that have the potential to increase transmission to each other, and some can escape infection or vaccine-induced immunity [8].

The Omicron variant was first detected in specimens collected on November 11, 2021, in Botswana and on November 14 in South Africa [9]. This variant is causing concern worldwide because it's able to evade the immunity generated by the vaccine, where in a meta-analysis a study showed that the vaccine's effectiveness was reduced to 40% against symptoms and 80% against severe disease [10]. The most common symptoms in Omicron sufferers are cough, fatigue and nasal congestion or runny nose [8]. This makes it difficult for people to move outside the home for fear of contracting the omicron virus. As a result of the large number of people who are unable to carry out their activities due to the relentless spread of the virus, it has created an economic crisis [11,12] due to reduced people's income and reduced food supply [13] which also affects the global supply chain [14].

Various computer-assisted techniques have been conducted for detection and treatment of COVID cases; an automatic method for detecting COVID-19 from chest x-ray images based on deep learning networks is presented by **Sabahi et al.** [15]. For the deep learning network proposed in this work, the combination of convolutional networks with type 2 fuzzy activation functions is used in order to deal better with noise. Also, to increase the data, adversarial networks have been used in this

research. The final accuracy obtained for the classification of the first scenario (health and COVID-19) and the second scenario (health, pneumonia and COVID-19) is about 99 and 95%, respectively.

On the other study [16], 9 key features and three top data mining techniques were identified. The results of the experiments show that the prediction model provided using the identified key features and the best data mining technique obtained provides an accuracy of 83.19% for the diagnosis of corona disease. In a review by **kazemi et al.** [17] there is summarized studies examining the description of Artificial intelligence (AI) summarized coronavirus and symptoms and how to transmit, the advantages of using AI in the crisis of the coronavirus pandemic and contributing to rapid diagnostics, monitoring of treatment stages, Epidemiology and rejection of infected areas, management of the pharmaceutical system, reducing the workload of medical personnel, AI-based definitive diagnostic methods of COVID-19 and laboratory programs with an emphasis on AI Is. Artificial intelligence can be used in rapid diagnosis, monitoring of people with COVID-19, treatment Development, Research facilitation, prevention, and drug and vaccine design. AI-based software uses smart grids designed to mimic human thought processes. AI can recognize a pattern beyond defined rules and analyze a significant amount of information from what humans can manage.AI-based systems make it possible to identify suspected cases of COVID-19 with minimal contact of therapeutic personnel with patients, to monitor the continuation of the therapeutic process.

Based on this background, this study uses data mining clustering techniques with the K-Means method to perform clusterization. So that what happened in this study was to divide several clusters of the spread of the omicron virus so that decisions could be made on those clusters where the lowest cluster could reduce their concerns about the spread of the omicron virus and in the highest cluster, they could implement strategies to prevent the spread of the omicron virus.

Literature review

Knowledge discovery in database (KDD)

Knowledge discovery in database (KDD) is the process of finding knowledge, similarities, patterns or trends from previously unknown implicit information [18,19]. Data mining is a process or

steps that are important in finding knowledge, similarities, patterns or trends in useful databases. In KDD and data mining have different requirements, namely in KDD refers to the whole process to find useful knowledge from data while data mining focuses on finding new patterns from the amount of data with suitable and useful algorithms [20]. Determining is increasingly spreading and growing rapidly in recent times due to its ability to mine useful patterns and trends from existing databases. There are still many companies that have large data sets but don't know what to do, even though it is possible to extract information from the data set by extracting valuable information such as valuable information in the data repository [21]. With the help of data mining to look for patterns in this large data set, companies can use it to study the company's data more deeply and make decisions from the results of data mining to develop the company. Therefore, data mining has a big role in making a company's decision.

Knowledge is not represented explicitly, therefore finding useful information from complex and large volumes of data is developed in database technology that can process data [22-24]. The stages in the data mining process are done by selecting data from the source and then improving the quality of the data, after that transforming it into the desired form then data mining and evaluation are carried out [25].

Based on **figure (1)**, in KDD there are several processes that will be passed, there are [26].

a) Data selection

In the data mining process, the first step that needs to be done is data selection. Data selection is the stage for selecting and selecting data from a set of data.

b) Pre-processing/cleaning

Before the data is processed, it is necessary to carry out a cleaning process to remove duplication of data, repair or eliminate inconsistent and incomplete data.

c) Transformation

Data transformation is the process of converting or combining data into a predetermined format.

d) Data mining

Data mining is a process to find an unknown and interesting pattern, similarity or trend from a selected data set using the selected method or technique. The selection of the right method or technique greatly influences the desired results in data mining.

e) Interpretation/evaluation

The results of the data mining process in the form of information patterns and so on need to be displayed in a form that is easily understood by interested parties. This stage aims to examine the results of data mining whether the patterns or information found contradict the facts or pre-existing hypotheses.

The objectives in KDD can be divided into 2, namely verification and discovery [27]:

1) Verification, the goal is to verify the results of data mining with the hypothesis.

2) Discovery, the goal is that the system can find new patterns that are useful for the user.

These goals can be achieved by several methods in data mining, including [27]:

a. Classification: this method serves to map (classify) data items into one of the pre-defined groups.

b. Regression: this method serves to map data items to real-valued predictive variables and to find functional relationships between variables.

c. Clustering: clustering is a probability density estimation method consisting of techniques for estimating from the data the combined multi-variant probability density function of all variables/fields in the database.

d. Summarization: to find a concise description for a subset of data.

e. Dependency modeling: find a model that describes a significant dependence between variables.

f. Change and deviation detection: find the most significant changes in data from previous data or normative values.

Clustering

Clustering is one of the existing methods in data mining and is often used in implementing data mining in business intelligence activities. Clustering is a data mining technique that is used to group data based on characteristics into several groups so that data that has a level of similarity or similarity that is close to other data will be in a group and data that has a level of similarity or similarity that is far or low will be in a group. in different groups [19].

The clustering method is often considered the most important method in unsupervised learning. Any similar problem has to do with finding structures in unlabeled data sets. In clustering there are several terms that are often used, namely cluster analysis,

automatic classification, numerical taxonomy, botrological and typological analysis [28]. There are 2 types of similarity from clustering, namely intraclass similarity and interclass dissimilarity. Intraclass similarity is similar to objects in the same cluster while interclass dissimilarity is objects that are not similar to objects in other clusters [29].

a. Partitioning method is a method that produces a set of clusters with each object in one cluster.

b. Hierarchical clustering or also known as connectivity-based clustering is a method that connects objects to form clusters based on distance.

c. Density-based clustering is a method that defines a cluster as an area with a higher density than the remaining data set. Objects in this sparse area required to separate clusters are usually considered to be the limit of density affordability.

d. Grid-based clustering is a method that quantizes the data space into several multi-dimensional cells which then perform the desired operation on the object. Usually used for density-based clustering which is done by separating data into groups according to the structure [30].

e. Model-based clustering is a method that refers to the use of a mixed model that focuses on clustering, each observation emerging from a finite mixture [31,32].

K-Means

K-means is one of the techniques or algorithms that can be used in the clustering method in implementing data mining, K-means is often used because it is able to group data properly which separates data into different groups [16,18,27]. The K-Means algorithm is known as one of the oldest and most popular partitioning methods [33]. Data mining clustering method will produce optimal results if the K-means algorithm is combined with the hierarchical clustering algorithm. K-means has a weakness caused by determining the initial center of the [21].

The use of the K-means algorithm in the clustering process depends on the data obtained and the conclusions to be reached at the end of the process. So that in the use of the K-means algorithm there are the following rules [31]:

1) The desired number of clusters.

2) Attributes must be of numeric type.

There are several problems that need to be considered in using the K-means method including

different clustering models, selecting the most appropriate model for the analyzed dataset, failure to converge, outliers' detection, the shape of each cluster and overlapping problems [22].

As for the process of the K-Means algorithm, as follows [27]:

1) The dataset is divided into clusters and data points are randomly selected into clusters resulting in clusters with approximately the same number of data points.

2) For each data point:

a. Calculate the distance from the data point to each cluster.

b. If the data point approaches the cluster where it is located, it will remain and does not change but if the data point does not approach the cluster in its place, it will be moved to the nearest cluster.

3) Keep repeating step two until all data points do not move from one cluster to another. In this process the cluster point is stable, and the process ends.

4) The choice of partition at the beginning greatly affects the final cluster results, namely in terms of distance and cohesion between clusters and intraclusters.

The K-Means algorithm has the following properties [27]:

- Always have K Cluster

- Always have at least one item per cluster.

- existing clusters do not overlap and are not hierarchical.

- Each member of the cluster must be closer to the cluster that belongs to other clusters because proximity does not always involve the "center" of the cluster.

Figure 1. Stage of KDD [19].
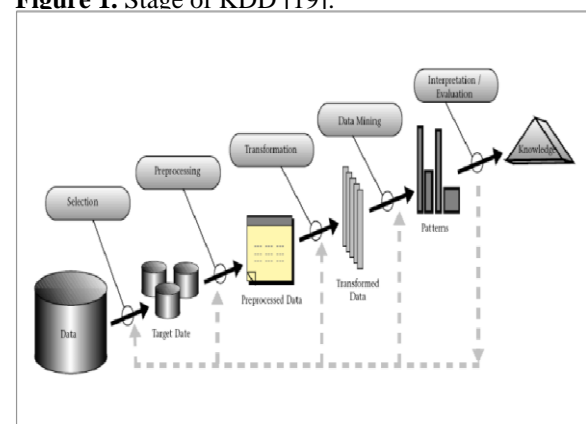
Methodology

This methodology is used to provide an understanding of the processes that occur in this paper. Based on **figure (2),** it is known that in this research there are stages of research that are passed:

1) Identify problems, this stage identifies the problems obtained which is where the problem we get is omicron

2) Analyze the problem, this stage analyzes the problem, from the problems obtained, it is better to analyze the data in what form it is.

3) Set goals, this stage determines the goals to be achieved, where the aim of this paper is to obtain the results of an analysis of the distribution of the omicron virus

4) Study literature, this stage conducts a study of the literature. Literature review is very important to know the application of the method to solve the problem [33].

5) Collecting data, at this stage, data collection is carried out where data is taken from kaggle.com

6) Selecting data, from the data that has been collected, the appropriate attributes will be selected.

7) Processing data with the K-Means Clustering algorithm, we perform data mining using the clustering method with the K-Means algorithm. Clustering with the K-Means algorithm is a very popular algorithm used because it is able to work on various types of data, can be used on images, text and numbers [30].

8) Analyze result, from the processing results will be analyzed and given an understanding of these results.

Figure 2. Research stages [16].
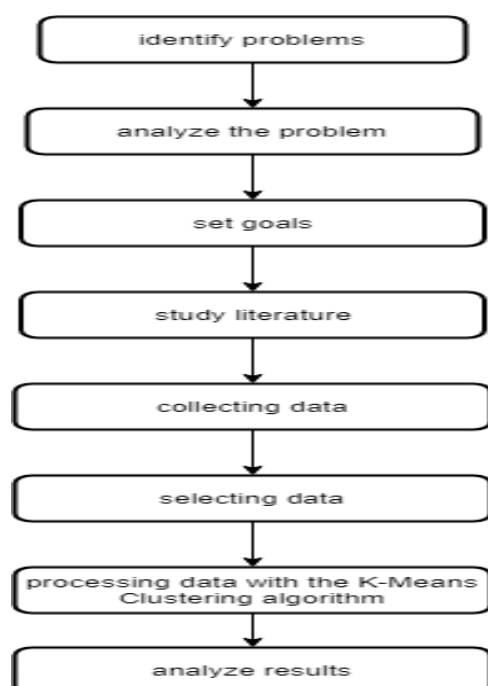
Results and discussion

In this section it will contain the contents of the selected data such as attributes and descriptions followed by explanations and also contains the implementation of the data mining process using the rapid miner application and provides the results.

The dataset taken in this paper is titled "Omicron daily cases by country (COVID-19 variant)" which was taken on the kaggle.com. From the dataset taken with the title "Omicron daily cases by country (COVID-19 variant)" it contains 6 columns and 100417 rows of data at the time this report was made. The dataset has attributes or features that describe each instance. The following is a list of attributes in this study and their explanation **(Table 1).**

Implementation of the dataset in rapidminer

The dataset obtained from the Kaggle website with the title Omicron daily cases by country (COVID-19 variant) will be used for research. Data mining with the clustering method and using the K-means algorithm in the rapid miner application. For the implementation of the dataset using the K-Means algorithm in the rapid miner, it is necessary to import data, which will be used as testing data and training data into the repository selected in the rapid miner.

In importing data, it is necessary to pay attention to the existing attributes and provide data types that match the attributes they have because otherwise they will produce inappropriate results. Due to the research using the K-means algorithm using the K-means operator where the operator can only accept numeric data types when processed, the data that will be used is only on num_sequences to be divided into several clusters.

Before entering the data mining process, an understanding of the omicron daily cases by country (COVID-19) data was carried out. This understanding is done so that the author can understand the contents of the dataset and can determine the path or direction of the implementation of data mining. So that the purpose of using data mining with the clustering method of the dataset used is to find out the division of groups in the dataset used. The division of groups in this report serves to identify areas with high, low and medium numbers.

To find out information about data and statistical values from the dataset used, you can put an operator

containing data in the process section, then connect the operator to the result point to produce the desired result like for it can be done by pressing 2 times the data in the repository. Then the results will be the same as **figure (4)**.

All patients with at least one symptom were classified into the symptomatic group (776 patients). In relation to the symptoms the sensitivity and specificity of PCR test in the symptomatic group were 88.7% and 13.9% respectively with increased sensitivity up to 97.7% when both symptoms and rapid Ag test were added together while specificity was 13.7% with a positive predictive value of 67.1% and a negative predictive value of 76.8%.

From this information, it can be obtained the names of the existing attributes, the data type of the attribute, the amount of missing data, the graph, the least value, most value, the average and deviation value for the numeric data type, as well as the existing data values in the polynomial data type. So, from the information provided by this rapid miner, the author can use it to carry out the data mining process.

Based on **table (2),** There are several types of data that exist in the dataset that is used in data training and data testing.

Process implementation using rapidminer

First of all, in implementing rapid miner, a repository is made which is a place to store processes and data sets that will be used. After that, enter the dataset that was taken into the repository that was created. Find the retrieve operator then drag the operator to the design process section if you have selected the data used in the data repository. After that enter the remove duplicates operator and filter examples into the design process section.

Connect retrieve with the duplicate's operator after that the duplicates operator is concatenated with filter examples. in the process of removing duplicates this is part of the pre-processing/cleaning where remove duplicates serves to remove or eliminate duplication of data so that when processed there is no data that has the same value.

After the process of removing duplicates is complete, the results will be sent to the filter examples operator for filtering the data. Because in this case clustering is done to find out the high, medium, low group of numbers in the case of the omicron variant only, the filter examples aim to filter from the data sent by the remove duplicates operator to only the omicron variant case.

The parameters that need to be filled in so that the desired results are as desired. In the filter examples operator parameter press add filters to perform a filter and select the attribute you want to filter after that select the condition and enter the value of the condition.

In this case, the desired attribute is a variant with the condition is in, which is a condition where the attribute contains the same content as the selected value, then the value option contains omicron as in **figure (6).**

The results of this filter process will be sent to the attribute select operator to be transformed into the data form. The form of the data is transformed because the K-means algorithm can only accept numeric values for processing.

The parameters that must be filled in to obtain results can be further processed by the K-means algorithm, namely the filter type attribute is filled with single and selects the attribute you want to retrieve, similar with method of **Tjhai et al.** [34]. Because in this case num_sequence, we will choose num_sequences as the selected attribute.

The results of the select attribute will be sent to the clustering for processing with the K-means algorithm. There are various operators provided for the clustering process. However, in this case, the K-means algorithm is chosen.

There are also parameters that must be filled in to obtain the desired results, as shown in **figure (10)**.

In the clustering operator parameter (K-means), things that need to be considered are checking the add cluster attribute, inputting the number of k as the number of groups to be divided, and max run as the number of experiments performed, similar with method of **Mohassel et al.** [35] . After that connect the output of the clustering operator to result and select the output attribute ori to result to display the original data before transforming. If it is correct then the arrangement will be seen in **figure (11)**.

From **figure (12),** it is known that there is a table cluster display that has been predicted from the number of cases (num_sequences). There is also a display of information on other tabs as shown in **figure (13).**

From **figure (13)**, it is known that there are 3 clusters that have been divided, namely:

a. Cluster 0 (Low level) has a value of 4181 data that is in cluster 0. It is called low level because after viewing the contents of the table cluster 0 has a range of the number of omicron cases

a. starting from 0 and less than 5,000

b. Cluster 1 (High Level) has a value of 1 data that is in cluster 1. It is called high level because after looking at the table, cluster 1 has the highest number of omicron cases, which is 52,456 cases. A total of 52,456 cases of Omicron took place in the United Kingdom on 27 January 2021.

c. Cluster 2 (medium level) has a value of 2 data that are in cluster 2. It is called medium level because after looking at the table, cluster 2 has a number of omicron cases that are less than 25,000 and more than 5,000, namely the number of cases 11956 and 24681. In the case of 11956 Omicron cases were located in the United Kingdom location on January 13, 2021 and 24681 Omicron cases were located at the United States location on 27 January 2021.

d. From the results of the clustering, a decision can be made that in countries affected by cluster 1 or affected by Omicron at most, a lockdown can be carried out and people who are still in that country are prohibited from traveling until there is news that the distribution rate has decreased and it is necessary to test before traveling, and for cluster 2, alert status can be exercised and reduce activity in the crowd. In cluster 0, there is no need to lock down and limit activities, but the community is still required to comply with health protocols.

Table 1. Attributes and description.

No	Attribute	Description
1.	Location	(Country of patient) nominal: -Angola -Argentina -Australia and more
2.	Date	(The date the report was submitted) date: -Fri Jan 29, 2021 -Wed Jan 13, 2021 - and more
3.	Variant	(a proven variant) nominal: -Omicron -Alpha -Beta and more
4.	Num_sequences	(Number of patients infected with the virus) numeric: -0 "no one is exposed to the virus" -1 "someone has a virus" - The more numbers mean the more patients exposed to the virus
5.	Perc_sequences	(Perc_sequences is the percentage number of cases affected by a variant in a time period (date) and in a location) Real: -the higher the number of people affected by the variant within the specified time
6.	Num_sequences_ total	(Num_sequences_total is the total of num_sequences or the total number of cases affected by various variants of the virus in one time period (date) and in a location)

Table 2. Type data.

Instance	Type data
Location, Variant	Nominal
Date	Date
Num_sequences, Num_sequences_total	Numeric
Perc_sequences	Real

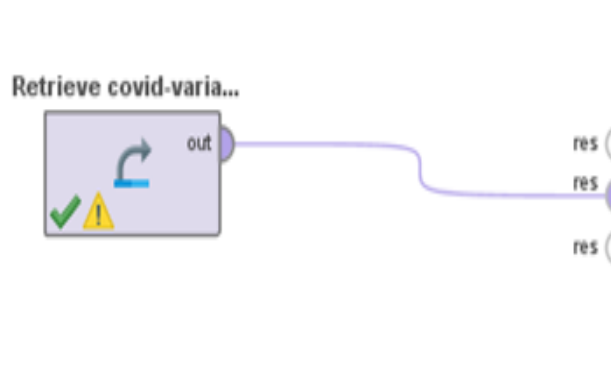**Figure 3.** Display of Retrieve to resultpoint
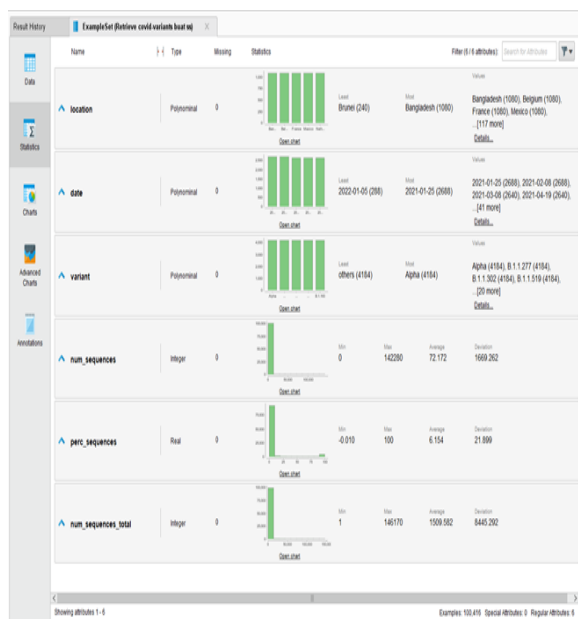
Figure 4. ExampleSet.

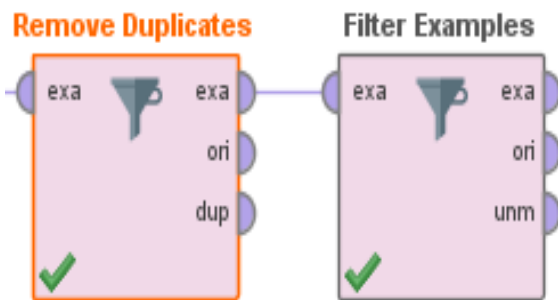**Figure 5.** Display of remove duplicates.

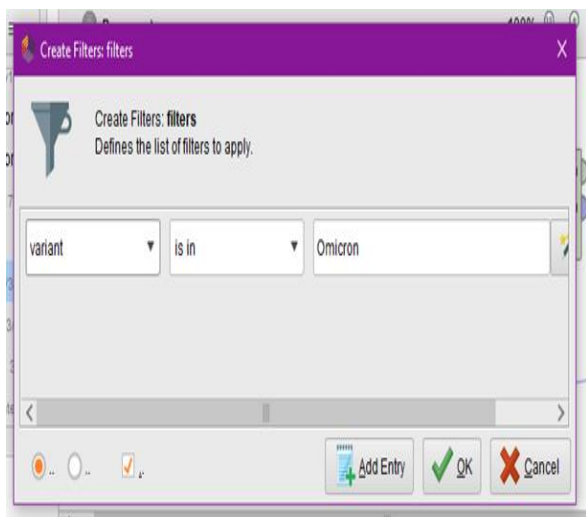**Figure 6.** Filter parameter display

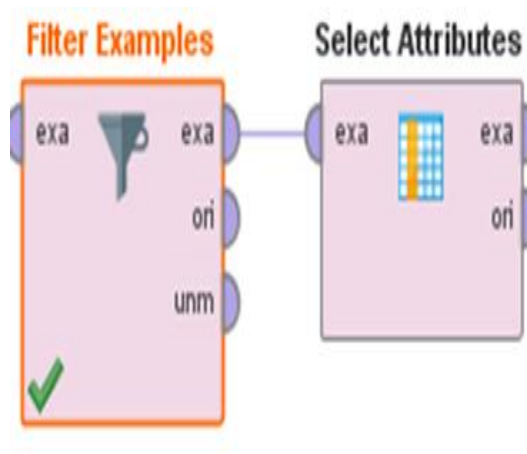

Figure 7. Display of select attributes operator.

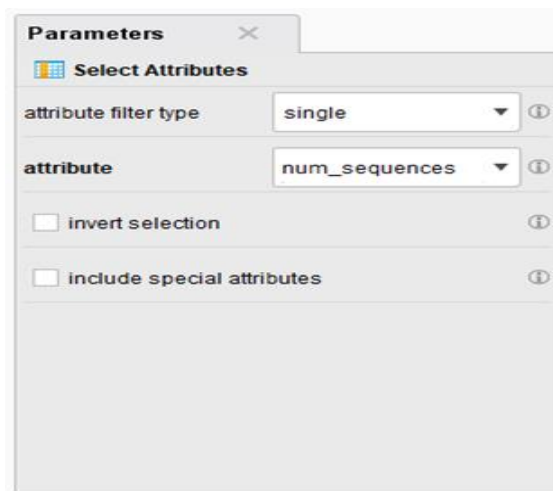**Figure 8.** Display of attribute select operator parameter.
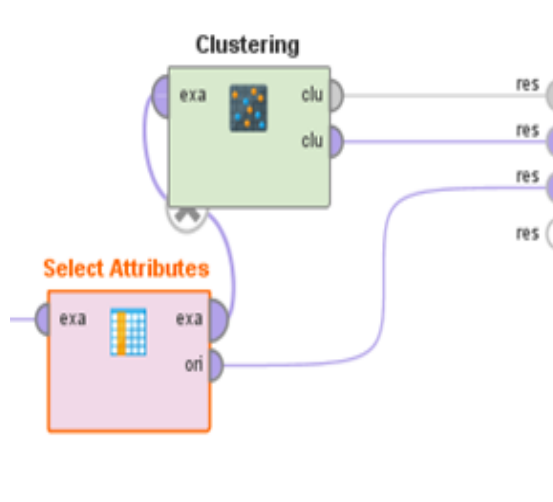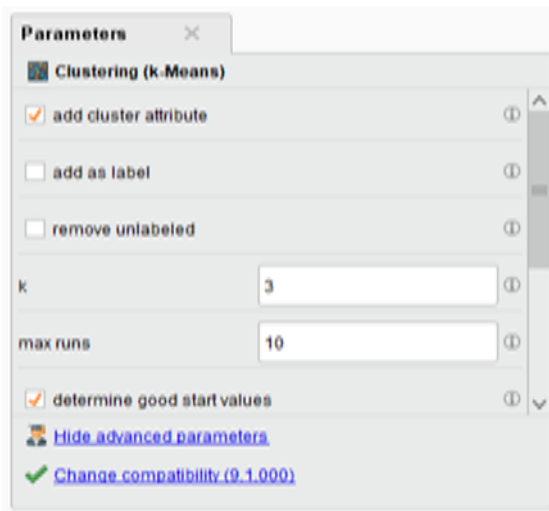

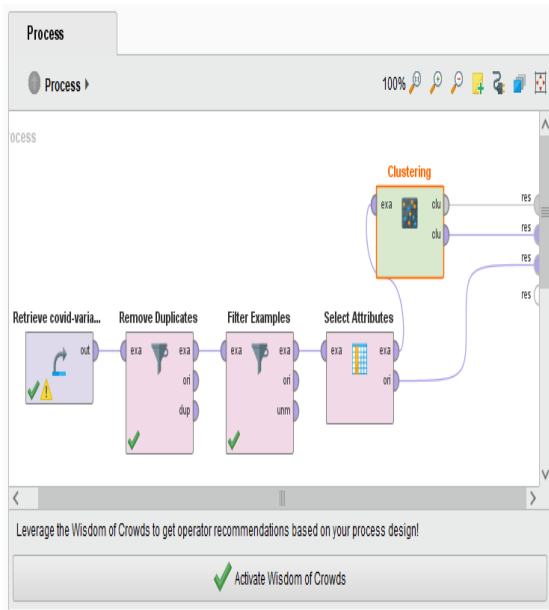**Figure 9.** Display of clustering process.

Figure 10. Display of clustering parameter with K-Means.
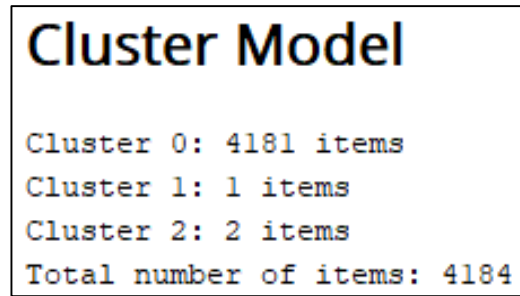

Figure 11. Display of data mining progress with K-Means Clustering

Figure 12. Display of clustering with K-Means.

Cluster Model

Cluster 0: 4181 items
Cluster 1: 1 items
Cluster 2: 2 items
Total number of items: 4184

Conclusion

It can be conducted that an analysis of the Omicron virus case using a data mining clustering method with the K-Means algorithm using a rapid miner application with a dataset taken from the Kaggle website with the title "Omicron daily cases by country (COVID-19 variant). From the results received in this method, it is summarized that this method is used to divide the data into 3 clusters of case distribution of the Omicron variant which has been understood as a level in the distribution of cases where cluster 0 is low level, cluster 1 is high level, and cluster 2 is medium level. Currently, in high-level cases there are 52,456 cases which are located in the United Kingdom on January 27, 2021, and in moderate cases there are 2 cases with a total of 11,956 originating from the United Kingdom on January 13, 2021, and 24,681 cases originating from the United States on January 27, 2021. Then the decision was taken to lock down in cluster 1, in cluster 2 there were restrictions on activities outside the home and in cluster 3 they were given supervision to comply with health protocols. There are still many limitations in our research, such as the lack of good datasets that we take for analysis using other methods such as forecasting and classification where the accuracy level is very low and the predictions from the analysis results are still inaccurate. So that the clustering method with the K-Means algorithm was chosen because it can provide good results and is easy to understand for researchers and readers.

Recommendations

Because the data distance of each cluster is very far, the level of cluster division is not very good, so it can be seen in the data that cluster 1 and cluster 2 have minimal members and are not evenly distributed. It is hoped that in further research, researchers can continue this research by finding better datasets and being able to properly process these datasets so as to provide good results for

readers. Also, so that in the future research can be carried out with other methods such as classification and forecasting with a good level of accuracy of results, so that they can be analyzed according to the objectives in order to provide good and appropriate decision results for decision makers. We hope that in the future we can do better research and can provide useful benefits for the world.

References

1- **Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al.** Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet 2020;395(10223):497-506.

2- **Alinia-Ahandani E, Sheydaei M.** Overview of the introduction to the new coronavirus (COVID19): A Review. J Med Biol Sci Res 2020;6(2):14-20.

3- **Aydin E, Demir PA.** The effect of the pandemic on the consumption of animal products: the case of Kafkas university of Turkey. Online J Anim Feed Res 2022; 12(1):37-45.

4- **Wiguna H, Nugraha Y, Andika A, Kanggrawan JI, Suherman AL**. Kebijakan Berbasis Data: Analisis dan Prediksi Penyebaran COVID-19 di Jakarta dengan Metode Autoregressive Integrated Moving Average (ARIMA). Jurnal Sistem Cerdas 2020;3(2):74-83.

5- **Chen LD.** Effects of ambient temperature and humidity on droplet lifetime–A perspective of exhalation sneeze droplets with COVID-19 virus transmission. International Journal of Hygiene and Environmental Health 2020 229:113568.

6- **Kar SK, Arafat SY, Sharma P, Dixit A, Marthoenis M, Kabir R.** COVID-19 pandemic and addiction: Current problems and future concerns. Asian journal of psychiatry 2020;51:102064.

7- **Berkhout B, Herrera-Carrillo E.** SARS-CoV-2 Evolution: On the Sudden Appearance of the Omicron Variant. Journal of Virology 2022;96(7):e00090-22.

8- **Cdc COVID, R. T.** Morbidity And Mortality Weekly Report Sars-Cov-2 B.1.1.529 (Omicron) Variant-United States 2021.

9- **Jansen L, Tegomoh B, Lange K, Showalter K, Figliomeni J, Abdalhamid B, et al**. Investigation of a SARS-CoV-2 B. 1.1. 529 (omicron) variant cluster—Nebraska, November–December 2021. Morbidity and Mortality Weekly Report 2021;70(51-52):1782.

10- **Lyngse FP, Mortensen LH, Denwood MJ, Christiansen LE, Møller CH, Skov RL, et al.** SARS-CoV-2 Omicron VOC transmission in Danish households. Medrxiv 2021.

11- **Mofijur M, Fattah IR, Alam MA, Islam AS, Ong HC, Rahman SA, et al**. Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic. Sustainable production and consumption 2021;26:343-59.

12- **Zenker S, Kock F.** The coronavirus pandemic–A critical discussion of a tourism research agenda. Tourism management 2020;81:104164.

13- **Workie E, Mackolil J, Nyika J, Ramadas S.** Deciphering the impact of COVID-19 pandemic on food security, agriculture, and livelihoods: A review of the evidence from developing countries. Current Research in

Environmental Sustainability 2020;2:100014.

14-Guan D, Wang D, Hallegatte S, Davis SJ, Huo J, Li S, et al. Global supply-chain effects of COVID-19 control measures. Nature human behavior 2020;4(6):577-587.

15-Sabahi, K., Sheykhivand, S., Mousavi, Z., & Rajabioun, M. Recognition COVID-19 cases using deep type-2 fuzzy neural networks based on chest X-ray image. Computational Intelligence in Electrical Engineering *2022; In press.*

16-Arabzadeh S A A, Jamshidi V, Saeed M, Yazdani R, Jamshidi M. Automated detection of coronavirus disease (COVID-19) by using data-mining techniques: a brief report. Tehran University Medical Journal 2022; 79 (10):822-830.

17-kazemi E, mehrabi N, hajipurtalebi A. An Overview of the Artificial Intelligence Applications in Identifying and Combating the COVID-19 Pandemic. Paramedical Sciences and Military Health 2022; 16 (3) :52-65.

18-Ayobami AS, Rabi'u S. Knowledge Discovery in Database: A knowledge management strategic approach. In Proceedings of 6th Knowledge Management International Conference (KMICe), Johor Bahru, Malaysia 2012.

19-Fatmawati K, Windarto AP. Data Mining: Penerapan rapidminer dengan K-means cluster pada daerah terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. CESS (Journal of Computer Engineering, System and Science). 2018;3(2):173-8.

20-Sharma N, Saxena A. A Survey Of Data Mining And Knowledge Discovery Process Model And Its Applications In Database. IJRAR-International Journal of Research And Analytical Reviews 2018;6 (1):21Z-33Z.

21-Madyatmadja ED, Marvell M, Andry JF, Tannady H, Chakir A. Implementation of Big Data in Hospital using Cluster Analytics. In 2021 International Conference on Information Management and Technology (ICIMTech) 2021;1: 496-500. IEEE Proceeding.

22-Seerat B, Azam F. Opinion mining: Issues and challenges (a survey). International Journal of Computer Applications 2012 49(9):42-51.

23-Sitorus T, Tannady H. Synergy, System IT, Risk Management and The Influence on Cyber Terrorism and Hoax News Action. Journal of Theoretical and Applied Information Technology,2021; 99(8), 1802-1814.

24-Gunawan FE, Andry JF, Tannady H, Meylovsky R. Designing enterprise architecture using togaf framework in meteorological, climatological, and geophysical agency. Journal of Theoretical and Applied Information Technology 2019;97(20):2376-85.

25-Soni Madhulatha T. An Overview On Clustering Methods 2012; 2(4), 719–725.

26-Bijuraj LV. Clustering and its Applications. InProceedings of National Conference on New Horizons in IT-NCNHIT 2013; 169:172).

27-Bungkomkhun P, Auwatanamongkol S. Grid-based Supervised Clustering Algorithm using Greedy and Gradient Descent Methods to Build Clusters. International Journal of Computer Science Issues (IJCSI) 2012;9(3):73.

28-Fop M, Murphy TB. Variable selection methods for model-based clustering. Statistics Surveys 2018;12:18-65.

29-Mcnicholas PD. Model-Based Clustering. Journal of Classification 2016; 33: 331–373.

30-Andry JF, Silaen FM, Tannady H, Saputra KH. Electronic health record to predict a heart attack used data mining with Naïve Bayes method. Int J Inf & Commun Technol 2252(8776):8776.

31-Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. IEEE Access 2020; 8:80716-27.

32-Liliana DY, Maulana H, Setiawan A. Data Mining untuk Prediksi Status Pasien COVID-19 dengan Pengklasifikasi Naïve Bayes. Multinetics 2021;7(1):48-53.

33-Hossain MZ, Akhtar MN, Ahmad RB, Rahman M. A dynamic K-means clustering for data mining. Indonesian J Elec Eng Comp Sci 2019;13(2):521-6.

34-Tjhai GC, Furnell SM, Papadaki M, Clarke NL. A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm. Computers & Security 2010;29(6):712-23.

35- Mohassel P, Rosulek M, Trieu N. Practical privacy-preserving k-means clustering. Cryptology ePrint Archive 2019.