Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
www.bircu-journal.com/index.php/birci
Budapest Institute

The Application of Prediction Data Miningbed Occupancy Rate of Covid-19 Patients in West Java

Amelia Hani[1], Ayi Ratna K[2], Cristina Juliane[3]
[1,2,3]*Master's Program, Business Information Systems, STMIK LIKMI, Bandung, Indonesia*
ameliahoneyhani@gmail.com, arkristi38@gmail.com, chistina.juliane@likmi.ac.id

Abstract

The Covid-19 pandemic in 2020 is a complex health problem and requires fast handling. Covid-19 patients who receive treatment in hospitals have different conditions and severity. This affects the handling actions that will be carried out by medical officers. The large number of patients and the lack of medical personnel and the availability of beds have resulted in the need for technological support to help predict the status of patient bed availability based on their condition so that treatment is concentrated on patients who are very critical and require rapid treatment. This research applies prediction techniques from data mining disciplines to predict a spike or decrease in BOR (Bed Occupancy Rate). Prediction using the C4 algorithm. 5 was applied to build a model based on the Covid-19 bed availability dataset. The Covid-19 BOR (Bed Occupancy Rate) dataset in West Java was obtained from Opendata.jabarprov.go.id and applied using Rapid Miner. The model built can predict the status of bed availability based on the occupancy of the patient's hospitalization. The results of this study indicate that predictions using the C4.5 Algorithm method have a high level of accuracy of %.

Keywords
Covid-19; prediction; BOR; C4.5 algorithm

I. Introduction

Corona Virus Disease 2019 or Covid-19 for short is a new disease that emerged in 2019 and can cause pneumonia and respiratory problems. Most people infected with the Covid-19 virus will experience mild to moderate respiratory illness and recover without needing treatment (C. Long, 2020), This disease is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and can cause death, people who have problems medical conditions such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illnesses. This disease has been a global health crisis pandemic since March 2020 (S. Sanche, 2020). This figure continues to creep up. Considering the Covid-19 outbreak is a global problem in parts of the world, including in West Java. The number of patients in the hospital and the capacity of medical beds are the main problems faced in various regions (M. Sukmana, 2020). Patients with high urgency require treatment priority over patients with moderate or asymptomatic symptoms (M. Abed, 2020).

Data mining is a technique used to build machine learning models. Machine learning (machine learning) is a modern artificial intelligence technique that learns to build models using empirical data (T. Miler, 2019). Data mining is used to find patterns in large sets of raw

data. Data Mining applies Machine Learning techniques to draw knowledge on data. In this study, the authors apply data mining techniques to classify Covid-19 datasets using the C4.5 Algorithm because the C4.5 Algorithm has been successfully applied in many classification tasks or conditional probability-based predictions on data populations (A. Fattah, 2019).

This study aims to provide a solution to automatically predict the status of Covid-19 room availability. Sihombing (2020) state that Covid-19 pandemic caused everyone to behave beyond normal limits as usual. The outbreak of this virus has an impact especially on the economy of a nation and Globally (Ningrum, 2020). The problems posed by the Covid-19 pandemic which have become a global problem have the potential to trigger a new social order or reconstruction (Bara, 2021). One of the techniques from Data Mining that can be used to predict data on the availability of beds for COVID-19 patients. The Regional Government and their staffs also established regional regulations addressing the problem of Covid-19 in their respective regions as an exertion to avoid the feast of Covid-19 (Rahmaniah, 2021). The Government of the Republic of Indonesia was formed to protect the whole of the Indonesian people (Angelia, 2020). To that extend, one of the important elements to consider is Human Resources (HR). According to Simamora who stated that HR management is a process of utilizing raw materials and human resources to achieve the goals set (Simamora, 2011). The C4.5 algorithm is an algorithm used to build a decision tree (decision making). The main benefit of using a decision tree is its ability to break down complex decision-making processes into simpler ones so that decision-makers will better interpret solutions to problems (Elmande, 2012). The C.45 algorithm is one of the decision tree induction algorithms, namely ID3 (Iterative Dichotomiser 3). ID3 was developed by J. Ross Quinlan. In the ID3 algorithm procedure, the input is a training sample, training label and attributes. C4 algorithm. 5 is the development of ID3. Some of the developments carried out in C4.5 are, among others, being able to overcome missing values, being able to overcome continue data, and pruning (Faradullah, 2013). So that it is hoped that it can help predict the status of the availability of beds for Covid-19 patients to get the right treatment directly by the medical team.

II. Research Methods

Data mining is a process of gathering important information from big data. The collection of important information is carried out through several processes which include statistical methods, mathematics and artificial intelligence technology. More specifically, data mining is defined as a tool and application that uses statistical analysis of data and filters and stores as much of that data as possible. Data mining has various functions, including the main functions, namely descriptive and predictive. For more details about this function, the following will be given an explanation.

Prediction is generally considered as an action that explains about the future. This is different from guessing simply by considering experience, opinions, and other information in making forecasts. The term commonly associated with 'prediction' is 'forecasting'. Although many people believe that the two terms are synonymous, there is a subtle but very important difference between the two. 'Prediction' is generally opinion and experience based, 'forecasting' is based on data and models. That is, in order of reliability, people will sort the terms like this: 'guessing', 'predicting', and 'forecasting'. In data mining terminology, 'prediction' and 'forecasting' are used synonymously, and the term prediction is used as a general representation.

This research methodology is carried out systematically which can be used as a guide for researchers in carrying out research so that the results achieved do not deviate and the

desired goals can be carried out properly and in accordance with the goals that have been previously set as shown in Figure 1.

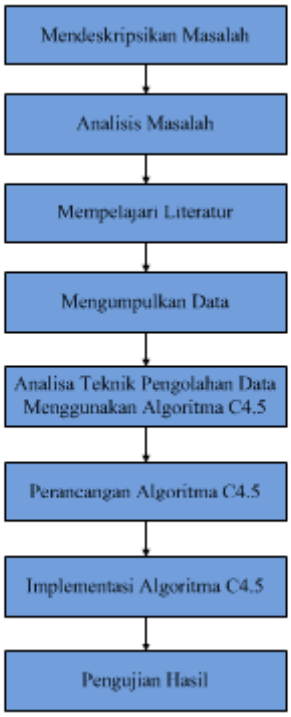

Figure 1. *Research Framework*

Based on the framework in Figure 1, each step can be described as follows:

a. Describing the Problem

Describing the problem to be researched needs to be determined first. Describe the problem in research by determining and defining the boundaries of the problem to be studied, so that it helps in getting the best solution to the problem. So, this first step is the most important initial step in this research.

b. Problem Analysis

The problem analysis step is a step to be able to understand the problem that has been determined by its scope or limitations. By analyzing the problem that has been determined, it is hoped that the problem can be understood properly.

c. Studying Literature

In order to achieve this goal, several literatures that are expected to be used are studied. Then the literature studied is selected to determine which literature will be used in research

d. Collecting data

The dataset of Bed Availability (BOR) in West Java was obtained from the West Java Open Data website (opendata.jabarprov.go.id). This dataset will be trained to produce predictive models for the availability of beds for COVID-19 patients. Datasets are prepared and cleaned where only the attributes are relevant. The COVID-19 bed availability dataset (BOR) consists of 14,823 rows with 34 attributes.

	tanggal_update	kabupaten_kota	tersedia_all	terpakai_all	bor_percent
2	2020-10-06 00:00:00	Bandung	83	59	39,91
3	2020-10-06 00:00:00	Bandung Barat	86	7	5,58
4	2020-10-06 00:00:00	Bekasi	565	332	32,25
5	2020-10-06 00:00:00	Bogor	696	470	43,69
6	2020-10-06 00:00:00	Ciamis	25	7	10,00
7	2020-10-06 00:00:00	Cianjur	106	60	37,74
8	2020-10-06 00:00:00	Cirebon	177	134	53,12
9	2020-10-06 00:00:00	Garut	69	31	11,31
10	2020-10-06 00:00:00	Indramayu	60	30	16,31
11	2020-10-06 00:00:00	Karawang	322	98	19,42
12	2020-10-06 00:00:00	Kota Bandung	635	379	34,48
13	2020-10-06 00:00:00	Kota Banjar	41	11	8,94
14	2020-10-06 00:00:00	Kota Bekasi	992	693	40,53
15	2020-10-06 00:00:00	Kota Bogor	354	196	36,81
16	2020-10-06 00:00:00	Kota Cimahi	132	70	24,23
17	2020-10-06 00:00:00	Kota Cirebon	149	68	14,02
18	2020-10-06 00:00:00	Kota Depok	513	368	45,22
19	2020-10-06 00:00:00	Kota Sukabumi	85	29	24,65
20	2020-10-06 00:00:00	Kota Tasikmalaya	65	49	19,13
21	2020-10-06 00:00:00	Kuningan	60	34	12,00
22	2020-10-06 00:00:00	Majalengka	17	9	23,14
23	2020-10-06 00:00:00	Pangandaran	19	0	0,00
24	2020-10-06 00:00:00	Purwakarta	92	89	42,28
25	2020-10-06 00:00:00	Subang	58	8	1,96

Figure 2. *Covid-19 Bed Availability Dataset (BOR)*
(opendata.jabarprov.go.id)

The attributes used are update date, city district, ICU without negative pressure with a ventilator available, ICU with negative pressure with a used ventilator, ICU without negative pressure without a ventilator available, ICU without negative pressure without a ventilator in use, ICU negative pressure with an available ventilator, ICU negative pressure with ventilator used, negative pressure ICU without ventilator available, negative pressure ICU without ventilator in use, isolation without negative pressure available, isolation without negative pressure used, negative pressure isolation available, negative pressure isolation in use, total ICU covid available, total ICU covid used, total covid isolation available, total covid isolation used, available covid nicu, used covid nicu, available covid trigger, used covid trigger, all available, all used, drill percent,igd covid available, igd covid used.

a. Analysis of Data Processing Techniques Using the C4.5. Algorithm
   The data obtained from the research site was then analyzed and processed using the C4.5 algorithm.
b. C4.5. Algorithm Design
   At this stage, the design process of the system model with the C4.5 algorithm will be carried out so as to form a decision tree and produce a predictive rule for students who repeat the course.
c. C4.5 . Algorithm Implementation
   The steps in this stage are:
   a) Specifies the attribute as the root and calculates the value of the attribute gain information.
   b) Compile the initial tree.
   c) Constructing an advanced Tree.
   d) Turning a tree into a rule

   As shown in Figure 5

d. Test Results

At this stage, the authors conducted testing and system design results using the Data Mining Rapid Miner software. The system is tested with procedures to explore and model the existing data so as to obtain a hidden relationship from the data.

The Stages Used in Making a Decision Tree Using

The C4.5 algorithm in this study is: 1. Prepare training data, which can be taken from historical data that has happened before and has been grouped into certain classes. 2. Determine the root of the tree by calculating the highest gain value for each attribute or based on the lowest entropy index value. Previously, the entropy index value was calculated, with the formula

$$Entropy\,(i) = -\sum_{j=1}^{m} f(i,j).\log 2\, f[(i,j)]$$

Information:
i = case set
m = number of partitions i
f(i,j) = the proportion of j to i

Gain Concept
Gain is the acquisition of information from attribute A. Calculate the gain value with the formula:

$$Entropy\ split = -\sum_{i=1}^{p} \frac{n1}{n}\ .\ IE\,(i)$$

Information:
p = number of attribute partitions
ni = the proportion of ni to i
n = number of cases in n
Repeat step 2 until all records are partitioned the decision tree partitioning process will stop when:

a. All tuples in records in node m get the same class
b. No attributes in partitioned record anymore
c. There are no records in the empty branch.

III. Discussion

Data mining implementation is done using RapidMiner software. The dataset for training and testing the C4.5 Algorithm model was obtained from West Java Covid-19 patient bed availability (BOR) data available on the West Java Open Data website. The attributes contained in the availability of beds for Covid-19 patients are the update date, city district, ICU without negative pressure with available ventilator, ICU with negative pressure with a used ventilator, ICU without negative pressure without a ventilator available, ICU without negative pressure without a used ventilator, negative pressure ICU with ventilator available, negative pressure ICU with ventilator in use, negative pressure ICU without ventilator available, negative pressure ICU without ventilator in use, isolation without negative pressure available, isolation without negative pressure used,

There are three patient bed statuses, namely all available, all used, percentage BOR

The main purpose of this study is to predictprobability valuewith the C4.5 algorithm which is used to predict the bed availability status of Covid-19 patients. In RapidMiner

Design, at the training stage there is a Read Excel operator to select the dataset and a C4.5 algorithm operator to predict the dataset. While in the test there is an operator *Apply Model*to run the C4.5 algorithm model and the Performance operator to measure the predicted performance of the NBC model.

Prediction resultmodel classificationthe C4.5 algorithm using RapidMiner software is shown in Figure 3

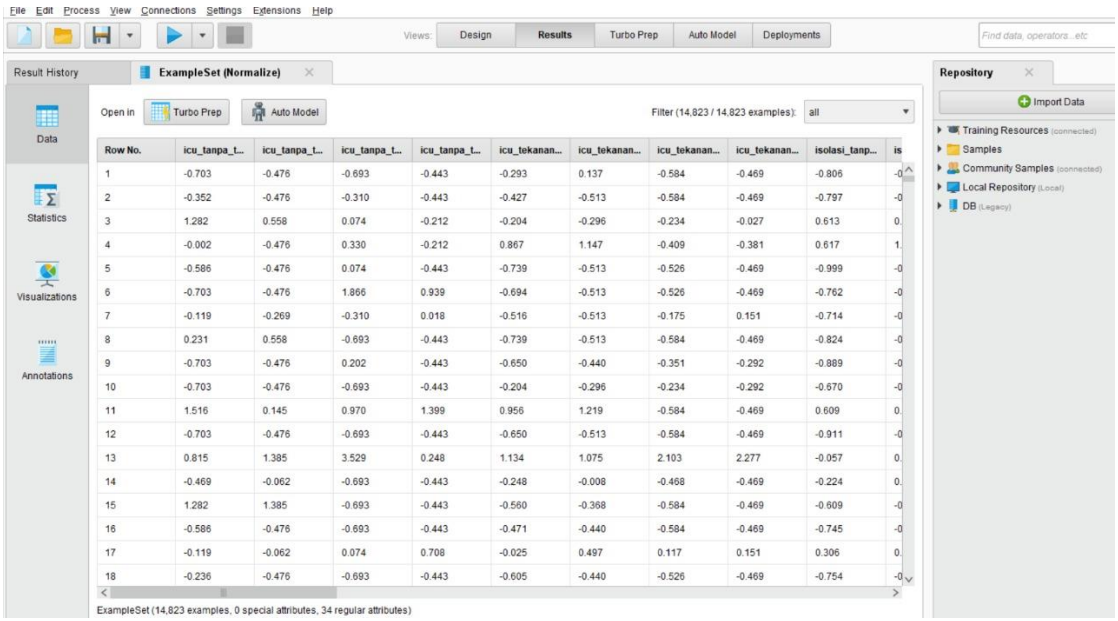

Figure 3. *Prediction Results with Rapidminer*

Furthermore, to make it easier to generate data, a cleaning stage is carried outIn this work stage, the data is cleaned through several processes such as filling in missing values, smoothing noisy data, and resolving inconsistencies found. Data can also be cleaned by dividing into segments of similar size and then smoothing (binning) as shown in Figure 4.

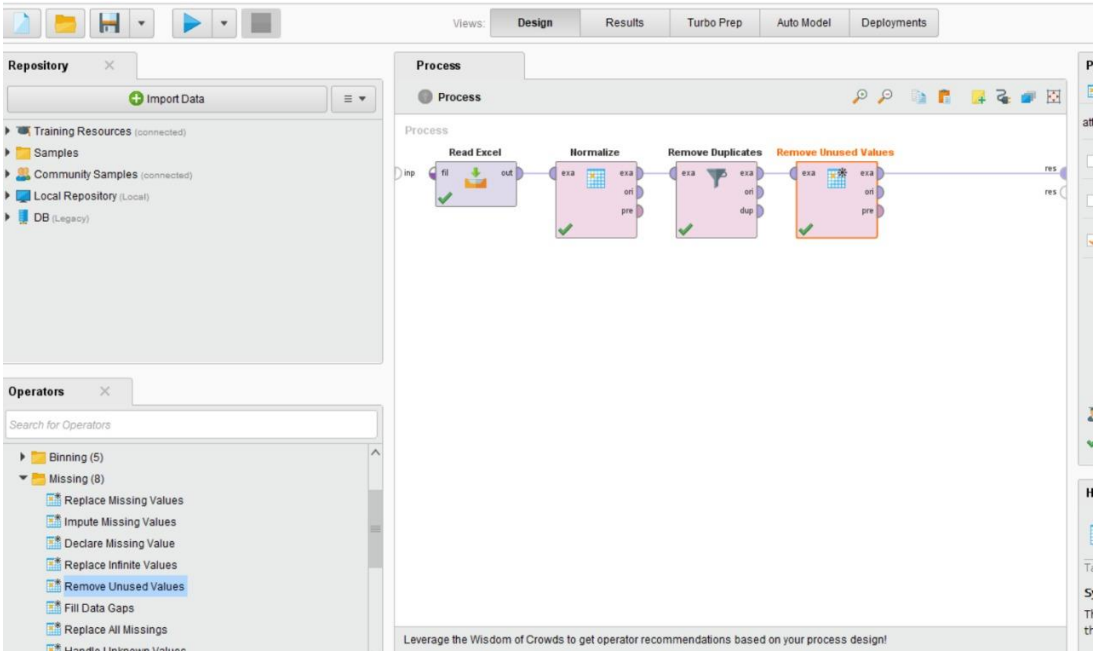

Figure 4. *Cleaning Results Using Rapidminer*

The C4.5 algorithm is a model for converting data into a decision tree with its rules. A decision tree or better known as a decision tree is an implementation of a system that humans have developed in finding and making decisions for these problems by taking into account various factors related to the scope of the problem. In general, the decision tree is a modeling description of a problem which consists of a series of decisions that lead to the resulting solution following the decision tree model of the C4.5 Algorithm in Figure 5

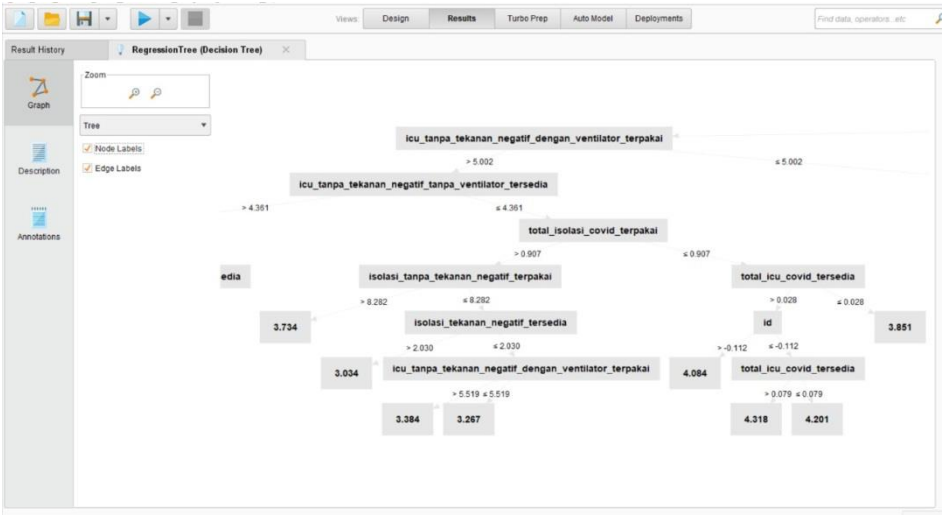

Figure 5. *Decision Tree*

The precision value shows how accurate the prediction is in recognizing the data according to the original class, as in Figure 6, while the recall value shows the classifier's ability to redefine information according to the original class.

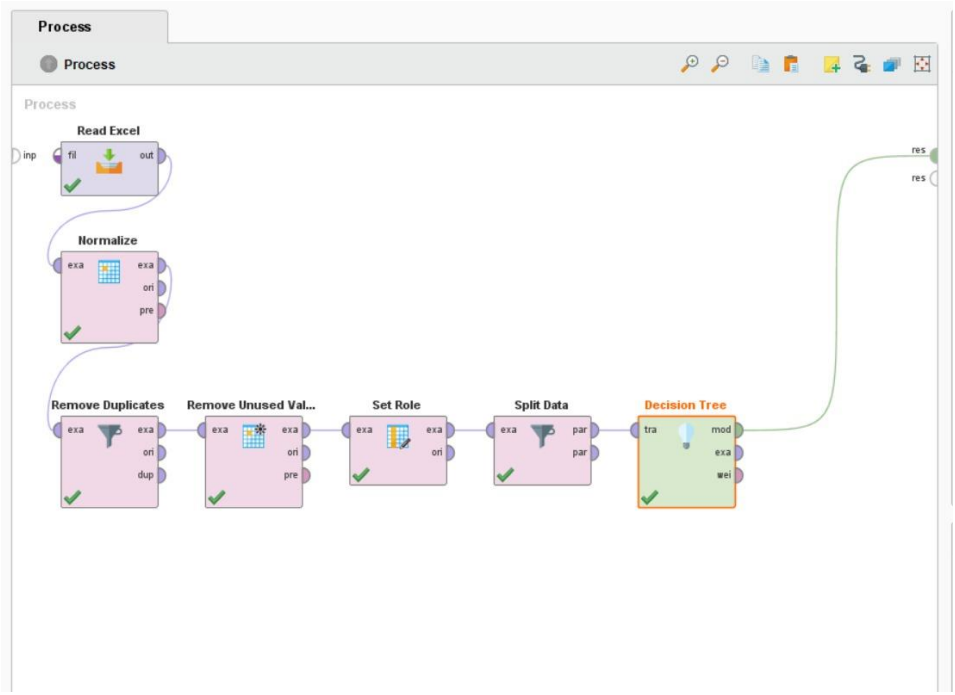

Figure 6. *The Classifier's Ability*

Accuracy indicates the overall performance of the built model. The test results in the form of a confusion matrix of three patient status classes can be analyzed in the form of average precision, average recall and accuracy as shown in table 1.

Table 1. Average Precision, Average Recall and Accuracy

Test Result Performance	
Parameter	mark
Precision	92%
Recall	98.72%
Accuracy	96.67%

Based on Table 1, it can be seen that the classification accuracy with C4.5 reaches 96.67%. The classification model has a recall average that is higher (98.72%) than the average precision (92%). This means that the built model has good sensitivity in recognizing the existing test data classes, while precision indicates the accuracy of the prediction results. Precision and recall Together are used to show how well a classifier predicts a model.

IV. Conclusion

In this study, a Data Mining model was developed to predict the status of Covid-19 patients using the C4.5 algorithm. C4.5 algorithm is used because of its reliability in classifying data based on its attributes, both in the form of numeric and categorical values. The model was built using the Covid-19 BOR (Bed Occupancy Rate) patient bed availability dataset obtained from the websiteopendata.jabarprov.go.idand implemented using RapidMiner software. The results of the C4.5 model for predicting the bed availability status of Covid-19 patients BOR (Bed Occupancy Rate) in West Java gave results measured in terms of precision, recall, and accuracy, the values were 92%, 88.72%, and 96.67, respectively. %. The results of this study are useful to be applied to real situations, to help medical personnel determine future actions, the number of real and large datasets with a balanced proportion of the value of each class is very good to get higher prediction accuracy.

References

A. Fattah and R. Setyadi, “Teknologi informasi dan pendidikan,” J. Teknol. Inf. dan Pendidik., vol. 12, no. 2, pp. 1–7, 2019

Albert Verasius Dian Sano, S.T., M.Kom, Cara Kerja Data Mining – Seri Data Mining For Business Intelligence (3)

Angelia, N. (2020). Analysis of Community Institution Empowerment as a Village Government Partner in the Participative Development Process. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3 (2): 1352-1359.

binus.ac.id/malang/2019/01/cara-kerja-data-mining-seri-data-mining-for-business-intelligence-3/

Bara, A., et.al. (2021). The Effectiveness of Advertising Marketing in Print Media during the Covid 19 Pandemic in the Mandailing Natal Region. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 4 (1): 879-886.

C. Long et al., “Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?,” Eur. J. Radiol., vol. 126, p. 108961, May 2020, doi: 10.1016/j.ejrad.2020.108961.

Dama, M., et.al. (2021). Implementation of Green Government by the Regional Government of East Kalimantan Province as a Form of Ecological Principles (Case Study of the

Impact of the Implementation of Coal Mining Policy in Samarinda City). Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 4 (3): 4445-4457.

Elmande, Yusuf. 2012. Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk PenentuanKualitas Beras : Studi Kasus Pada Perum BulogDivre Lampung. Jurnal TELEMATIKA MKOM

Faradillah, Sarah. 2013. Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer DenganMenggunakan Algoritma C4.5. Pelita Informatika Budi Darma, Volume : V, Nomor: 3.

S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2," Emerg. Infect. Dis. J., vol. 26, no. 7, 2020, doi: 10.3201/eid2607.200282.

L. K. Kumar and P. J. A. Alphonse, "Automatic Diagnosis of COVID-19 Disease using Deep Convolutional Neural Network with Multi-Feature Channel from Respiratory Sound Data : Cough , Voice , and Breath Reference : To appear in : Received Date : Revised Date : Accepted Date : Abstract :," Alexandria Eng. J., 2021, doi: 10.1016/j.aej.2021.06.024.

M. Sukmana, M. Aminuddin, and D. Nopriyanto, "Indonesian government response in COVID-19 disaster prevention," East Afrian Sch. J. Med. Sci., vol. 3, no. 3, pp. 81–6, 2020, doi: 10.36349/EASMS.2020.v03i03.025.

M. Abed Alah, S. Abdeen, and V. Kehyayan, "The first few cases and fatalities of Corona Virus Disease 2019 (COVID-19) in the Eastern Mediterranean Region of the World Health Organization: A rapid review," J. Infect. Public Health, vol. 13, no. 10, pp. 1367–1372, 2020, doi: 10.1016/j.jiph.2020.06.009.

Ningrum, P.A., Hukom, A., and Adiwijaya, S. (2020). The Potential of Poverty in the City of Palangka Raya: Study SMIs Affected Pandemic Covid 19. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3 (3): 1626-1634.

Opendata.jabarprov.go.id

Rahmaniah, S.E., Syarmiati, and Paramita, R.R. (2021). Community Resilience and Digital Literacy Model during the COVID-19 Outbreak in Indonesia. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 4 (4): 5339-5346.

Sihombing, E.H., and Nasib. (2020). The Decision of Choosing Course in the Era of Covid 19 through the Telemarketing Program, Personal Selling and College Image. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3 (4): 2843-2850.

T. Miller, "Explanation in artificial intelligence : Insights from the social sciences," Artif. Intell., vol. 267, pp. 1–38, 2019, doi: 10.1016/j.artint.2018.07.007.