Vol. 2 No. (2) (December 2024) 88-94

**Journal of Data Insights**

http://journalnew.unimus.ac.id/index.php/jodi

# Decision Tree Classification Prediction of Covid-19 Cases in Indonesia

**Amaliah Sholeha Arafat**[1], **Aprilla Anawai Basman**[2], **Fatkhurokhman Fauzi**[3], **Saeful Amri**[4]
[1]Universitas Muhammadiyah Semarang, Indonesia
[2]Universitas Muhammadiyah Semarang, Indonesia
[3]Universitas Muhammadiyah Semarang, Indonesia
[4]Universitas Muhammadiyah Semarang, Indonesia

**DOI**: https://doi.org/10.26714/jodi.v2i2.213

## Article Information

Article History:
Submitted: 28th October 2024
Revised: 06th December 2024
Accepted: 31th December 2024

*Keywords:*
*Forecasting, COVID-19, Classification; Decision Tree.*

***Abstract***

*Forecasting is the prediction of current and future events using data from past events. The goal of forecasting is to minimize the forecast error so as to provide a greater level of confidence. In the context of the COVID-19 pandemic, forecasting the number of cases can help anticipate spikes in cases so that preparations can be more mature and the impact can be minimized. Forecasting methods can be divided into three general classifications, namely qualitative, time series, and causal methods. Time series methods are divided into statistical and machine learning methods. Machine learning methods are better at forecasting because they can accommodate non-linear and complex relationships between inputs and outputs. One of the machine learning methods used is the decision tree, which is a prediction model in the form of a tree structure or hierarchy. Decision tree is a data processing way to predict the future by building classification and regression models in the form of a tree structure. In this study, decision tree classification is used to forecast positive cases of COVID-19 in Indonesia using the Python programming language.*

✉ Corresponding Author:
E-mail: fatkhurokhmanf@unimus.ac.id

e-ISSN: 2988 - 2109

## INTRODUCTION

Forecasting or commonly referred to as forecasting is the prediction of an event in the present and future data using data on past events [1]. Forecasting is an important medium in a design or planning to produce effective and efficient results [2]. Forecasting has the goal of minimizing the error in forecasting activities (forecast error) which is usually measured by mean square error, mean absoulute error, and so on. Minimizing the error in forecasting can provide a greater level of confidence [3]. Through forecasting, the Indonesian government can consider policy preparation in dealing with future covid-19 cases [4]. The emergence of the Covid outbreak at the end of 2019 which is an infectious disease and can attack the respiratory organs which is very deadly in China. This infectious disease was first discovered in Wuhan, Hubei Province spreading rapidly throughout China and around the world [5]. On March 11, 2020, Covid-19 was officially designated by WHO (World Health Organization or World Health Organization) as a pandemic, which means that the Corona Virus has spread rapidly to various countries around the world [6].

Indonesia as a developing country with the 4th largest population in the world could receive a huge impact from COVID-19 if the pandemic is not handled properly and lasts long [7]. Indonesia's healthcare capacity and number of tests are inadequate, while a total lockdown is not possible. The government and hospitals need to anticipate a surge in the number of cases to avoid a surge in deaths like in Italy [8]. Forecasting the number of COVID-19 cases can help anticipate the surge of cases so that preparations can be more mature and minimize the impact [9], [10]. COVID-19 can spread through respiratory droplets produced when a person coughs, sneezes, talks or breathes. Common symptoms include fever, cough, shortness of breath, fatigue, and impaired taste or smell. Some cases can develop into serious illnesses, such as pneumonia, and can lead to death, especially in vulnerable people such as the elderly and those with pre-existing medical conditions

There are three classifications of forecasting methods in general, namely qualitative, time series and causal methods [11]. Time series forecasting methods are divided into two, namely statistical and machine learning methods [12]. Statistical methods use statistical approaches, such as linear regression, and exponential smoothing [13]. Machine learning methods are algorithms that allow users to find and describe structural patterns in data, so that by using the structural patterns in the data, machine learning performs forecasting of future data [14]. Compared to traditional statistical methods, machine learning methods perform better forecasting because they can accommodate non-linear and complex relationships between inputs and outputs [15].

Classification is the process of finding models or functions that explain or distinguish concepts or classes of data, with the aim of being able to estimate the class of an object whose lab is unknown. there are several methods contained in data mining classification, including decision trees, neaural networks, k-Nearest Neighbor Classifiers, Case-Based Reasoning and genetic algorithms. One of the most popular and easy to interpret classification methods is the decision tree. This method is one of the functional data mining methods that uses a tree representation to determine classification rules. Decision tree is a flowchart structure that resembles a tree, where each internal node indicates a test on the attribute, each branch represents a class or class distribution [16].

A decision tree is a prediction model using a tree structure or hierarchical structure. The concept of a decision tree is to convert data into decision trees and decision rules [17]. The Decision Tree method is one way of data processing in predicting the future by building classification and regression models in the form of a tree structure. Decision Tree is a visual model to enable understanding of the structured, gradual and rational decision-making process. In the decision tree diagram the structured decision-making process can be seen visually [18].

Decision tree is the most popular classification method used. Apart from being relatively fast to build, the results of the model built are easy to understand. In the decision tree there are 3 types of nodes, namely: (a) Root Node is the topmost node, at this node there is no input and can have no output or have more than one output. (b) Internal Node is a branching node, in this node there is only

one input and has a minimum of two outputs. (c) Leaf node or terminal node is the final node, in this node there is only one input and has no output [19]. In this study, Classification decision tree is used to forecast positive cases of Covid-19 in Indonesia. With the software that will be used, namely Python

## METHODS

In classification techniques, there are several classification algorithms that can be used, including: decision tree, rule based, neural network, support vector machine, naive bayes, rough set, and nearest neighbor. One of the most commonly used data mining classification techniques is the decision tree. A decision tree is a flowchart structure that resembles a tree, where each internal node signifies a test on an attribute, each branch represents the test result, and the leaf nodes represent classes or class distributions. The flow in a decision tree is traced from the root node to the leaf node that holds the class prediction for that instance. Decision trees are easy to convert to classification rules. Decision tree techniques are easier to use, for several reasons:

a) Compared to a JST or Bayesian classifier, a decision tree is easy for humans to interpret/handle.
b) While JST training can take a lot of time and thousands of iterations, decision trees are efficient and appropriate for large data sets.
c) Algorithms with decision trees do not require additional information other than that contained in the training data (i.e., domain knowledge of distributions in the data or classes).
d) Decision trees show good classification accuracy compared to other techniques.

Transform data into decision trees and rules.

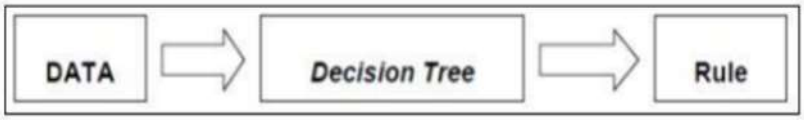


**Figure1.** The Concept of Decision Tree

The concept of decision tree data, namely:

a) Data is expressed in tabular form with attributes and records.
b) Attributes express a parameter that is made as a criterion in tree formation.
c) Attributes have values called instances.

## RESULTS AND DISCUSSION

Figures 2 and 3 show the model comparison between the target data and the assumed data on the total spread of COVID-19 cases in Indonesia. The blue line shows the total number of people infected with COVID-19, the orange line shows the total number of active people.

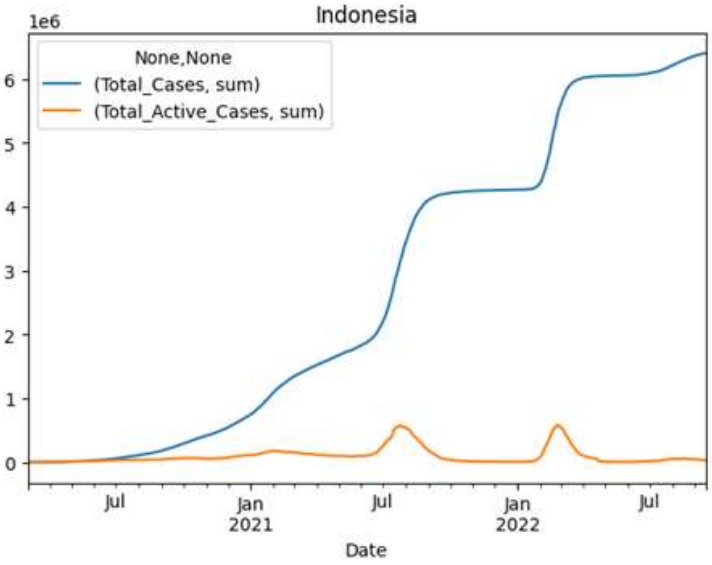


**Figure 2.** Graph of total target covid-19 cases

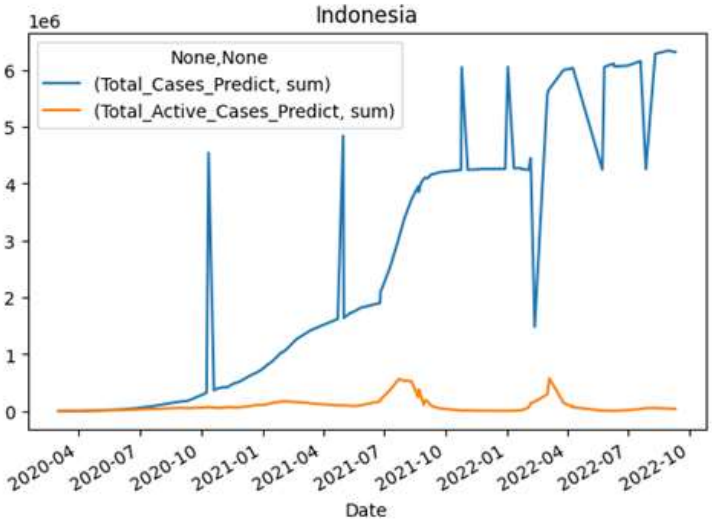


**Figure 3**. Assumption graph of total covid-19 cases

Figures 4 and 5 show the model comparison between the target data and the assumption data for the total spread of COVID-19 cases in Indonesia. The blue line shows the total number of people infected with COVID-19, the orange line shows the total number of people who died from COVID-19.

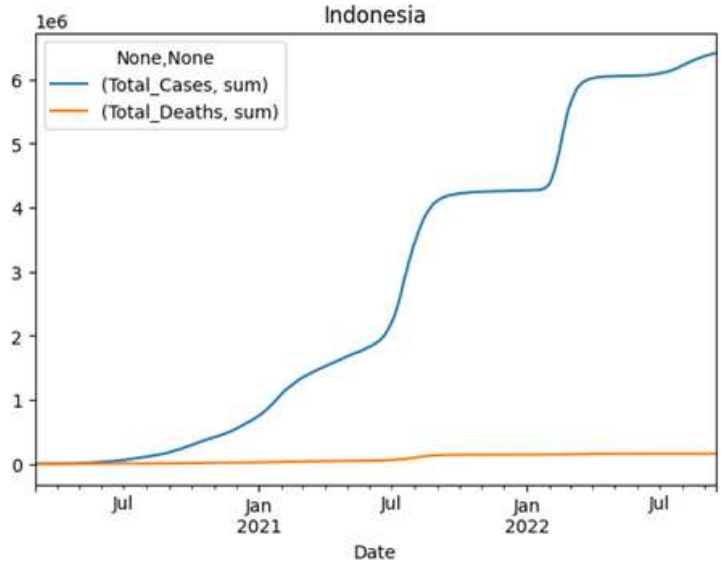


**Figure 4**. Graph of target total covid-19 cases

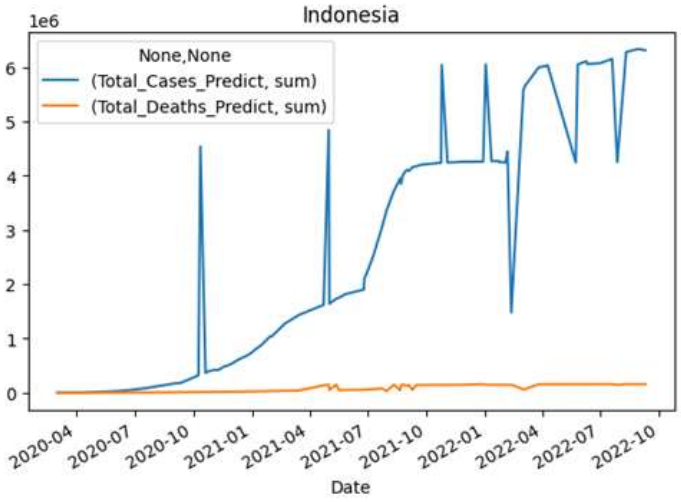


**Figure 5**. Assumption graph of total covid-19 cases

Figures 6 and 7 show the model comparison between the target data and the assumed data on the total spread of COVID-19 cases in Indonesia. The blue line shows the total number of people who recovered from COVID-19, the orange line shows the total number of people who died from COVID-19.

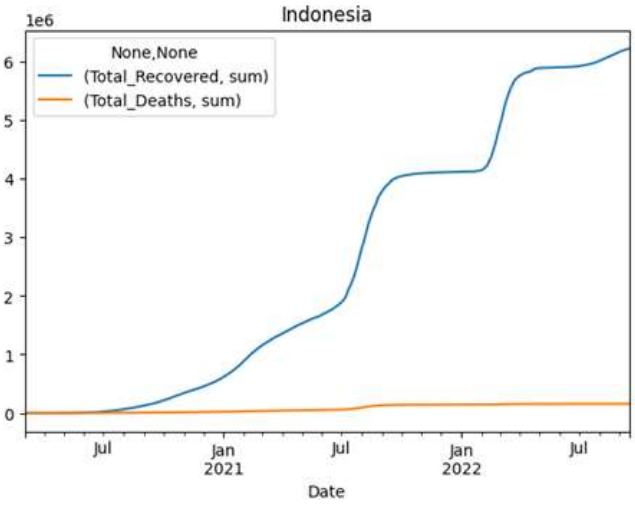


**Figure 6**. Graph of total target covid-19 cases

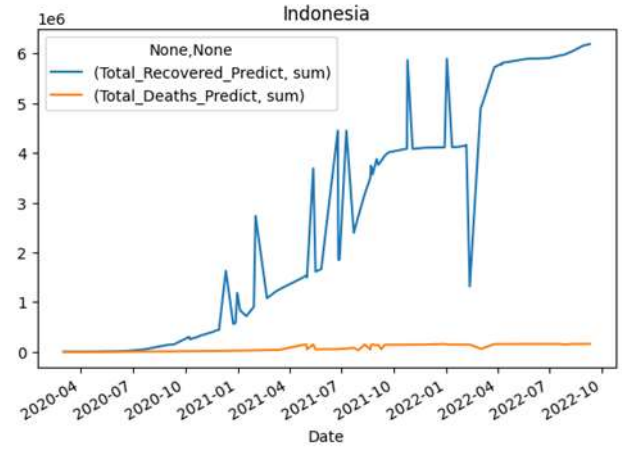


**Figure 7**. Assumption graph of total covid-19 cases

## CONCLUSION

Based on the results of the analysis using the Decision Tree method, it can be concluded that the results of the assumption data resemble the target data, where the target graph and the assumption graph have a similar distribution graph, namely the peak of the spread of the highest covid-19 cases in Indonesia is in 2022. So it can be said that the Decision Tree method is effectively used for forecasting covid-19 cases in Indonesia.

## REFERENCE

[1] B. P. Jumadil Nangi, Siti Hartinah Indrianti, "Peramalan Persediaan Obat Menggunakan Metode Triple Exponential Smoothing (Tes) (Studi Kasus : Instalasi Farmasi Rsud Kab. Muna)," vol. 4, no. 1, pp. 135–142, 2018.

[2] F. I. Y. Dwi Hilda Anjasari, Eko Listiwikono, "Perbandingan Metode Double Exponential Smoothing Holt Dan Metode Triple Exponential Smoothing HoltWinters Untuk Peramalan Wisatawan Grand Watu Dodol PERBANDINGAN," J. Pendidik. Mat. Mat., vol. 2, no. 2, pp. 12–25, 2018.

[3] K. S. K. Dewi and K. R. Suwena, "Analisis Peramalan Tingkat Jumlah Tamu Menginap Menggunakan Metode Single Exponential Smoothing Di Villa X Di Desa Gobleg, Kabupaten Buleleng Tahun 2018," J. Pendidik. Ekon. Undiksha, vol. 9, no. 2, p. 335, 2019.

[4] F. R. Perdana, Daryanto, and H. Wahyu, "Perbandingan Metode DES (Double Exponential Smoothing) Pada Peramalan Penjualan Rokok (Studi Kasus Toko Utama Lumajang)," Jur. Tek. Inform. Fak. Tek. Univ. Muhammadiyah Jember, no. 1110651142, pp. 1–8, 2018.

[5] K. Amboro, "Kontekstualisasi Pandemi Covid-19 dalam Pembelajaran Sejarah," Yupa Hist. Stud. J., vol. 3, no. 2, pp. 90–106, 2020.

[6] T. C. Morphology, Merekam Pandemi Covid 19. 2021.

[7] R. Djalante et al., "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," Prog. Disaster Sci., vol. 6, pp. 100091, 2020, doi: 10.1016/j.pdisas.2020.100091.

[8] I. Rudan, "A cascade of causes that led to the COVID-19 tragedy in Italy and in other European union countries," J. Glob. Health, vol. 10, no. 1, 2020, doi: 10.7189/JOGH.10.010335.

[9] F. Petropoulos dan S. Makridakis, "Forecasting the novel coronavirus COVID-19," PLoS One, vol. 15, no. 3, pp. 1–8, 2020, doi: 10.1371/journal.pone.0231236.

[10] J. Devaraj et al., "Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?," Results Phys., vol. 21, no. 2, pp. 1–25, 2021, doi: 10.1016/j.rinp.2021.103817.

[11] Render, B., Stair Jr, R. M., & Hanna, M. E. (2012). Quantitative Analysis For Management.

[12] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE, 13(3), 1–26. doi:10.1371/journal.pone.0194889

[13] Abraham, B., & Ledolter, J. (2009). Statistical methods for forecasting (Vol. 234): John Wiley & Sons.

[14] Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31(1), 76-77.

[15] Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1): Springer series in statistics New York.

[16] Andrew W. Moore, "Decison Trees", Carnegic Mellon University. http://www.cs.cmu.edu/awm

[17] Nugroho, K. W. (2002). Eksentrik Digraf dari Graf Star, Graf Double Star dan Graf Komplit Bipartit. Jember: Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember

[18] Junaedy, Musdar, I. A., & Angriani, H. (2017). Implementasi Metode Decision Tree Dalam Menentukan Pemberian Kredit Mobil Menggunakan Visual Basic (Studi Kasus UD PUTRA MAS Makassar). Jtriste, 4(1), 55–67.)

[19] Han, Jiwei and Michelin Kamber. (2001). Data Mining : Concepts and Techniques. Intelegent Databases Systems Research Lab, School of Computing Science, Simon Fraser University.