JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)
Vol. 6 No. 2, February 2023 E-ISSN : 2580-2879

Analysis of Method C5.0 in Identifying Factors in the Number of Covid-19 Increasing or Decreasing After Getting the Vaccine

Alferedo Tanzil[1], Randy Aldany Barasa[2], Yonata laia[3], Jepri Banjarnahor[4]
[1,2,3,4] Faculty of Technology and Computer Science, Prima University of Indonesia Medan
Jl. Belanga No.1, Sei Putih Tengah, Kec. Medan Petisah, Kota Medan, Sumatera Utara 20118
E-mail: yonata@unprimdn.ac.id

ABSTRACT- Because we now know that many things make everyone sick, such as fever, flu, cough, and other diseases that are said to be easily transmitted, we need a system that can overcome the above problems. This study uses the K-NN method to examine what factors influence the increase in the number of people infected with Covid-19. The factors tested in this study were frequent violations of health practices, overcrowding, and weak immune systems. The K-NN method can overcome the problem of knowing the factors causing the increase in Covid-19 patients after vaccination.

Key words : Covid19, Identify, Vaccine, Covid-19.

1. INTRODUCTION

One of the reasons for the increase in COVID-19 cases in Indonesia is poor hygiene practices. Another factor is re-vaccination in Indonesia which has not been widely carried out, especially in vulnerable groups such as the elderly [1]. Several provinces in Indonesia that experienced a significant increase in COVID-19 cases were DKI Jakarta, West Java, Banten, Bali, and East Java. According to Dicky Budiman, an epidemiologist at Griffith University in Australia, the increase in COVID-19 cases in Indonesia is due to a weakening of protection. The source above author obtained information from www.halodoc.com [1].

The number of people infected with the coronavirus, or Covid-19 for short, has increased, bringing the total number of infections of the deadly virus to 181 million nationwide in June 2021 and the death toll to 3.93 million, while Indonesia has. See an increase in cases. They are associated with. Data on 2,115,304 deaths in 57,138 patients and 1,850 patients recovered in 34 provinces [2, 3, 4]. The Government of the Republic of Indonesia has reported 6,627,538 people infected with COVID-19 as of 24 November 2022. 159,524 deaths related to the COVID-19 virus have been notified, and 6,403,551 patients have recovered from the disease. WHO is working with the Indonesian government to control the situation and prevent the spread of this disease [5-10].

Researchers previously designed a system to process and store sales data at the Cimah post office using the C5.0 algorithm. The C5.0 algorithm is used to determine the highest information gain value to determine the best root to be used as a node or branch [10-15]. The research problem is that it is difficult to determine why the number of Covid-19 patients is increasing. In accordance with the discussion above, the author wants to use the C5.0 method to identify the causes of the increase in Covid-19 cases in Indonesia, especially in North Sumatra.

2. LITERATURE STUDY

Patients diagnosed with Covid-19 in Indonesia came from cases in Jakarta, where the patient had contact with Japanese foreigners living in Malaysia. After removal, the patient complained of fever, cough, and shortness of breath [16]. This epidemic has been declared a global health emergency. This virus has made it difficult for everyone's daily activities. Quarantine alone may not be enough to prevent the spread of the COVID-19 virus, and the global impact of infection with this virus is more worrying [17].

3. METHOD

The research in this work is design research or also called development research, and as a system development methodology, the SDLC *(Systems Development Life Cycle)* method is used.

The decision tree is a predictive classification technique. Decision trees can turn a database into a rulebook with a decision tree representation. The rules are easy to understand, so decision trees are an easy-to-learn and popular classification technique. The Decision tree is a classification technique for objects or data with a decision tree representation. Decision trees are generally used to examine patterns and see correlations between multiple input and output variables. The decision tree diagram consists of three parts, namely:

1. Root Node: The root node at the top of the decision tree.
2. Internal nodes: namely branch nodes,
3. Leaf Node: a decision node has only one input and no output, only one input and at least two outputs.

The requirements for implementing a decision tree algorithm are a training dataset that provides a

target such as a machine learning trainer or teacher, the training material must be large and varied, and the target must be definite.

The C5.0 algorithm is a data mining technique with a decision tree-based classification algorithm. Ross Quinlan refined this algorithm from the ID3 and C4.5 algorithms in 1987. This algorithm is considered better than the previous algorithm in terms of accuracy and memory. The speed of model creation is considered very fast compared to other algorithms. This algorithm can also handle attributes with discrete or continuous values. This advantage makes the C5.0 algorithm superior to other algorithms. This algorithm starts with all the attributes that are the roots of the decision tree. Then the attribute with the highest gain value is selected as the root node. In addition, other attributes are evaluated in the same way to get the next root node. The equation for calculating the information utility of all cases is as follows:

The C5.0 algorithm is a method used by the author to determine potential customers at the Cimahi Post Office. In the process of forming a decision tree, the highest gain information value will be selected as the root for the next node. Formula (1) is a formula to find out the overall entropy and entropy of each attribute [9]:

$$Entropy\ (S) = \sum_{i=1}^{n} -pi * \log_2 pi \quad (1)$$

Where S is the number of cases, n is the number of parts of S, and pi is the ratio of Si to S.

Formula (2) is a formula
to find out the gain of each attribute:

$$Gain(S,A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (Si) \quad (2)$$

Where S is the number of cases, A is the attribute used, n is the number of attribute sections, |Si| is the number of cases in I, and |S| part is the number of cases in S.

Formula (3) is a formula
to calculate the gain ratio:

$$Gain\ Ratio = \frac{Gain\ (S,A)}{\sum_{i=1}^{n} Entropy\ (Si)} \quad (3)$$

Where is the gain value of the variable, and $\sum_{i=1}^{n} Entropy\ (Si)$ is the sum of the entropy values of the variable, the criteria used?

4. RESULTS AND DISCUSSION

Factors leading to an increase in Covid-19 cases

1. Healthy habits are often not followed.
2. The Human Crowd
3. Weak stamina

Table of Confirmed Covid-19 Case Data for 2020-2023

The table below was created to find out confirmed cases of Covid-19 in 2020-2023. Factors causing the high number of cases and deaths in various provinces are delays in conveying information to the public. The lack of implementation of health appeals, starting from not using masks and not keeping a distance in public places, so the Covid-19 case is increasing.

Province	Case	Healed	Death	Isolation
DKI Jakarta	1.539,191	1,522,544	15,941	562
Banten	363,348	358,533	2,982	0
Jawa Barat	1,233,566	1,215,383	16,102	0
Jawa Timur	637,477	604,999	32,288	87
Sumatera Utara	163,712	160,262	3,390	0

Table of Covid-19 Case Data After Vaccines for September 2021

The table below was created to establish data on Covid-19 cases after the launch of the vaccine in January, which increased from place to place due to public discipline in implementing health protocols and many large-scale actions. Others, such as airports, concerts, concerts, and shopping centers.

Province	Case	Healed	Death	Isolation
DKI Jakarta	857,765	842,510	13,539	1,195
Banten	131,475	128,002	2,670	1,666
Jawa Barat	702,722	685,219	14,624	499
Jawa Timur	395,475	364,251	29,413	268
Sumatera Utara	104,706	100,076	2,837	0

5. CONCLUSIONS

Based on the problems identified by industry research, it is difficult to determine what factors are contributing to the increase in Covid-19 patients. In this study, the K-NN method was used to obtain the results of the reasons for the increasing number of people infected with the coronavirus.

6. REFERENCES

[1] Azanella L. A. (2020). Apa itu PSBB Hingga Upaya Pencegahan Covid-19.

https://www.kompas.com/tren/read/2020/04/13/153415265/apa-itu-psbb-hingga-jadi-upaya-pencegahan-Covid-19. Diakses pada 10 Juni 2020.

[2] Bender L. (2020). Pesan dan Kegiatan Utama Pencegahan dan Pengendalian COVID-19 di Sekolah. Publikasi UNICEF.

[3] Bnpb.go.id. (2020). Sebaran Daerah PSBB.,https://bnpb.go.id/infografis/update-sebaran-daerah-psbb-Covid19. Diakses pada 17 Mei 2020.

[4] Budianto, Y. (2020). Memahami Karakter Virus dan Penyakit Covi-19. https://bebas.kompas.id/baca/riset/2020/03/14/memahami-karakter-virus-dan-penyakit-korona-Covid-19/. Diakses tanggal 1 Juni 2020.

[5] Fadli, A. (2020). Mengenal Covid-19 Dan Cegah Penyebarannya Dengan "Peduli
Lindungi" Aplikasi Berbasis Andorid. Artikel Pengabdian Kepada
Masyarakat Jurusan Teknik Elektro

[6] Hardani et al. (2020). Metode Penelitian Kulaitatif dan Kuantitatif. CV. Pustaka
Ilmu: Yogyakarta.

[7] Hartono. (2010). Statistik untuk Penelitian. Yogyakarta: Pustaka Pelajar.
Health.detik.com.(2020). Cara Penyebaran Virus Corona Covid-19 menurut
WHO. Diakses pada 12 Juni 2020, dari
https://theconversation.com/survei-pengetahuan-dan-partisipasi-masyarakat-selama-psbb-masih-rendah-perlu-ada-perbaikan-selama-memulai-pelonggaran-140083.

[8] Coronavirus.jhu.edu. Dashboard by the Center for Systems Science and Engineering (CSSE). https://coronavirus.jhu.edu/map.html. Diakses pada
Tanggal 8 Mei 2020.

[9] Penyebab Melonjaknya Kasus Covid-19 di Indonesia Halaman all - Kompas.com
Covid-19 (jakarta.go.id)
Corona Provinsi Banten (bantenprov.go.id)
Pikobar - Pusat Informasi dan Koordinasi COVID-19 Jawa Barat (jabarprov.go.id)
JATIM TANGGAP COVID-19 (jatimprov.go.id)
Covid19 - Beranda (sumutprov.go.id)
COVID-19 di Indonesia @kawalcovid19 - Google Drive