GOVERNMENT POLICIES MODELING IN CONTROLLING INDONESIA'S COVID-19 CASES USING DATA MINING

Ultach Enri [1*)] ; Eka Puspita Sari [2]

Teknik Informatika[1]
Universitas Singaperbangsa Karawang[1]
unsika.ac.id[1]
ultach@staff.unsika.ac.id[1]

Teknologi Komputer[2]
Universitas Bina Sarana Informatika[2]
bsi.ac.id[2]
eka.eps@bsi.ac.id[2]

(*) Corresponding Author

Abstract— *Since the positive case of covid-19 in Indonesia, the government has taken several policies with the purpose of controlling the spread of the covid-19 virus, which has been regulated in Government Regulation No. 21 of 2020. The purpose of research is to obtain a model of government policy in controlling cases of covid by using data mining classification techniques and obtain attributes that have the greatest weight, as well as look at the impact of policies that have been carried out by the government on the cases of covid-19 in Indonesia. The methodology used in the research is Knowledge Discovery In Database (KDD). Based on the research that has been done, it can be concluded that the policies that have been done by the government in controlling cases of covid-19 can be said to be successful, the C4.5 algorithm is the algorithm that gives the best results compared to the Deep Learning algorithm, as well as the attribute that has the greatest weight is canceled public events. Secondary data will be used in this research.*

Keywords: *data mining, covid-19, government policies.*

Abstrak— Sejak tedapat kasus positif covid-19 di Indonesia pemerintah telah mengambil beberapa kebijakan dengan tujuan untuk mengendalikan penyebaran virus covid-19, yang telah diatur pada Peraturan Pemerintah RI No. 21 Tahun 2020. Tujuan dari penelitian adalah untuk mendapatkan model kebijakan pemerintah dalam mengendalikan kasus covid dengan menggunakan teknik klasifikasi data mining, dan mendapatkan atribut yang mempunyai bobot terbesar, serta melihat dampak kebijakan-kebijakan yang telah dilakukan pemerintah terhadap jumlah covid-19 di Indonesia. Metodologi yang digunakan pada penelitian adalah Knowledge Discovery in Database (KDD). Pada penelitian ini dapat disimpulkan bahwa kebijakan-kebijakan yang telah dilakukan pemerintah dalam mengendalikan kasus covid-19 dapat dikatakan berhasil, algoritma C4.5 adalah algoritma yang memberikan hasil terbaik dibandingkan dengan algoritma Deep Learning, serta atribut yang mempunyai bobot terbesar adalah cancel public events. penelitian ini menggunakan data sekunder.

Kata Kunci: *data mining, covid-19, kebijakan pemerintah.*

INTRODUCTION

Since the positive case of covid-19 in Indonesia, both local and central governments have taken several policies with the purpose of controlling the spread of the covid-19 virus. One of the policies triggered is Social distancing or physical distancing has been implemented in Indonesia since March 16, 2020 (Parhusip, 2020), which restricts the activities of certain residents in a zone suspected of being contaminated with Covid-19 in such a way as to avoid the conceivable spread of the infection (Anung Ahadi Pradana, Casman, 2020). Based on Government Regulation No. 21 of 2020 concerning large-scale social restrictions, it is stated that PSBB can be done when the number of cases and/or the number of passing increases and spreads rapidly to various districts, and there is an epidemiological affiliation with similar events in other districts or nations. (Lengkong et al., 2021). Blavatnik School of Government, University of Oxford has systematically collected information collected based on policies taken by governments in more than 180 countries including Indonesia.

Data mining classification techniques have been widely used in various fields of science, among which are often used is C4.5 algorithm such as diabetes prediction (Noviandi, 2018), heart disease prediction (Rohman, Suhartono, & Supriyanto,

2017), predictions of potential blood donors (Wahono & Riana, 2020), where C4.5 is a development of id3 algorithm. C4.5 can handle missing data, handling continuous data, pruning, rules, and also use the gain ratio as a solving criterion (Santosa & Ardian, 2018) In addition to the C4.5 algorithm, there is also a Deep Learning(DL) algorithm that has been widely used in Artificial Intelligence (AI), such as dermatology applications, where DL is a subset of machine learning, which can complete tasks or can answer specific questions (Murphree et al., 2020)

In some studies that have been conducted by some researchers, the C4.5 algorithm provides considerable accuracy including 86.59% with an AUC value of 0.982(Rohman et al., 2017), accuracy value of 70.32% (Noviandi, 2018), accuracy value of 90% (Bahri, Marisa Midyanti, Hidayati, Sistem Komputer, & Mipa, 2018)and accuracy value of 91.76% (Hijrah, Mukhlizar, & Pandria, 2020). However, some studies stated the C4.5 algorithm performs less well compared to other algorithms, and of them is DL (Mutrofin, Machfud, Satyareni, Ginardi, & Fatichah, 2020). DL also provides high accuracy of 98.54% (Amelia, Eosina, & Setiawan, 2018). Therefore the author will use the C4.5 and DL algorithms.

The purpose of research is to obtain a model of government policy in controlling covid-19 cases by using data mining classification techniques by comparing C4.5 and DL algorithms, and obtaining attributes that have the greatest weight, as well as looking at policies that have been done by the government, whether it succeeds in decreasing cases of covid-19 in Indonesia.

MATERIALS AND METHODS

The data used in this study is secondary data from The Oxford COVID-19 Government Response Tracker, Blavatnik School of Government, University of Oxford. The data was accessed on February 26, 2021, the data used is time series data and only took Indonesia data, which amounts to 420 records, with attributes as seen in table 1.

Table 1. Attribute description

No	Attribute	Description
1	School closing	closings of schools and universities
2	Workplace closing	closings of workplaces
3	Cancel public events	canceling public events
4	Restrictions on gatherings	limits on gatherings
5	Close public transport	closing of public transport
6	Stay at home requirements	stay at home requirements
7	Restrictions on internal movement	restrictions on internal movement between cities/regions
8	International travel controls	restrictions on international travel for foreign travelers
9	Testing policy	testing policy for current infection (PCR tests)
10	Contact tracing	government policy on contact tracing after a positive case
11	facial_coverings	policies on the use of facial coverings outside the home
12	vaccination_policy	policies for vaccine delivery for different groups

Source: (Oxford University, 2021)

The methodology that will be used in the research is Knowledge Discovery in Database (KDD), where there are three main steps, namely pre-processing data, modeling data, and post-processing data (Benhar, Idri, & L Fernández-Alemán, 2020). Where Data preprocessing is the most important and crucial step in the KDD process (Idri, Benhar, Fernández-Alemán, & Kadi, 2018)

Figure 1 explains the research framework of this study, where after the dataset is obtained The next step is to pre-process the data, with the detection of missing values and also outliers, continues with selecting attributes and labeling because the algorithm will be used as an algorithm that belongs to the supervised technique that requires labels. Next, the author will use a correlation matrix to be able to know the relationship between attributes (Ibrahim, 2017), and will compare between C4.5 and DL algorithms. In the post-processing stage, evaluation of modeling will be conducted using 10-fold cross-validation, accuracy, and AUC values.

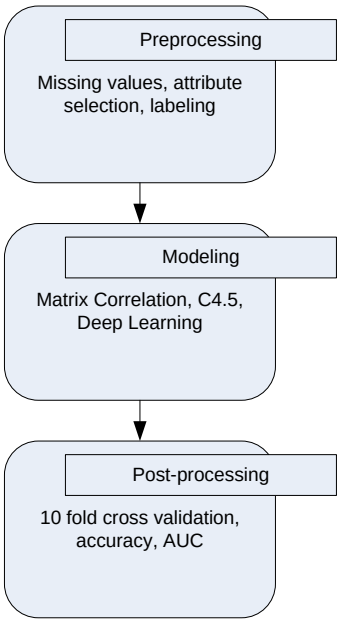

Figure 1. Research framework

RESULTS AND DISCUSSION

Matrix Correlation will give a result in the weighting of each attribute as shown in figure 2, where the cancellation of general activities has the greatest weight followed by the international travel ban, with the smallest weight is the restrictions on gatherings.

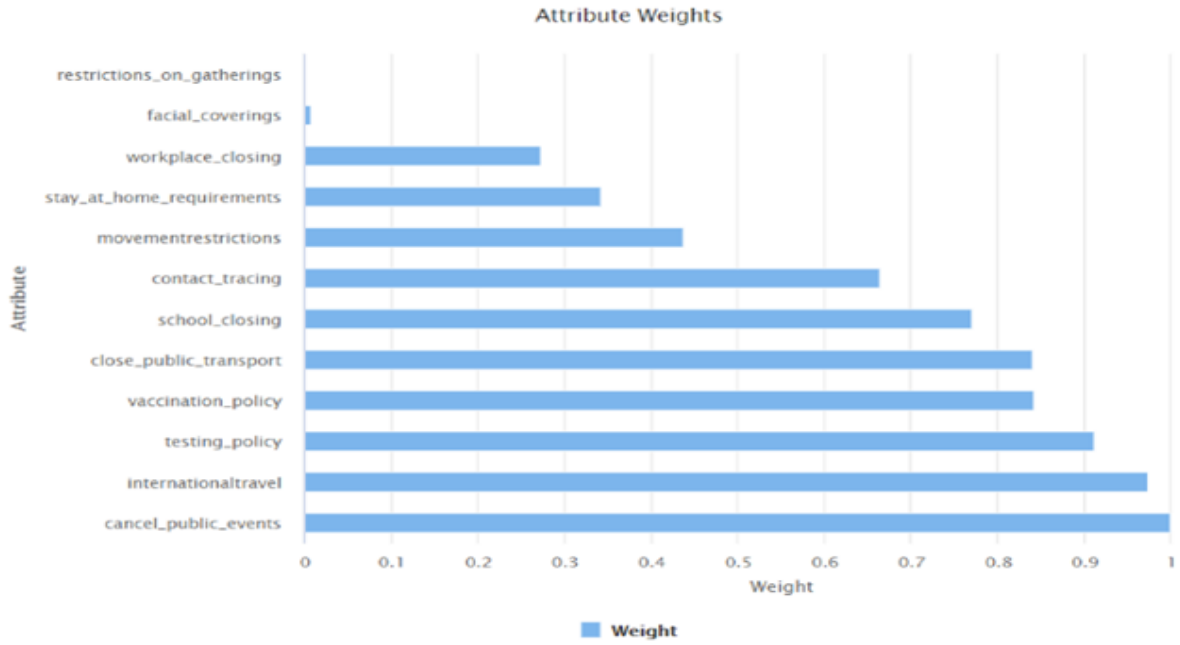

Figure 2. Attribute Weights

In figure 3, there is a correlation between attributes and the new cases attribute as labels, from the figure there is no significant difference, and it can be said that all the attributes are equally effected.

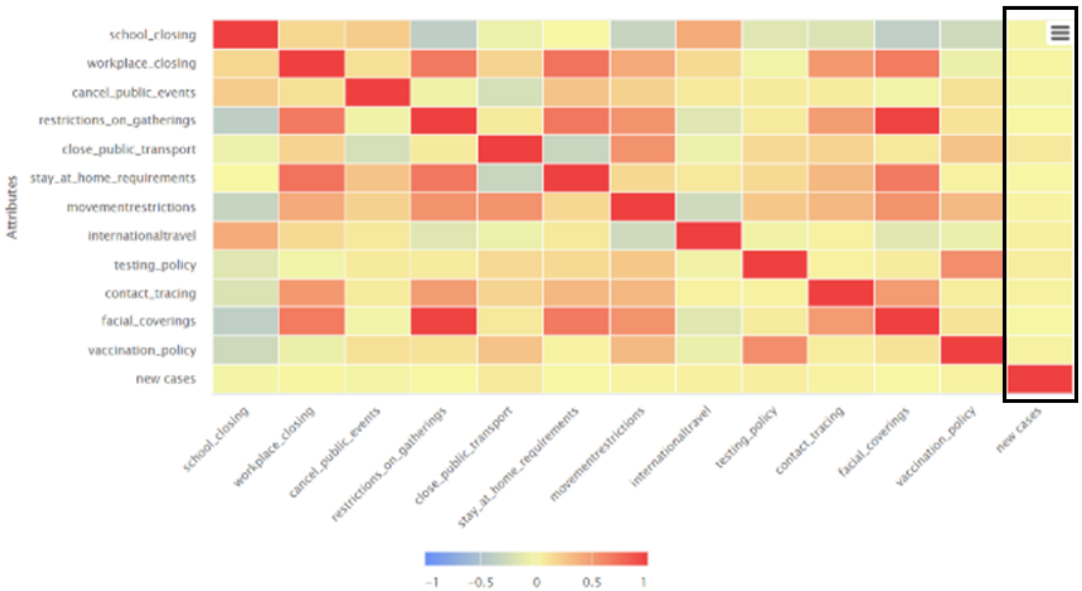

Figure 3. Matrix Correlation

In figure 4, the correlation of attributes between cancel public events and new cases is presented, there is no significant difference, but when applied the order to cancel the activity occurs a decrease in cases although not significant

.

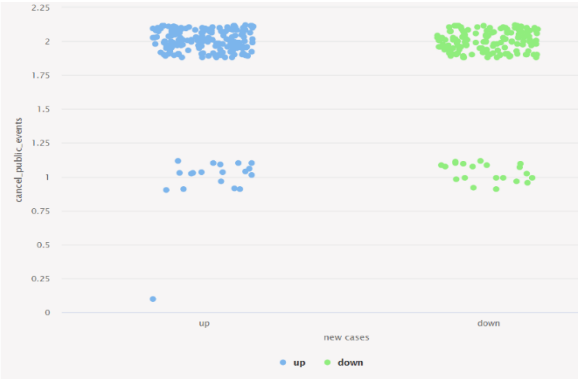

Figure 4. Attribute correlation between cancel public events and new cases

In figure 5 is the correlation between the attributes of international travel and new cases, where there is a ban on travel for foreigners, seen when only screening activities on arrival there is a rise in cases, and at the time of closing of arrivals for some countries, there is a decrease in cases although not significant.

Figure 5. Correlation of attributes between international travel and new cases

Figure 6 is the correlation of attributes between the restriction on gathering attribute and also the new cases attribute, there is no significant difference whether it is enforced the prohibition of gathering or not.

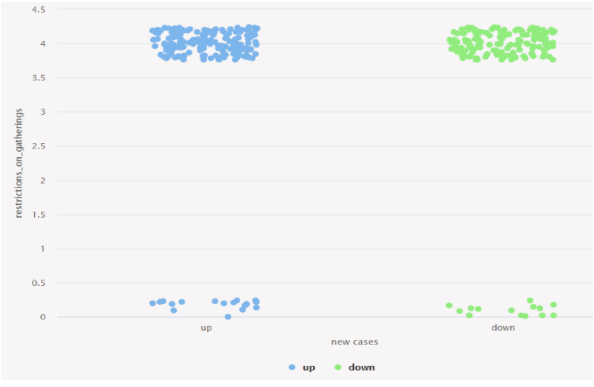

Figure 6. Correlation attributes between the restriction on gathering and new cases

Figure 7 is the correlation of attributes between school closing attribute and new cases attribute, there is no significant difference between school closures or not this is possible because not all schools in Indonesia are enforced closures, only in certain zones, such as red or black zone.

Figure 7. Correlation of attributes between school closing and new cases

In figure 8, there is a correlation between workplace closing and new cases attributes, seen when there is no closure, there is a rise in cases, and when work from home (WFH) is applied there is a decrease in cases.

Figure 8. Attribute correlation between workplace closing and new cases

Table 2 shows the results after the modeling process using the C4.5 algorithm as well as the DL, where there is better accuracy for the c4.5 algorithm than the DL of 53.25% with an AUC value of 0.969 and can be classified into excellent classification (Gorunescu, 2011). And both algorithms give the same result for positive class i.e. down. Small accuracy values are possible due to the large number of attributes used.

Table 2. Comparison of evaluation results

No	Algorithm	Accuracy	AUC	Positive Class
1	C4.5	53.25%	0.969	Down
2	Deep Learning	45.89%	0.495	Down

Based on table 2, and it is obtained that the best algorithm is C4.5, and based on that, the model to be used is the model obtained from the C4.5 modeling process as seen in figure 9.

```
testing_policy = 1
|   cancel_public_events = 0: up {up=1, down=0}
|   cancel_public_events = 1: down {up=18, down=20}
|   cancel_public_events = 2
|   |   facial_coverings = 0
|   |   |   close_public_transport = 0: up {up=8, down=6}
|   |   |   close_public_transport = 1
|   |   |   |   internationaltravel = 3: down {up=0, down=1}
|   |   |   |   internationaltravel = 4
|   |   |   |   |   movementrestrictions = 0: up {up=3, down=3}
|   |   |   |   |   movementrestrictions = 1: down {up=3, down=4}
|   |   facial_coverings = 1: up {up=1, down=0}
|   |   facial_coverings = 4
|   |   |   restrictions_on_gatherings = 0: up {up=1, down=0}
|   |   |   restrictions_on_gatherings = 4
|   |   |   |   close_public_transport = 0: up {up=68, down=49}
|   |   |   |   close_public_transport = 1
|   |   |   |   |   stay_at_home_requirements = 0: up {up=2, down=1}
|   |   |   |   |   stay_at_home_requirements = 1: up {up=39, down=34}
|   |   |   |   |   stay_at_home_requirements = 2
|   |   |   |   |   |   workplace_closing = 2: down {up=3, down=5}
|   |   |   |   |   |   workplace_closing = 3
|   |   |   |   |   |   |   internationaltravel = 3: up {up=16, down=12}
|   |   |   |   |   |   |   internationaltravel = 4
|   |   |   |   |   |   |   |   movementrestrictions = 1: up {up=16, down=15}
|   |   |   |   |   |   |   |   movementrestrictions = 2: down {up=4, down=5}
testing_policy = 3: down {up=6, down=9}
```

Figure 9. C4.5's tree

CONCLUSION

The conclusion based on the research that has been done, that the policies that have been done by the government in controlling cases of covid-19 can be said to be successful with the accuracy value obtained from the C4.5 algorithm is 53.25% with an AUC value of 0.969 that belongs to the excellent classification. And based on weighting on matrix correlation, a very influential attribute is the canceled public events and the least influential attribute is the restrictions on gathering.

REFERENCE

Amelia, Y., Eosina, P., & Setiawan, F. A. (2018). Perbandingan Metode Deep Learning Dan Machine Learning Untuk Klasifikasi (Uji Coba Pada Data Penyakit Kanker Payudara). *Seminar Nasional Teknologi Informasi*, *1*, 789–796.

Anung Ahadi Pradana, Casman, N. (2020). Pengaruh Kebijakan Social Distancing pada Wabah COVID-19 terhadap Kelompok Rentan di Indonesia. *Jurnal Kebijakan Kesehatan Indonesia : JKKI*, *9*(2), 61–67. Retrieved from https://jurnal.ugm.ac.id/jkki/article/view/55575

Bahri, S., Marisa Midyanti, D., Hidayati, R., Sistem Komputer, J., & Mipa, F. (2018). Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 24–31.

Benhar, H., Idri, A., & L Fernández-Alemán, J. (2020). *Data preprocessing for heart disease classification: A systematic literature review. Computer Methods and Programs in Biomedicine* (Vol. 195). https://doi.org/10.1016/j.cmpb.2020.105635

Gorunescu, F. (2011). *Data Mining Concepts, Models and Technique*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19721-5

Hijrah, Mukhlizar, M., & Pandria, T. M. A. (2020). Perbandingan Teknik Klasifikasi Untuk Memprediksi Kualitas Kinerja Karyawan. *Jurnal Optimalisasi*, *6*(1), 10–21. Retrieved from http://jurnal.utu.ac.id/joptimalisasi/article/view/1990

Ibrahim, D. (2017). Analisis Hubungan antar Faktor dan Komparasi Algoritma Klasifikasi pada Penentuan Penundaan Penerbangan. *Senit*, (September), 15–17.

Idri, A., Benhar, H., Fernández-Alemán, J. L., & Kadi, I. (2018). A systematic map of medical data preprocessing in knowledge discovery. *Computer Methods and Programs in Biomedicine*, *162*, 69–85. https://doi.org/10.1016/j.cmpb.2018.05.007

Lengkong, N. C., Safitri, O., Machsus, S., Putra, Y. R., Syahadati, A., & Nooraeni, R. (2021). Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg. *Jurnal Teknoinfo*, *15*(1), 20. https://doi.org/10.33365/jti.v15i1.866

Murphree, D. H., Puri, P., Shamim, H., Bezalel, S. A., Drage, L. A., Wang, M., … Comfere, N. (2020). Deep Learning for Dermatologists: Part I Fundamental Concepts. *Journal of the American Academy of Dermatology*. https://doi.org/10.1016/j.jaad.2020.05.056

Mutrofin, S., Machfud, M. M., Satyareni, D. H., Ginardi, R. V. H., & Fatichah, C. (2020). Komparasi Kinerja Algoritma C4.5, Gradient Boosting Trees, Random Forests, dan Deep Learning pada Kasus Educational Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(4), 807. https://doi.org/10.25126/jtiik.2020742665

Noviandi. (2018). Implementasi Algoritma Decision Tree C4.5 Untuk Prediksi Penyakit Diabetes. *Inohim*, *6*(1), 1–5.

Oxford University. (2021). Coronavirus Government Response Tracker | Blavatnik School of Government (ox.ac.uk). Retrieved

February 26, 2021, from https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker

Parhusip, H. A. (2020). Study on COVID-19 in the World and Indonesia Using Regression Model of SVM, Bayesian Ridge and Gaussian. *Jurnal Ilmiah Sains*, *20*(2), 49. https://doi.org/10.35799/jis.20.2.2020.28256

Rohman, A., Suhartono, V., & Supriyanto, C. (2017). Penerapan Agoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung. *Jurnal Teknologi Informasi*, *13*, 13–19.

Santosa, B., & Ardian, U. (2018). *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka.

Wahono, H., & Riana, D. (2020). Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5. *JURIKOM (Jurnal Riset Komputer)*, *7*(1), 7. https://doi.org/10.30865/jurikom.v7i1.1953