

ANALISIS DAN PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENERIMA VAKSINASI *COVID-19* DI KOTA BENGKULU

Pindo Putra Pratama¹, Harry Witriyono², Rozali Toyib³, Anisya Sonita⁴
Universitas Muhammadiyah Bengkulu^{1,2,3,4}, Bengkulu, Indonesia
Korespondensi : pindoputrapratama263@gmail.com

Abstrak- Fenomena wabah baru yang bernama Covid-19 merupakan pandemi yang meresahkan kehidupan dunia sehingga memberi tantangan besar dalam upaya peningkatan kesehatan masyarakat. Hingga sekarang vaksin untuk covid-19 terus didistribusikan keseluruh penjuru dunia bahkan sampai ke Indonesia. Di Kota Bengkulu sendiri proses vaksinisasi covid-19 terus dijalankan pemerintah. Dengan adanya pemberian vaksin tersebut maka akan dibutuhkan proses untuk pendistribusian vaksin, oleh karena itu dalam penelitian ini dibuatlah sistem klasifikasi pengelompokan untuk penerima vaksin. Permasalahan yang muncul dalam perencanaan pemberian vaksin covid-19 kepada masyarakat adalah tidak adanya kesadaran masyarakat untuk melakukan proses vaksin covid-19 dan masyarakat masih kurang pemahaman bahwa vaksin ini merupakan pencegahan penularan virus, oleh karena itu perlu dibuat pengelompokan atau penjadwalan vaksin kepada masyarakat sehingga nantinya vaksin ini dapat digunakan oleh masyarakat. Salah satu solusi untuk penerima vaksin ini adalah dengan teknik klasifikasi dengan datamining. Tujuan dari penelitian yang dilakukan ini adalah untuk mengetahui siapa saja yang berhak menerima vaksin terhadap masyarakat kota Bengkulu dan menerapkan aplikasi tools phyton untuk mengetahui hasil analisa cluster penerima vaksin covid-19. Algoritma C4.5 berhasil diterapkan pada Analisis dan Penerapan Algoritma C4.5 Untuk Klasifikasi Penerima Vaksinasi Covid-19 Di Kota Bengkulu, dengan memperoleh hasil tingkat akurasi sebesar 58%. Algoritma C4.5 adalah model klasifikasi yang baik dalam melakukan klasifikasi *data mining*. *Root* atau akar pohon keputusan yang diperoleh dari hitung manual diperoleh atribut status vaksin dengan nilai *gain ratio* 0,779284418.

Kata kunci : Vaksin, Covid-19, Algoritma C4.5, Klasifikasi, Phyton.

Abstract- The new outbreak phenomenon called Covid-19 is a pandemic that disturbs the life of the world so that it presents a major challenge in efforts to improve public health. Until now the vaccine for covid-19 continues to be distributed throughout the world even to Indonesia. In the city of Bengkulu itself the process of vaksinisasi covid-19 continues to be run by the government. With the provision of the vaccine, a process will be needed for the distribution of vaccines, therefore in this study a grouping classification system was created for vaccine recipients. The problem that arises in the planning of giving the covid-19 vaccine to the community is the absence of public awareness to carry out the covid-19 vaccine process and the public still lacks understanding that this vaccine is a prevention of virus transmission, therefore it is necessary to make grouping or scheduling vaccines to the community so that later this vaccine can be used by the community. One solution for recipients of this vaccine is with classification techniques with datamining. The purpose of this research is to find out who is entitled to receive vaccines against the people of Bengkulu city and apply phyton tools application to find out the results of the analysis of clusters of covid-19 vaccine recipients. C4.5 algorithm was successfully applied to the Analysis and Application of C4.5 Algorithm for classification of Covid-19 vaccination recipients in Bengkulu City, by obtaining an

accuracy rate of 58%. The C4.5 algorithm is a good classification model for classifying data mining. Root or root tree decisions obtained from manual calculations obtained attributes of vaccine status with a gain ratio value of 0.779284418. Keywords: System, DSS, MAUT, Assistance, Seeds

Keyword : *Vaccine, Covid-19, Algorithm C4.5, Classification, Phyton.*

1. Pendahuluan

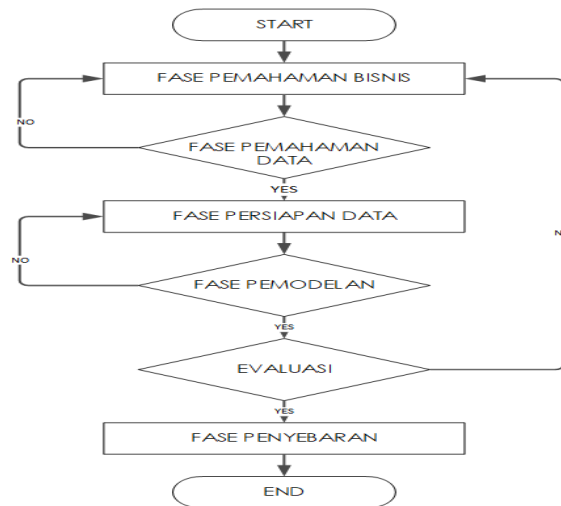
Virus baru yang menghebohkan seluruh warga dunia muncul pada awal tahun 2019, tepatnya pada tanggal 31 Desember 2019, Khusus pertama muncul di China, kasus virus yang belum diketahui penyebabnya ini terletak di Kota Wuhan, Provinsi Hubei. Kejadian virus baru ini diberi nama Covid-19 [1]. Sampai saat pandemi Covid-19 ini sangat meresahkan kehidupan dunia sehingga memberi tantangan besar bagi negara-negara yang terdampak virus ini. Berbagai macam upaya dilakukan untuk mengurangi penyebaran covid-19 ini. Salah satu cara untuk mengakhiri dampak dari covid-19 ini adalah dengan mencari vaksin yang tepat untuk mengurangi penyebarannya.

Pada akhir tahun 2020 banyak negara-negara yang mengembangkan vaksin covid-19. Penelitian calon vaksin Covid-19 membuahkan hasil positif. Sejumlah negara bersiap melakukan penyuntikan vaksin corona secara massal kepada warganya Di negara kita sendiri yaitu Indonesia, Presiden Joko Widodo (*Jokowi*) menyebutkan saat ini Indonesia sudah mendapatkan akses *vaksin Corona* sebagai salah satu cara untuk mengakhiri wabah Covid-19. Indonesia sudah mendapatkan akses sebanyak 290 juta vaksin COVID-19 Di indonesia sendiri sudah mendapatkan akses vaksin Covid-19 sehingga terjadilah permasalahan-permasalahan yang muncul dalam pemberian vaksin covid-19 kepada masyarakat. Masalah-masalah tersebut dapat kita lihat dari segi siapa saja yang menerima vaksin dan tidak adanya kesadaran masyarakat untuk melakukan proses vaksin covid-19 dan yang terakhir, masyarakat masih kurang pemahaman bahwa vaksin ini merupakan pencegahan penularan virus oleh karena itu dibuatlah pengelompokan vaksin kepada masyarakat sehingga nantinya vaksin ini dapat digunakan oleh masyarakat.

Salah satu solusi untuk penyebaran penerima vaksin ini adalah dengan teknik klasifikasi data mining. Dengan adanya klasifikasi pemberian vaksin ini diharapkan penyebaran penerima vaksin dapat tersebar merata kepada masyarakat Indonesia khususnya Kota Bengkulu agar dapat mengurangi dampak penyebaran Covid-19 Metode yang ditawarkan dalam penelitian ini adalah data mining dengan algoritma C4.5 yang merupakan proses menemukan korelasi data baru dan pola sekelompok data dalam jumlah yang besar dengan menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika Tujuan dari penelitian yang dilakukan ini adalah untuk mengetahui siapa saja yang berhak menerima vaksin terhadap masyarakat kota Bengkulu dan menerapkan aplikasi tools phyton untuk mengetahui hasil analisa cluster penerima vaksin covid-19 [2][3][4][5].

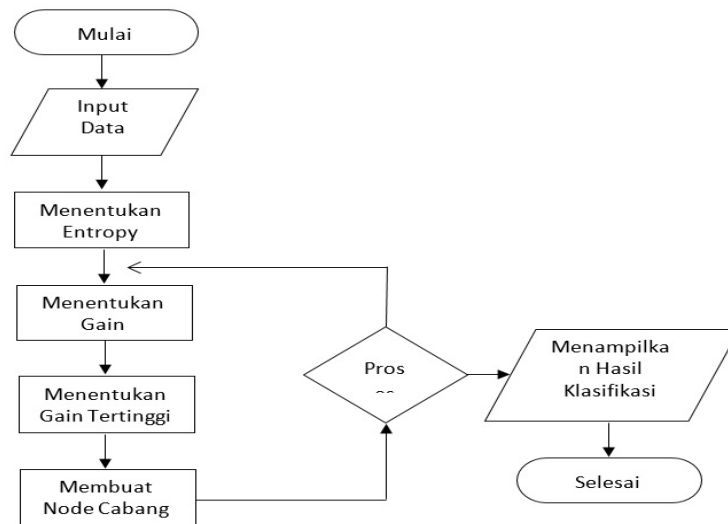
2. Metodologi Penelitian

Metode penelitian ini menggunakan teknik *data mining* yaitu *Cross Industry Standard Process Data Mining* (CRISP-DM) yang merupakan standar pada proses *data mining*, dalam metode CRIPS-DM terdapat enam fase yaitu *Business Understanding Phase* (Fase Pemahaman Bisnis), *Data Understanding Phase* (Fase Pemahaman Data), *Data Preparation* (Persiapan Data), *Modeling Phase* (Fase Pemodelan), *Evaluation Phase* (Fase Evaluasi), dan *Deployment Phase* (Fase Penyebaran) [6].



Gambar 1. Tahapan penelitian metode *CRISP-DM*

1. Fase pemahaman bisnis (*business understanding phase*)
Pada tahapan pemahaman bisnis ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan data mining
2. Fase pemahaman data (*data understanding phase*) .
Pada tahapan pemahaman data, beberapa hal yang perlu dilakukan antara lain, mengumpulkan data awal, mendeskripsikan data, mengeksplorasi data, dan memverifikasi kualitas data Mengumpulkan data, data yang digunakan merupakan data penerima vaksin covid-19 di kota Bengkulu.
3. Fase persiapan data (*data preparation phase*) Pada tahapan persiapan data, teknik persiapan data agar kualitas data diperoleh lebih baik. Pada tahap persiapan ini juga, mencari referensi jurnal yang berhubungan dengan algoritma C4.5. Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal, bisa mengatasi missing data, bisa mengatasi data kontinu dan pruning [7][8]. Serta informasi mengetahui tentang python sebagai system yang akan digunakan. Python merupakan sebuah bahasa pemrograman yang cukup terkenal yang memiliki banyak manfaat untuk mendukung pemrograman yang berorientasi objek dan dapat berjalan diberbagai macam platform sistem operasi seperti PCs, Macintosh [9].
4. Pemodelan Pada penelitian yang menggunakan Algoritma C4.5 dengan *operator Cross Validation* dan dengan metode pemilihan *split* menggunakan *gain ratio*. Untuk melakukan pengukuran dalam akurasi penelitian ini akan menggunakan *Phyton*. Pada fase pemodelan ini juga rancangan yang berhubungan dengan system nantinya akan dibuat, untuk mempermudah proses pembuatan system maka diperlukan kerangka atau flowchart dari algoritma C4.5 sebagai referensi dari pembuatan system. Berikut Flowchart Algoritma C4.5 :



Gambar 2. Flowchart Algoritma C4.5

5. Evaluasi (evaluation)

Pada tahap ini dilakukan evaluasi terhadap keefektifan dan keakuratan dari model yang terbentuk sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase pemahaman bisnis.

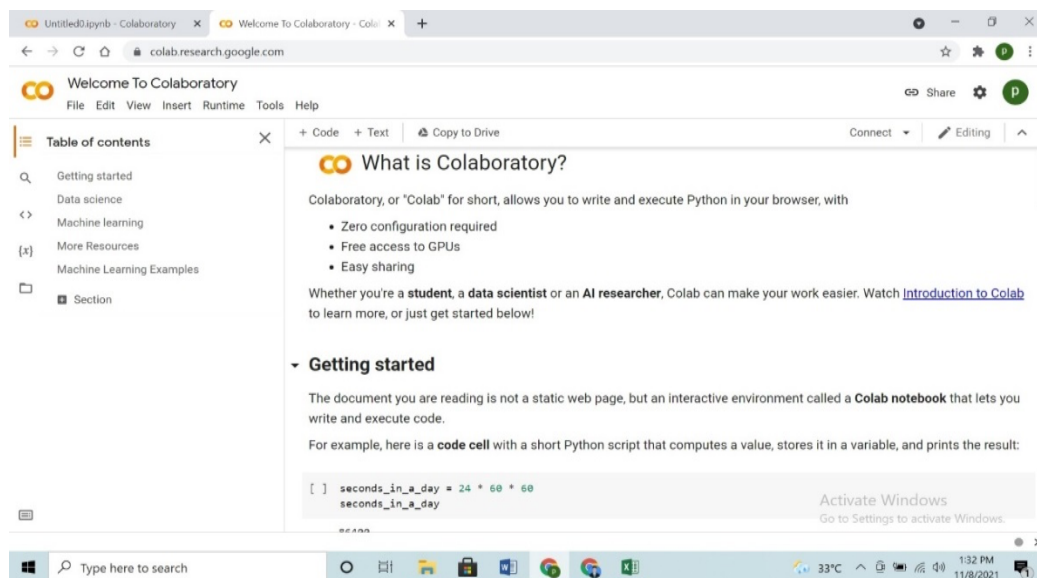
6. Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan dipresentasikan dalam bentuk laporan sehingga dapat diketahui atau dibaca oleh orang awam. Hasil dari penelitian ini berupa klasifikasi penerima vaksin covid-19 di kota Bengkulu.

3. Hasil Penelitian

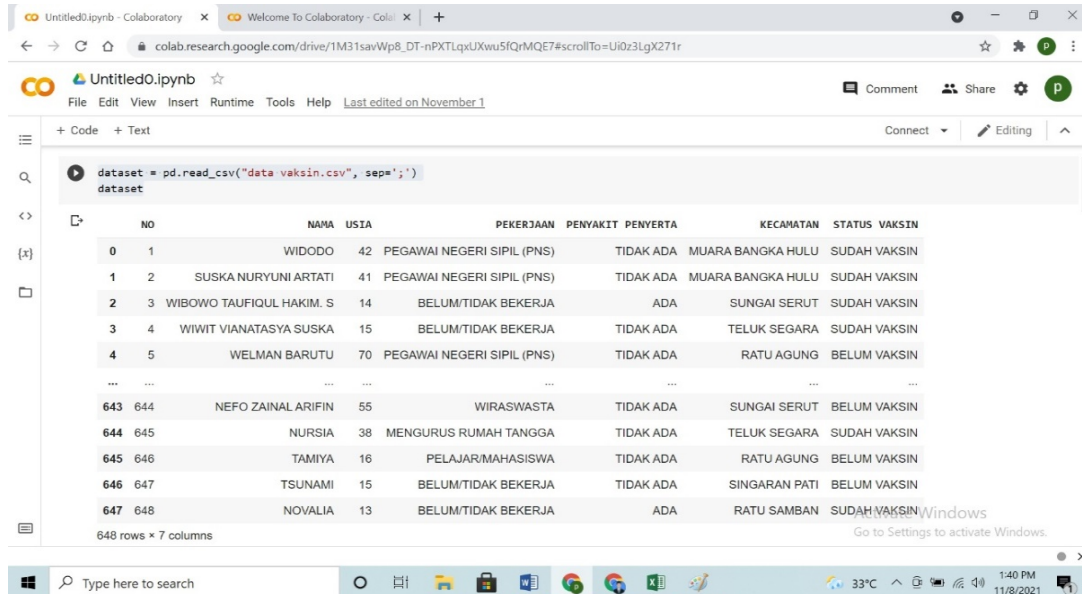
Pengelompokan Penerima Vaksinasi Covid-19 Di Kota Bengkulu Menggunakan Algoritma c4.5 . System yang dibuat menggunakan software resmi phyton yaitu google colab, Adapun tampilannya sebagai berikut:

Proses pertama mengakses halaman website



Gambar 3. Halaman Colab Phyton

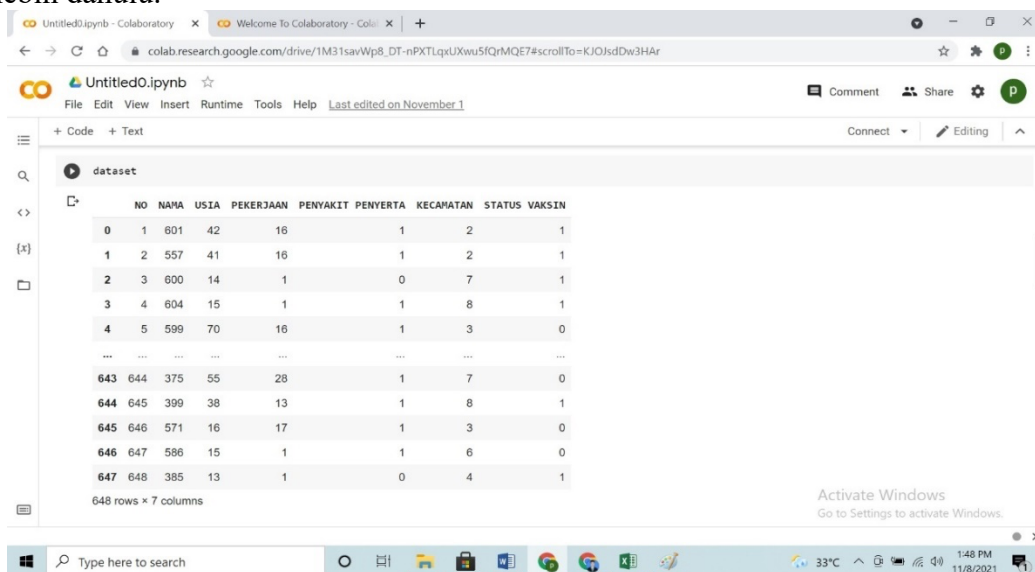
Dalam website ini merupakan proses untuk menganalisa penerima vaksin covid-19 di Kota Bengkulu. Pada colab ini nantinya akan diinputkan script untuk mengakses data penerima vaksin. Proses selanjutnya memanggil data vaksin yang diperoleh pada saat penelitian sebanyak 648 data set yang telah dibuat bentuk table excel.



Gambar 4. Form data penerima vaksin

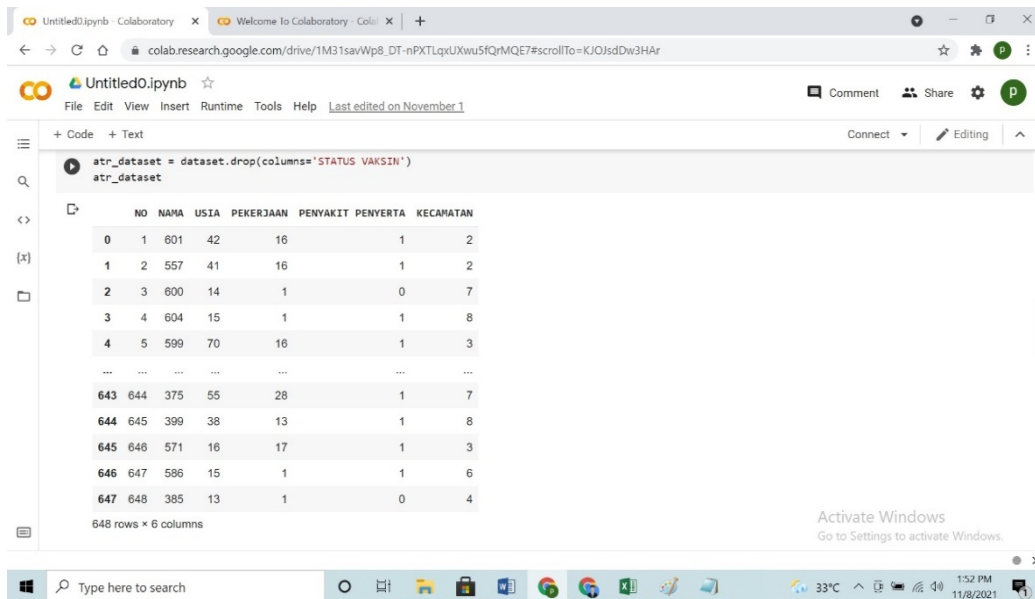
Proses selanjutnya, dengan menerapkan Algoritma C4.5 mengubah data set penerima vaksin tersebut.

Langkah awal yaitu menghitung jumlah kasus keseluruhan penerima vaksin Covid-19 berdasarkan atribut yang digunakan yaitu, Nama, Usia, Jenis Pekerjaan, Penyakit Penyerta, Kecamatan, Status Vaksin. Selanjutnya menghitung nilai *entropy*, kemudian lanjutkan dengan menghitung nilai *entropy* dan *gain* tertinggi pada setiap nilai-nilai dalam variabel. Pada penelitian ini menggunakan *gain ratio* makan pertama mencari nilai *gain ratio* sesuai dengan alur algoritma C4.5 yaitu menentukan *entropy* dan *gain* terlebih dahulu.



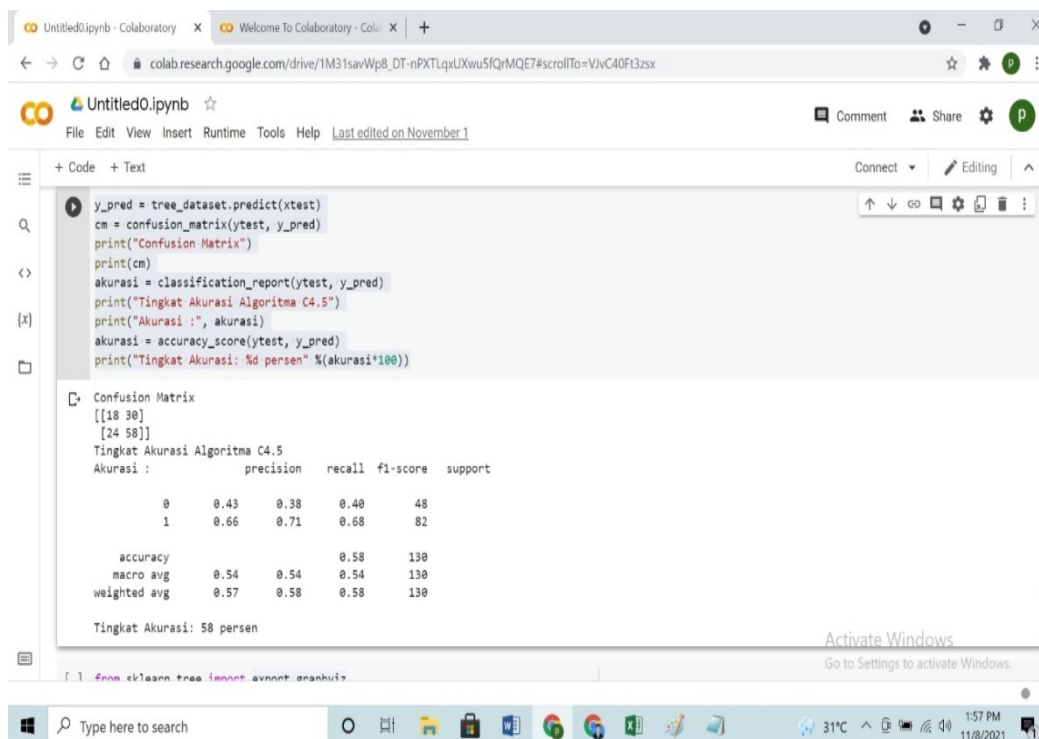
Gambar 5. Dataset Penerima Vaksin

Selanjutnya data tersebut diubah kedalam bentuk class untuk klasifikasi.



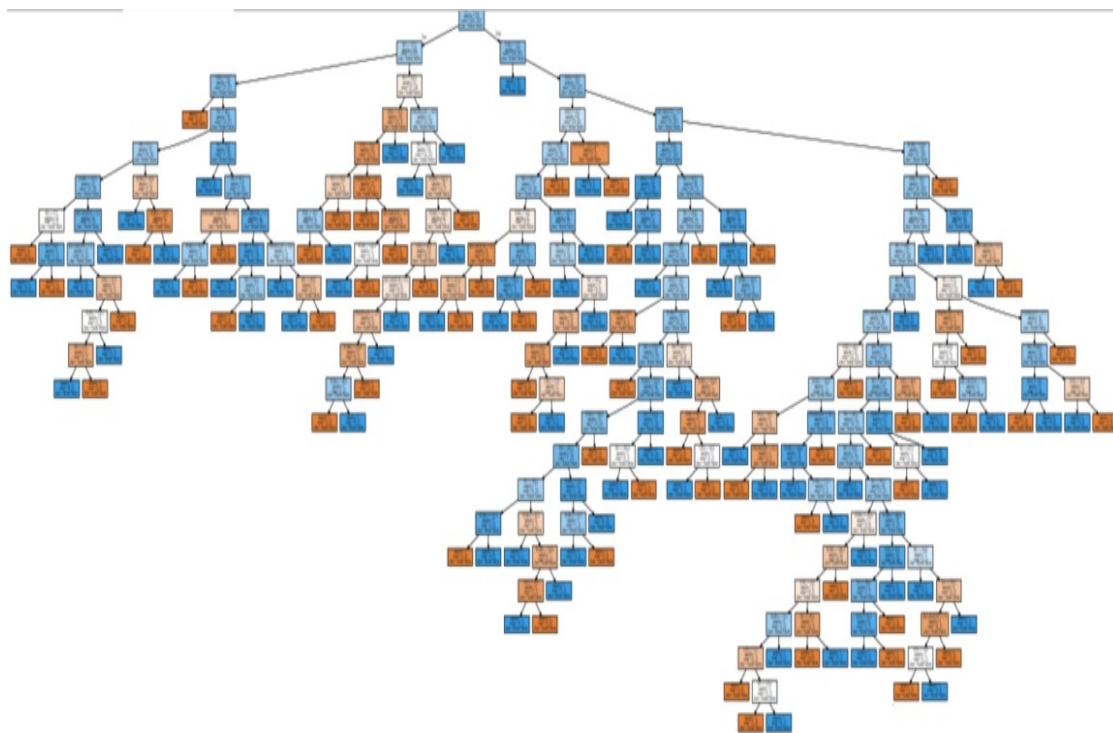
Gambar 6. Data class Status Vaksin

Setelah class data didapat proses selanjutnya adalah menghitung tingkat akurasi menggunakan confusion matrix.



Gambar 7. Tingkat akurasi

Untuk nilai tingkat akurasi yang di dapat yaitu sebesar 58%. Selanjutnya mendapatkan Diagram Pohon.



Gambar 8. Pohon Keputusan Hasil Analisa Penerima Vaksin

Root atau akar pohon keputusan, yang memiliki nilai *gain ratio* tertinggi adalah variable Status Vaksin dengan nilai *gain ratio* 0,779284418 dan dijadikan *root*. Dalam atribut Status Vaksin memiliki nilai yaitu, “Belum Vaksin”, “Sudah Vaksin,” dapat dilihat dari hasil perhitungan bahwa untuk atribut Status Vaksin dengan nilai “Belum

Vaksin” dan “Sudah Vaksin” masing-masing nilai seluruhnya memiliki 6 kategori rawan sehingga dapat disimpulkan kedua nilai tersebut menjadi daun atau *leaf*

Dari keputusan diatas dapat dijelaskan bahwa parameter dari klasifikasi penerima vaksin covid-19 di Kota Bengkulu terdiri dari berbagai atribut, mulai dari nama, Usia, pekerjaan, penyakit penyerta, kecamatan dan status vaksin. Dalam analisa ini atribut dengan nilai tertinggi dikategorikan gain untuk pohon keputusan. Gain tertinggi tersebut terdapat pada atribut status vaksin. Variabel dari status vaksin tersebut adalah sudah vaksin dan belum vaksin. Pohon keputusan diatas terdapat atribut belum vaksin dengan total data 518 orang yang belum vaksin dari 648 data dengan nilai value 162, 356. Diikuti dengan atribut belum vaksin lainnya dan atribut sudah vaksin.

4. Kesimpulan

Dari hasil penelitian mengenai “Analisis dan Penerapan Algoritma C4.5 Untuk Klasifikasi Penerima Vaksinasi Covid-19 Di Kota Bengkulu” maka dapat disimpulkan bahwa Algoritma C4.5 berhasil diterapkan pada Analisis dan Penerapan Algoritma C4.5. Klasifikasi Penerima Vaksinasi Covid-19 Di Kota Bengkulu, dengan memperoleh hasil tingkat akurasi sebesar 58%. *Root* atau akar pohon keputusan yang diperoleh dari hitung manual diperoleh atribut status vaksin dengan nilai *gain ratio* 0,779284418. Dari hasil analisa penerima vaksin terdapat atribut belum vaksin dengan total data 518 dengan nilai 162,356 dan di ikuti dengan atribut sudah vaksin. Dengan adanya sistem ini warga kota Bengkulu dapat melihat siapa saja yang berhak untuk menerima vaksin berdasarkan atribut status vaksin dan Dengan adanya klasifikasi pemberian vaksin ini diharapkan

penyebaran penerima vaksin dapat tersebar merata kepada masyarakat Kota Bengkulu agar dapat mengurangi dampak penyebaran Covid-19.

5. Daftar Pustaka

- [1] M. Johnson, "Wuhan 2019 Novel Coronavirus - 2019-nCoV," *Mater. Methods*, vol. 10, no. January, pp. 1–5, 2020, doi: 10.13070/mm.en.10.2867.
- [2] H. A. Diah Handayani, Dwi Rendra Hadi, Fathiyah Isbaniah, Erlina Burhan, "Penyakit Virus Corona 2019," *RESPILOGI Indones.*, vol. 40, no. 2, p. 128, 2020.
- [3] A. Wikanto, "Vaksin Corona Ditemukan, 4 Negara Ini Segera Vaksinasi Massal," *Internasional Kontan.co.id*. 2021, [Online]. Available: <https://internasional.kontan.co.id/news/vaksin-corona-ditemukan-4-negara-ini-segera-vaksinasi-massal>.
- [4] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, p. 108, 2016, doi: 10.31154/cogito.v2i2.19.108-119.
- [5] D. Bayu Febriyanto, L. Handoko, and H. Aisyah, "Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kepuasan Pembeli Online Shop," *J. Ris. Komput.*, vol. 5, no. 6, pp. 569–575, 2018, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C569>.
- [6] A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 260–270, 2015, doi: 10.28932/jutisi.v1i3.406.
- [7] R. S. Putra, "Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 Kota Pagar Alam," *Jukomika*, vol. 4, no. 1, pp. 23–35, 2021.
- [8] S. Turnip and P. Siltionga, "Analisis Pola Penyebaran Penyakit dengan Menggunakan Algoritma C4.5," *J. Tek. Inform. Unika St. Thomas*, vol. 03, no. 479, pp. 3–7, 2018.
- [9] V. S. Ginting, K. Kusri, and E. T. Luthfi, "Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 1–6, 2020, doi: 10.36294/jurti.v4i1.1101.