



UNIVERSITAS  
**MERCU BUANA**

**ALGORITMA *DECISION TREE* UNTUK MENGLASIFIKASI DATA  
JENIS VAKSIN *COVID-19* DI WILAYAH KELURAHAN GROGOL  
UTARA**

*TUGAS AKHIR*

Viky Nurhidayanti  
41518110139

UNIVERSITAS  
**MERCU BUANA**  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022



**ALGORITMA *DECISION TREE* UNTUK MENGLASIFIKASI DATA  
JENIS VAKSIN *COVID-19* DI WILAYAH KELURAHAN GROGOL  
UTARA**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Viky Nurhidayanti

41518110139

UNIVERSITAS  
MERCU BUANA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022

## LEMBAR PERNYATAAN ORISINALITAS

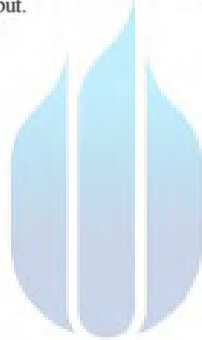
Yang bertanda tangan dibawah ini:

NIM : 41518110139

Nama : Viky Nurhidayanti

Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis  
Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 27 Juli 2022



Viky Nurhidayanti

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Viky Nurhidayanti  
NIM : 41518110139  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 27 Juli 2022

UNIVERSITAS  
MERCU BUANA



METERAI TEMPEL  
B3CPCAIX980428210

Viky Nurhidayanti

### SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Viky Nurhidayanti  
NIM : 41518110139  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: JEPIN (Jurnal Edukasi dan Penelitian Informatika)	
	ISSN	: eISSN : 2548-9364 / pISSN : 2460-0741	
	Link Jurnal	: <a href="https://jurnal.untan.ac.id/index.php/jepin/index">https://jurnal.untan.ac.id/index.php/jepin/index</a>	
	Link File Jurnal Jika Sudah di Publish	:	

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS  
MERCU BUANA

Jakarta, 27 Juli 2022



Viky Nurhidayanti

**LEMBAR PERSETUJUAN PENGUJI**

NIM : 41518110139  
Nama : Viky Nurhidayanti  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi  
Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan  
Grogol Utara

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



(Vina Ayumi, S.Kom., M.Kom)

UNIVERSITAS  
MERCU BUANA

#### LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110139  
Nama : Viky Nurhidayanti  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi  
Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan  
Grogol Utara

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



(Muhammad Rifqi, S.Kom, M.Kom)

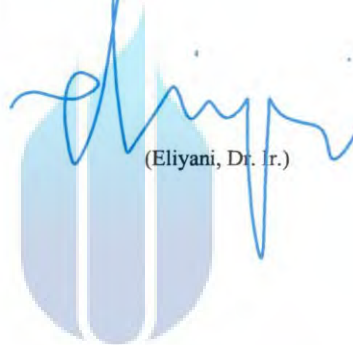
UNIVERSITAS  
MERCU BUANA

**LEMBAR PERSETUJUAN PENGUJI**

NIM : 41518110139  
Nama : Viky Nurhidayanti  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi  
Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan  
Grogol Utara

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022



(Eliyani, Dr. Ir.)

UNIVERSITAS  
MERCU BUANA




## LEMBAR PENGESAHAN

NIM : 41518110139  
Nama : Viky Nurhidayanti  
Judul Tugas Akhir : Algoritma *Decision Tree* Untuk Mengklasifikasi Data  
Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol  
Utara

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 27 Juli 2022

a.n. Menyetujui,

  
(Leonard Goeirmanto, Dr., ST,M.Sc)  
Dosen Pembimbing

Mengetahui,

  
(Wawan Gunawan, S.Kom, MT)  
Koord. Tugas Akhir Teknik Informatika

  
(Ir. Emil R. Kaburuan, Ph.D., IPM.)  
Ka. Prodi Teknik Informatika

## ABSTRAK

Nama : Viky Nurhidayanti  
NIM : 41518110139  
Pembimbing TA : Leonard Goeirmanto, Dr., ST,M.Sc  
Judul : Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara

Indonesia merupakan salah satu negara terdampak *Covid-19* selama dua tahun belakangan ini. *Covid-19* merupakan sebuah wabah yang menyerang hampir seluruh negara di dunia dan banyak memakan korban jiwa. Langkah pemerintah Indonesia untuk memerangi wabah *Covid-19* ini yaitu dengan mencanangkan program vaksinasi secara bertahap dan dimulai dari kategori masyarakat lansia. Program vaksinasi dilakukan secara dinamis di setiap kelurahan-kelurahan salah satunya di kelurahan Grogol Utara yang seringkali mengalami kendala dalam memprediksi jenis vaksin yang akan digunakan selanjutnya. Pada penelitian ini, Penulis menerapkan algoritma *Decision Tree C4.5* untuk mengklasifikasikan data berdasarkan jenis vaksin yang telah digunakan dengan cara membagi data menjadi dua jenis yaitu data training dan data testing. Data training merupakan data uji coba yang nantinya akan menjadi pembelajaran bagi data testing dalam menentukan hasil akhir yang ditentukan. Hasil dari penelitian menggunakan Algoritma *Decision Tree C4.5* mendapatkan nilai akurasi sebesar 99.23% dihitung menggunakan *gain ratio*. Kemudian hasil prediksi berdasarkan pohon keputusan yang terbentuk yaitu *COVID-19 Vaccine Pfizer* dapat dijadikan klasifikasi jenis vaksin yang dapat digunakan pada masa yang akan datang jika dilihat dari tanggal pendaftarannya.

Kata kunci:

Vaksinasi Covid-19, Klasifikasi, Algoritma *Decision Tree C4.5*

## ABSTRACT

Name : Viky Nurhidayanti  
Student Number : 41518110139  
Counsellor : Leonard Goeirmanto, Dr., ST,M.Sc  
Title : Algoritma *Decision Tree* Untuk Mengklasifikasi  
Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan  
Grogol Utara

Indonesia is one of the most affected countries by Covid-19 for the past two years. Covid-19 is an epidemic that has attacked almost all countries in the world and has claimed many lives. The Indonesian government movement to fight this Covid-19 plague is by launching a vaccination program in stages and starting from the category of the elderly. Vaccination program was dynamic in Grogol Utara urban village which often experienced problems in predicting the type of vaccine to be used. In this research, the author applied the *Decision Tree* C4.5 algorithm to classify data based on the type of vaccine that has been used by dividing data into two types of training data and testing data. Training data is a test data that will be a test data in determining the final results. The result of this research using *Decision Tree* C4.5 Algorithm get an accuracy value of 99.23% calculated using gain ratio. Then the prediction result based on the *Decision Tree* is COVID-19 Vaccine Pfizer can be used as a classification of vaccine types in the future based on the date registration

Key words:

Covid-19 Vaccinated, Classification, *Decision Tree* C4.5 Algorithm

UNIVERSITAS  
MERCU BUANA

## KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah Subhanahu Wa Ta'ala, atas berkat, rahmat dan karunia-Nya penulis dapat menyelesaikan penyusunan Tugas Akhir yang berjudul “Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara “. Penyusunan Tugas Akhir ini disusun dalam rangka memenuhi penilaian salah satu mata kuliah wajib dan sebagai syarat akademik untuk memperoleh gelar sarjana Ilmu Komputer di Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak yang telah mendukung penulis baik secara moral maupun material. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Leonard Goeiranto, Dr., ST,M.Sc selaku dosen pembimbing Tugas Akhir serta dosen pembimbing akademik, atas kesabaran, nasehat serta waktunya dalam membimbing dari awal penulisan hingga penulis dapat menyelesaikan penulisan tugas akhir ini. Segala kebaikan yang telah diberikan oleh Bapak Leonard akan selalu terkenang oleh penulis.
2. Allah Subhanahu Wa Ta'ala yang telah memberikan rahmat, nikmat serta hidayah-Nya sehingga penulis mampu menyelesaikan laporan jurnal serta tugas akhir ini.
3. Kedua orang tua tercinta yang tidak hentinya memberikan doa dan dukungan kepada penulis mulai dari penulis lahir sampai detik ini mampu menyelesaikan laporan jurnal dan tugas akhir.
4. Bapak Ir. Emil Robert Kaburuan, Ph.D., IPM selaku Kepala Program Studi Teknik Informatika Universitas Mercu Buana
5. Bapak Ibu Dosen Fakultas Ilmu Komputer Universitas Mercu Buana yang telah memberikan pelajaran serta ilmu yang bermanfaat selama penulis menempuh perkuliahan.
6. Kepada Kelurahan Grogol Utara dan Puskesmas Kelurahan Grogol Utara II yang telah memberikan kesempatan bagi penulis untuk melakukan

penelitian serta memperoleh data Vaksinasi *Covid-19* se-Kelurahan Grogol Utara.

Akhir kata, penulis berharap agar penulisan tugas akhir ini dapat bermanfaat bagi kita semua, penulis menyadari bahwa penyusunan tugas akhir ini jauh dari kata sempurna. Untuk itu penulis mengharapkan saran dan kritik yang membangun untuk penulisan tugas akhir ini untuk perbaikan yang akan datang

Jakarta, 5 Juli 2022

Penulis



## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ...	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	v
LEMBAR PENGESAHAN.....	viii
ABSTRAK.....	ix
ABSTRACT .....	x
KATA PENGANTAR .....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL.....	1
KERTAS KERJA .....	2
BAB 1. LITERATUR REVIEW .....	3
BAB 2. ANALISIS DAN PERANCANGAN.....	11
BAB 3. SOURCE CODE.....	13
BAB 4. DATASET .....	17
BAB 5. TAHAPAN EKSPERIMEN .....	18
BAB 6. HASIL SEMUA EKSPERIMEN .....	20
DAFTAR PUSTAKA .....	27
LAMPIRAN DOKUMEN HAKI .....	29
LAMPIRAN KORESPONDENSI .....	31

## NASKAH JURNAL

# Algoritma Decision Tree Untuk Mengklasifikasikan Data Jenis Vaksinasi Covid-19 Di Wilayah Kelurahan Grogol Utara

Viky Nurhidayanti

*Teknik Informatika, Universitas Mercu Buana  
JL. Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650  
41518110139.student@mercubuana.ac.id*

**Abstrak**— Indonesia merupakan salah satu negara terdampak *Covid-19* selama dua tahun belakangan ini. *Covid-19* merupakan sebuah wabah yang menyerang hampir seluruh negara di dunia dan banyak memakan korban jiwa. Langkah pemerintah Indonesia untuk memerangi wabah *Covid-19* ini yaitu dengan mencanangkan program vaksinasi secara bertahap dan dimulai dari kategori masyarakat lansia. Program vaksinasi dilakukan secara dinamis di setiap kelurahan-kelurahan salah satunya di kelurahan Grogol Utara yang seringkali mengalami kendala dalam memprediksi jenis vaksin yang akan digunakan selanjutnya. Pada penelitian ini, Penulis menerapkan algoritma *Decision Tree C4.5* untuk mengklasifikasikan data berdasarkan jenis vaksin yang telah digunakan dengan cara membagi data kedalam dua kategori yaitu training data dan testing data. Training data merupakan kumpulan data uji coba yang nantinya akan menjadi pembelajaran bagi data testing dalam menentukan hasil akhir yang ditentukan. Hasil dari penelitian menggunakan *Algoritma Decision Tree C4.5* mendapatkan nilai akurasi sebesar 52% dihitung menggunakan *gain ratio*. Kemudian hasil prediksi berdasarkan pohon keputusan yang terbentuk yaitu *COVID-19 Vaccine Pfizer* dapat dijadikan klasifikasi jenis vaksin yang dapat digunakan pada masa yang akan datang jika dilihat dari tanggal pendaftarannya.

**Abstract**— Indonesia is one of the most affected countries by *Covid-19* for the past two years. *Covid-19* is an epidemic that has attacked almost all countries in the world and has claimed many lives. The Indonesian government movement to fight this *Covid-19* plague is by launching a vaccination program in stages and starting from the category of the elderly. Vaccination program was dynamic in Grogol Utara urban village which often experienced problems in predicting the type of vaccine to be use. In this research, the author applied the *Decision Tree C4.5* algorithm to classify data based on the type of vaccine that has been used by dividing data into two types of training data and testing data. Training data is a test data that will be a test data in determining the final results. The result of this research using *Decision Tree C4.5* Algorithm get an accuracy value of 52% calculated using *gain ratio*. Then the prediction result based on the decision tree is *COVID-19 Vaccine Pfizer* can be used as a

classification of vaccine types in the future based on the date registration.

**Keywords**— Vaksinasi Covid-19, Klasifikasi, Algoritma Decision Tree C4.5

## I. PENDAHULUAN

Indonesia menjadi salah satu negara terdampak *Covid-19* yang mewabah di seluruh dunia hingga merenggut jutaan jiwa. Berawal dari Wuhan, China, virus ini menjadi trending topik dunia pada akhir 2019 lalu karena penyebarannya yang sangat cepat dan hampir seluruh negara di dunia terpapar virus tersebut. Pemerintah mengambil langkah untuk memulai program vaksinasi *Covid-19* di seluruh wilayah Indonesia. Terdapat beberapa jenis vaksin yang digunakan yaitu *Sinovac, Astra Zaneqa, Moderna dan Pfizer*[1]. Program vaksinasi *Covid-19* dilakukan dengan tujuan untuk menciptakan *herd immunity* (kekebalan kelompok) terhadap penyakit agar masyarakat dapat lebih produktif dalam melakukan kegiatannya.

Program kegiatan vaksinasi *Covid-19* di Indonesia dilaksanakan secara bertahap sejak Desember 2020 dan dilakukan secara dinamis di setiap Kelurahan maupun kecamatan salah satunya yaitu di Kelurahan Grogol Utara. Saat ini kegiatan vaksinasi *Covid-19* telah sampai pada tahap ke-4 yang merujuk pada masyarakat umum, remaja dan anak-anak. Dilihat dari akhir 2020 Kelurahan Grogol Utara telah melaksanakan proses vaksin bagi tenaga Kesehatan menggunakan jenis vaksin *Sinovac*, jika dilihat dari jumlah keseluruhan masyarakat umum yang ada di wilayah Kelurahan Grogol Utara (32.083 jiwa) dari total jumlah penduduk se-kecamatan yang berjumlah 309.500 jiwa pada tahun 2019[2].

Proses vaksinasi yang dilakukan secara dinamis memang menjadi salah satu cara baik untuk mengurangi keramaian antrian dan menjaga protokol kesehatan yang dicanangkan oleh pemerintah Indonesia. Namun, hal ini menimbulkan kendala-



kendala dalam mengetahui jenis vaksin yang akan digunakan pada dosis selanjutnya.

Berdasarkan permasalahan yang timbul dalam kasus ini, diperlukan solusi yang tepat untuk mengetahui perkembangan proses vaksinasi *Covid-19*. *Data mining* merupakan solusi yang tepat untuk menyelesaikan permasalahan dalam pengelompokan data. *Data mining* adalah proses untuk mengolah data untuk menghasilkan sebuah pola dari data yang dikumpulkan dengan menerapkan metode klasifikasi[3]. Klasifikasi merupakan proses untuk menemukan pola yang nantinya akan mengelompokkan *class* data hingga mendapatkan prediksi *class* objek yang tidak diketahui[4]. Teknik klasifikasi termasuk dalam teknik *data mining* yang sederhana, yaitu dengan melakukan penyusunan data kedalam kategori agar mudah digunakan. Dalam teknik klasifikasi terdapat beberapa algoritma yang bisa diterapkan diantaranya adalah *Decision Tree*, *K-NN*, *Super Vector Machine*, *Naïve Bayes* dan masih banyak lagi. Salah satu algoritma yang dapat diimplementasikan pada teknik klasifikasi ini adalah algoritma *Decision Tree* yang memiliki struktur seperti pohon, yaitu tiap-tiap simpul menggambarkan atribut, cabangnya menggambarkan hasil tes dan bagian daun dapat dianggap sebagai kelas atau distribusi kelas[5].

Berdasarkan penelitian yang telah dilakukan sebelumnya oleh Syah, Rama Dian[5] yang mengklasifikasikan hasil seleksi kompetensi dasar pada CPNS 2019 di Arsip Nasional Republik Indonesia. Hasil penelitian yang dilakukan pada 344 data peserta SKD yang terbagi menjadi 4 kelas dengan nilai akurasi sebesar 92,23%.

Penelitian lain terkait yang diteliti oleh Fida Maisa[6] yang memprediksi deteksi awal penyakit diabetes untuk menurunkan tingkat komplikasi diabetes di waktu yang akan datang. Hasil penelitian yang dilakukan pada 512 data masyarakat penderita diabetes, didapatkan hasil akurasi sebesar 97,12%, Precision sebesar 93,02% dan Recall sebesar 100,00%.

Pada penelitian ini, Teknik klasifikasi digunakan sebagai sebuah alat untuk memprediksi seseorang mendapatkan jenis vaksin apa yang tepat digunakan untuk tahap selanjutnya. Algoritma yang digunakan adalah *Decision Tree*.

## II. LANDASAN TEORI

### A. *Covid-19*

*Coronavirus Disease-19* atau dikenal dengan virus *Covid-19* adalah penyakit menular yang disebabkan oleh virus *Corona* yang menyebar melalui kontak langsung dari orang ke orang. *Coronavirus* adalah jenis virus yang masih satu turunan dengan *influenza* namun efek yang ditimbulkan dari *coronavirus* lebih besar jika dibanding dengan penyakit *influenza* lainnya. Virus ini dapat menular pada manusia melalui *droplet* pada saat manusia berbicara, batuk maupun bersin[7].

### B. Vaksinasi

Vaksinasi adalah pemberian dosis vaksin sesuai dengan ketentuan yang telah ditetapkan yang khusus diberikan dalam

rangka pembentukan imunitas seseorang terhadap penyakit. Vaksinasi bertujuan untuk melindungi dan memperkuat system perawatan kesehatan serta mempertahankan produktivitas guna mengurangi dampak sosial dan ekonomi terhadap masyarakat. Terdapat beberapa kelompok prioritas untuk mendapatkan vaksin yaitu: tenaga kesehatan, pelayan publik, lansia dan masyarakat umum[8]

### C. *Data Mining*

*Data mining* adalah suatu cara atau metode untuk mengolah informasi dalam jumlah besar atau sering disebut dengan *big data* dengan menerapkan analisis statistik, matematika dan kecerdasan buatan. Teknik pada *data mining* digunakan sebagai alat pemeriksa *big data* untuk menemukan pola terbaru dan pola tersebut dapat bermanfaat untuk masa depan. Teknik-teknik penambangan data digunakan untuk meningkatkan kemampuan sistem pencarian informasi.

Dalam proses pengolahan data dengan menerapkan *data mining* diperlukan adanya fokus terhadap metode yang digunakan. *Data mining* memiliki kategori kelompok berdasarkan metode atau tugas yang dilakukan, yaitu:

- a. Deskripsi  
Cara menggambarkan pola dalam suatu data untuk menghasilkan penjelasan mengenai pola atau kecenderungan yang ada.
- b. Estimasi  
Estimasi adalah contoh yang dikembangkan menggunakan kumpulan data lengkap yang mengembalikan nilai berdasarkan variable target dan merupakan nilai prediksi.
- c. Prediksi  
Prediksi adalah desain model untuk menganalisa data dan memberikan hasil prediksi yang akan datang.
- d. Klasifikasi  
Klasifikasi adalah teknik pengkategorian data dengan menggunakan kategori/target/kelas untuk menentukan hasil dari proses klasifikasi. Pada klasifikasi terdapat set data berlabel serta atribut untuk menentukan proses pembelajaran (*training*) terlebih dahulu.
- e. Klasterisasi  
Klasterisasi adalah teknik untuk menemukan kelompok atau cluster pada sebuah *dataset* yang memiliki kemiripan satu sama lain dalam satu *cluster* dan berbeda dengan *cluster* lain. Teknik *cluster* menggunakan pemodelan yang tidak ditentukan oleh *label dataset*. Teknik *cluster* dapat melakukan *summary generation*, *outlier detection*, menemukan duplikasi data dan diawali dengan *pre-processing*.
- f. Asosiasi  
Asosiasi adalah teknik *data mining* dalam pemberian informasi hubungan antar *atribut* data dalam *database*. Teknik asosiasi lebih cocok digunakan dalam proses penjualan untuk menentukan pola ketergantungan satu variabel terhadap variabel lain untuk mengambil tindakan bisnis yang sesuai.



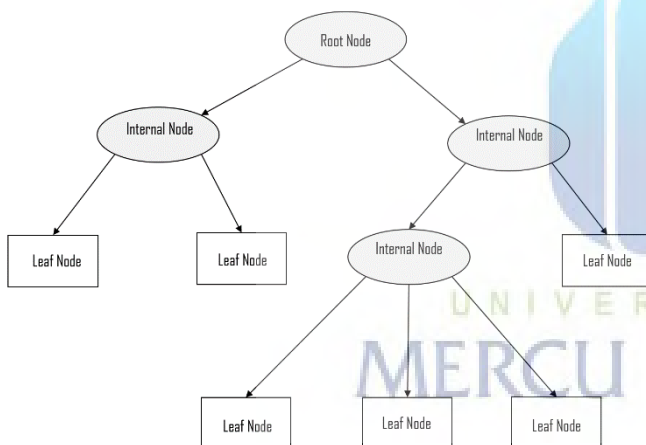
#### D. Klasifikasi

Klasifikasi merupakan sebuah proses untuk membangun data yang nilai kelas atau labelnya telah diketahui dan termasuk dalam *supervised learning*. Teknik klasifikasi termasuk pada teknik pengujian model yang mengukur sebuah akurasi prediksi[9]. Klasifikasi dapat didefinisikan sebagai proses pembelajaran model yang memetakan setiap kumpulan attribute input ke salah satu *class* yang telah didefinisikan sebelumnya[10].

#### E. Algoritma Decision Tree C4.5

*Decision Tree* adalah proses klasifikasi yang menggunakan metode pohon keputusan yang akan memberikan solusi berdasarkan aturan-aturan yang ditentukan. Kemampuan *Decision Tree* untuk menguji data, menemukan relasi antara variabel input dan variabel target melalui sebuah model, sehingga cocok digunakan sebagai langkah awal dalam proses pemodelan. Dalam *Decision Tree* terdapat tiga jenis *node*, yaitu:

1. *Root Node*, adalah node paling atas yang tidak memiliki input.
2. *Internal Node*, adalah node percabangan yang memiliki satu input dan memiliki minimal dua output.
3. *Leaf Node*, adalah node akhir yang hanya memiliki satu input dan tidak memiliki output[11].



Gambar 1 Bentuk Pohon Decision Tree

Terdapat banyak metode yang bisa digunakan dalam Algoritma *Decision Tree*, antara lain *CART*, *ID3* dan *C4.5*. Pada penelitian ini penulis menggunakan algoritma *Decision Tree C4.5* yang merupakan evolusi dari *ID3* sehingga prinsip dasar yang dimiliki algoritma *C4.5* hampir sama dengan *ID3*. Pohon atau Tree yang dibangun dengan Algoritma *Decision Tree C4.5* membagi data secara *rekursif* (mengulang dirinya sendiri) menjadi setiap bagian berisi data dari kelas yang sama[12]. Tahapan untuk membangun pohon keputusan adalah sebagai berikut[13]:

1. Tentukan *attribute* yang akan menjadi *root*
2. Membuat cabang dari tiap-tiap nilai
3. Membagi kasus menjadi cabang

4. Ulangi proses untuk setiap cabang hingga semua *instance* di cabang memiliki kelas yang sama.

Dalam menentukan *root* diperlukan perhitungan *Gain Ratio* untuk menentukan attribute apa yang cocok sebagai *root*. dengan nilai *gain ratio* tertinggi maka dijadikan sebagai *root*. Untuk menghitung nilai *gain ratio* menggunakan rumus sebagai berikut:

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

Keterangan:

- S = Kumpulan kasus
- A = Attribute
- N = Jumlah partisi attribute A
- |S<sub>i</sub>| = Jumlah kasus pada partisi ke-i
- |S| = Jumlah kasus dalam S

Evaluasi klasifikasi *data mining* disajikan dalam bentuk tabel dan disebut dengan *Confusion Matrix*[14]. *Confusion Matrix* merupakan sebuah metode yang digunakan untuk menghitung nilai akurasi. *Confusion Matrix* memiliki empat istilah dalam menghitung pengukuran kinerja yaitu:

1. *False Positif (FP)*, yaitu data negatif yang terprediksi sebagai data positif.
2. *False Negatif (FN)*, yaitu data positif yang terprediksi sebagai data negatif.
3. *True Positif (TP)*, yaitu data positif yang terprediksi benar.
4. *True Negatif (TN)*, yaitu data negatif yang terprediksi benar.

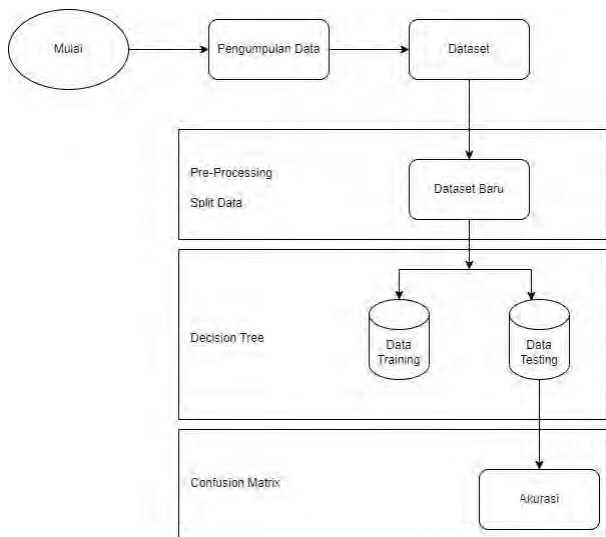
Rumus yang digunakan untuk menghitung tingkat akurasi data adalah sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\%$$

Kinerja model algoritma klasifikasi ditentukan saat model dihadapkan pada *data testing*, karena rata-rata model yang diterapkan dapat memprediksi dengan benar semua data yang menjadi *data trainingnya*[15].

### III. METODOLOGI PENELITIAN

Dalam penelitian ini penulis menggunakan metode eksperimen untuk menyelesaikan penelitian yang dilakukan. Penelitian ini melakukan perhitungan dataset dengan menerapkan algoritma *decision tree* untuk mengetahui prediksi dan tingkat akurasi *dataset*. Kerangka penelitian yang penulis gunakan seperti gambar berikut:



Gambar 2 Kerangka Penelitian

Penelitian ini dilakukan dalam beberapa tahap, tahap pertama yaitu persiapan dataset menggunakan proses pengumpulan data, kemudian data diseleksi dan dipilih agar sesuai dengan metode yang akan digunakan. Selanjutnya data akan di *cleaning* melalui proses *pre-processing* agar data menjadi data yang bersih dari gangguan. Setelah di *pre-processing*, data akan di *split* menjadi dua kategori yaitu *data training* dan *data testing*. Model yang digunakan adalah *Decision Tree* yang akan dievaluasi menggunakan *confusion matrix* untuk mendapatkan nilai akurasi.

#### A. Pengumpulan Data

Penelitian ini menggunakan data vaksinasi *Covid-19* yang diakses pada website pendataan vaksinasi Kelurahan Grogol Utara. Pengumpulan data ini dilakukan menggunakan metode observasi data dan mengumpulkan dataset vaksin yang didapatkan melalui website. Data-data tersebut memiliki attribute seperti no\_tiket, NIK, nama, usia, kategori, tanggal, status, nama vaksin, batch, dosis dan hasil observasi. Jumlah data yang digunakan dalam penelitian ini berjumlah 1142 data peserta vaksin dari bulan Agustus sampai September 2021.

No_Tiket	NIK	Nama_Lengkap	Usia	Kategori	Tanggal_Daftar	Status	Nama_Vaksin	Batch
J-0000OYM1	120515*****0006	CHAIRIL FUADI	35	MASYARAKAT UMUM	15/08/2021	Pemberian Vaksin Selesai	Astra Zeneca	210045
P-BZV5B2QV	367113*****0007	SUYANTO	49	MASYARAKAT UMUM	15/08/2021	Pemberian Vaksin Selesai	Astra Zeneca	57521401
Z-7621159F	317405*****0012	SUKARYONO	61		15/08/2021	Pemberian Vaksin Selesai	CoronaVac	24202021
Z-03FA4032	317405*****0012	Sulaeman	54		15/08/2021	Pemberian Vaksin Selesai	Astra Zeneca	57521401
P-PWB1K1PA	317405*****0001	R U D I	36	MASYARAKAT UMUM	18/09/2021	Pemberian Vaksin Selesai	COVID-19 Vaccine Pfizer	FG7900

Tabel 1 Record Data Peserta Vaksinasi Covid-19

#### B. Seleksi Data

Seleksi data dilakukan untuk menentukan *database* yang akan digunakan dalam *pre-processing* sebelum menjadi dataset yang siap diolah menggunakan *Decision Tree*. Pada tahap ini

penulis memilah data-data yang tidak sesuai dan tidak digunakan serta menentukan *attribute* yang akan digunakan sebagai *label data*. Hasil dari seleksi data didapatkan 651 data dari keseluruhan data sebanyak 1142 data.

#### C. Split Data

Pada tahap ini data-data yang terpilih dibagi menjadi dua bagian yaitu *data latih* dan *data uji*. Hasil seleksi data menghasilkan 651 data dan kemudian dibagi menjadi 80% *data latih* dan 20% *data uji*. *Data latih* digunakan sebagai data latih pembentuk pola dan *data uji* digunakan sebagai data pengujian pola yang telah terbentuk.

#### D. Decision Tree C4.5

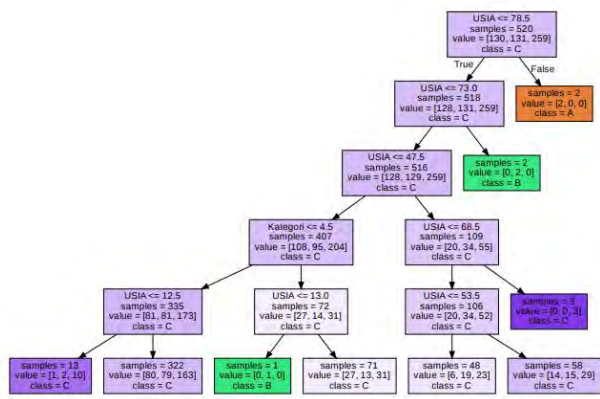
Tahap selanjutnya adalah modelling menggunakan metode klasifikasi dengan Algoritma *Decision Tree C.45* yaitu membuat pohon keputusan untuk menentukan prediksi dan akurasi sebuah data. Secara garis besar Algoritma *C4.5* dijelaskan sebagai berikut:

1. Tahap membentuk pohon keputusan dengan membentuk *root* dan dilanjutkan dengan pembagian data sesuai dengan *attribute* yang membentuk cabang (*leaf*).
2. Tahap pemangkasan pohon (*pruning*) untuk mengurangi ukuran pohon dan meminimalisir *error* pada prediksi.
3. Tahap pemebntukan aturan keputusan dari pohon keputusan yang telah terbentuk.

## IV. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan variabel data berupa data vaksinasi *Covid-19* Kelurahan Grogol Utara yang berjumlah 1142 data. Data di seleksi manual hingga didapatkan data sejumlah 651 data, data dibagi menjadi dua bagian yaitu *data latih* dan *data uji*. Persentase untuk pembagian data yaitu 80% untuk *data latih* dan 20% untuk *data uji* dengan jumlah data testing sebesar 130 data. Data tersebut kemudian dicari alur keputusannya menggunakan algoritma *Decision Tree C4.5* dengan struktur pohon keputusan.

Cara pertama untuk menentukan attribute sebagai root dilakukan dengan menghitung nilai *gain ratio*, yaitu mencari nilai *entropy* dari masing-masing *attribute*. *Attribute* dengan nilai *gain ratio* tertinggi maka ditetapkan sebagai *root* untuk membangun pohon keputusan. Pengujian dataset ini dilakukan menggunakan Bahasa pemrograman Python. Bentuk pohon keputusan yang didapatkan dari proses pengujian adalah sebagai berikut:



Gambar 3 Klasifikasi menggunakan *Decision Tree* untuk menentukan prediksi dampak vaksin

Hasil pengujian dari 130 data testing dan 521 data training dengan 1 label class *Nama\_Vaksin* menggunakan aplikasi *RapidMiner* adalah sebagai berikut:

**Accuracy: 52%**

		Precision	Recall	F1 Score
0	0.00	0.00	0.00	32
1	0.00	0.00	0.00	31
2	0.52	1.00	0.68	67

Gambar 4 Hasil Akurasi

Pengujian ini menghasilkan nilai akurasi yaitu 52%, nilai rata-rata *Precision* sebesar 52%, nilai rata-rata *Recall* sebesar 68%.

V. KESIMPULAN

- Kesimpulan dari penelitian ini adalah:
1. Dari hasil pohon keputusan didapatkan prediksi jika peserta vaksin dengan usia 50 tahun keatas (lansia) memiliki gejala demam setelah dilakukannya vaksin. Sedangkan peserta dengan usia 50 tahun kebawah tidak memiliki gejala (normal).
  2. Hasil pengujian yang dilakukan dengan menerapkan algoritma *Decision Tree C4.5* terhadap data Vaksinasi *Covid-19* Kelurahan Grogol Utara menghasilkan nilai *Accuracy* sebesar 52%, nilai rata-rata *Precision* sebesar 52%, nilai rata-rata *Recall* sebesar 68%.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan penelitian ini. Ucapan terimakasih juga penulis tujukan kepada Bapak

Leonard Goeiranto yang telah membimbing dan mengarahkan penelitian yang penulis lakukan sehingga penelitian ini dapat selesai tepat waktu. Tak lupa penulis mengucapkan terimakasih kepada kedua orang tua serta teman-teman seperjuangan yang selalu memberi semangat dan mendukung secara moril.

REFERENCES

- [1] "Peta Sebaran | Covid19.go.id." <https://covid19.go.id/peta-sebaran> (accessed Oct. 29, 2021).
- [2] Badan Pusat Statistik Kota Jakarta Selatan, "Jumlah Penduduk Menurut Kecamatan (ribu), 2010-2019," pp. 11–12, 2019, [Online]. Available: <https://jakselkota.bps.go.id/indicator/12/135/1/jumlah-penduduk-menurut-kecamatan.html>.
- [3] G. A. Prasetyo, R. G. Santosa, and A. R. Chrismanto, "Memprediksi Kategori Indeks Prestasi Mahasiswa," *Jurnal Terapan Teknologi Informasi*, vol. 3, no. 2, pp. 105–116, 2019, doi: 10.21460/jutei.2019.32.185.
- [4] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 3, no. July, p. 1, 2018, doi: 10.30645/jurasik.v3i0.60.
- [5] R. D. Syah, "Metode Decision Tree Untuk Klasifikasi Hasil Seleksi Kompetensi Dasar Pada Cpns 2019 Di Arsip Nasional Republik Indonesia," *J. Ilm. Inform. Komput.*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.
- [6] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [7] S. P. Kaur and V. Gupta, "COVID-19 Vaccine: A comprehensive status report," *Virus Res.*, vol. 288, p. 198114, Oct. 2020, doi: 10.1016/j.virusres.2020.198114.
- [8] Dinkes Provinsi Bali, "Yuk, Kenali lebih jauh Vaksinasi Covid-19 - Dinas Kesehatan Provinsi Bali," 2021, <https://diskes.baliprov.go.id/yuk-kenali-lebih-jauh-vaksinasi-covid-19/> (accessed Nov. 10, 2021).
- [9] P. Sharma, D. Singh, and A. Singh, "Classification algorithms on a large continuous random dataset using rapid miner tool," *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015*, no. October, pp. 704–709, 2015, doi: 10.1109/ECS.2015.7125003.
- [10] Z. Akhmad, "Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler," *Riau J. Of Computer Sci.*, vol. 02, no. 1, pp. 65–76, 2016.
- [11] P. Indriyanti, M. Fazalika Hismawan, and M. Mujiono, "Penerapan Decision tree dalam pengambilan keputusan untuk pemain Texas Holdem Poker," *J. Ilm. FIFO*, vol. 12, no. 2, p. 167, 2021, doi: 10.22441/fifo.2020.v12i2.006.
- [12] W. Syarif, "Aplikasi penentu hukum halal haram makanan dari jenis hewan berbasis web," 2011. <https://123dok.com/document/myj8e6kq-aplikasi-penentu-hukum-halal-haram-makanan-jenis-berbasis.html> (accessed Jun. 25, 2022).
- [13] S. Hamza, "Implementasi Data Mining Untuk Memprediksi Jumlah Peredaran Kendaraan Roda Empat Di Kota Ternate Menggunakan Metode C. 45," *Dinas Teknol.*, vol. 12, no. 1, pp. 52–56, 2019.
- [14] F. Gorunescu "Classification and Decision Trees," *Intell. Syst. Ref. Libr.*, vol. 12, no. 43, pp. 159–183, 2011, doi: 10.1007/978-3-642-19721-5\_4.
- [15] D. Istiawan and L. Khikmah, "Implementation of C4.5 Algorithm for Critical Land Prediction in Agricultural Cultivation Areas in Pemali Jratun Watershed," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 67, 2019, doi: 10.24014/ijaidm.v2i2.7569.

## KERTAS KERJA

### Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul Algoritma *Decision Tree* Untuk Mengklasifikasi Data Jenis Vaksin *Covid-19* Di Wilayah Kelurahan Grogol Utara. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, analisis dan perancangan, source code, dataset, tahapan eksperimen dan hasil semua eksperimen.

