IMPLEMENTASI KLASIFIKASI PENYAKIT GINJAL KRONIK MENGGUNAKAN DECISION TREE ALGORITMA CLASSIFICATION AND REGRESSION TREES (CART)

(Skripsi)

Oleh

FADHILAH GUSTRIANDINI NPM. 1917031087



JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

IMPLEMENTATION OF CHRONIC KIDNEY DISEASE CLASSIFICATION USING DECISION TREE CLASSIFICATION AND REGRESSION TREES (CART) ALGORITHM

By

FADHILAH GUSTRIANDINI

Classification is one of the data mining techniques by building a model whose results can be used to predict a class label and classify data. The decision tree is one of the classification methods that divides a set of data into several data sets with tree-shaped visualization. This research was conducted to classify patients who indicated chronic kidney disease by building a model of 11 variables that affect chronic kidney disease using one of the decision tree algorithms, namely the Classification and Regression Trees or CART algorithm. Model evaluation is built from three proportions of testing data and training data, which are 20:80, 30:70, and 40:60. Based on the results of the calculation, it is obtained that the highest accuracy rate is generated at the proportion of 30:70, which is 98.33%.

Key words: Data Mining, Decision Tree, Classification and Regression Trees (CART)

ABSTRAK

IMPLEMENTASI KLASIFIKASI PENYAKIT GINJAL KRONIK MENGGUNAKAN DECISION TREE ALGORITMA CLASSIFICATION AND REGRESSION TREES (CART)

Oleh

FADHILAH GUSTRIANDINI

Klasifikasi merupakan salah satu teknik dari *data mining* dan merupakan hasil dari pembentukan model yang digunakan untuk melakukan prediksi suatu label kelas serta mengelompokkan data. *Decision tree* adalah salah satu metode klasifikasi yang melakukan pembagian dari sekumpulan data menjadi beberapa himpunan data dengan visualisasi berbentuk pohon. Penelitian ini dilakukan untuk mengklasifikasikan pasien yang terindikasi penyakit ginjal kronik dengan membangun model dari 11 variabel yang mempengaruhi penyakit ginjal kronik menggunakan salah satu algoritma dari *decision tree*, yaitu algoritma *Classification and Regression Trees* atau CART. Evaluasi model dibangun dari tiga proporsi pembagian *data testing* dan *data training*, yaitu 20:80, 30:70, dan 40:60. Berdasarkan hasil perhitungan diperoleh bahwa tingkat akurasi tertinggi dihasilkan pada proporsi 30:70 yaitu sebesar 98.33%.

Kata Kunci: Data Mining, Decision Tree, Classification and Regression Trees (CART)

IMPLEMENTASI KLASIFIKASI PENYAKIT GINJAL KRONIK MENGGUNAKAN DECISION TREE ALGORITMA CLASSIFICATION AND REGRESSION TREES (CART)

Oleh

FADHILAH GUSTRIANDINI

Skripsi

Sebagai salah satu syarat untuk Mencapai Gelar SARJANA MATEMATIKA

Pada

Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS LAMPUNG BANDAR LAMPUNG 2023

Judul Skripsi

IMPLEMENTASI KLASIFIKASI PENYAKIT

GINJAL KRONIK MENGGUNAKAN

DECISION TREE ALGORITMA

CLASSIFICATION AND REGRESSION

TREES (CART)

Nama Mahasiswa

Fadhilah Gustriandini

Nomor Pokok Mahasiswa

1917031087

Program Studi

Matematika

Fakultas

: Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. Ahmad Faisol, S.Si., M.Sc.

NIP. 19800206 200312 1 003

Dr. Aang Nuryaman, S.Si., M.Si.

NIP. 19740316 200501 1001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.

NIP. 19740316 200501 1001

MENGESAHKAN

1. Tim Penguji

Ketua

Dr. Ahmad Faisol, S.Si., M.Sc.

MIS

Sekretaris

Dr. Aang Nuryaman, S.Si., M.Si.

Penguji

Bukan Pembimbing

Drs. Eri Setiawan, M.Si.

L

2. Plt. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

De Fing Heri Satria, S.Si., M.Si.

97110012005011002

Tanggal Lulus Ujian Skripsi: 31 Maret 2023

PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama Mahasiswa

Fadhilah Gustriandini

Nomor Pokok Mahasiswa

: 1917031087

Jurusan

Matematika

Judul Skripsi

: IMPLEMENTASI KLASIFIKASI PENYAKIT

GINJAL KRONIK MENGGUNAKAN

DECISION TREE ALGORITMA

CLASSIFICATION AND REGRESSION

TREES (CART)

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 11 April 2023

Yang menyatakan

Fadhilah Gustriandini NPM. 1917031087

RIWAYAT HIDUP

Penulis dilahirkan pada tanggal 4 Agustus 2001 di Bandar Lampung. Terlahir dari keluarga pasangan Bapak Hartadi dan Ibu Wasilah yang merupakan anak ketiga dari empat bersaudara.

Penulis menyelesaikan Pendidikan di Taman Kanak-Kanak Sari Teladan Bandar Lampung pada Tahun 2007. Pendidikan Sekolah Dasar di SDN 1 Beringin Raya Bandar Lampung pada tahun 2013. Pendidikan Sekolah Menengah Pertama di SMP Al-Kautsar Bandar Lampung pada tahun 2016. Pendidikan Sekolah Menengah Atas di SMA Al-Kautsar Bandar Lampung pada tahun 2019. Kemudian penulis melanjutkan Pendidikan S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) di Universitas Lampung pada tahun 2019 melalui jalur SBMPTN.

Selama menjadi mahasiswa penulis merupakan anggota Generasi Muda Penerus Himatika (GEMATIKA) pada periode 2019. Sebagai bentuk aplikasi bidang ilmu pengetahuan, penulis telah mengikuti Studi Independen dalam program Merdeka Belajar Kampus Merdeka (MBKM) di PT. MariBelajar selama empat bulan mulai dari bulan Februari sampai Juni 2022, kemudian penulis juga mengikuti program yang sama di PT. Hacktivate Teknologi Indonesia selama empat bulan mulai dari bulan Agustus sampai Desember 2022.

Sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah melaksanakan Kuliah Kerja Nyata (KKN) selama empat puluh hari di Kelurahan Karang Maritim, Kecamatan Panjang, Bandar Lampung pada bulan Januari sampai Februari 2022. Kemudian, sebagai bentuk aplikasi bidang ilmu di dunia kerja, penulis telah melaksanakan Kerja Praktik (KP) selama empat puluh hari di SAMSAT Bandar Lampung pada bulan Juni sampai Juli 2022.

KATA INSPIRASI

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya"

(Q.S. Al Bagarah: 286)

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan."

(Q.S. Al-Insyirah : 5-6).

"If you look at what you have in life, you'll always have more. If you look at what you don't have in life, you'll never have enough"

(Oprah Winfrey)

PERSEMBAHAN

Puji dan syukur kepada Allah SWT untuk rahmat, karunia dan Ridha-Nya, sehingga aku masih diberi kesempatan untuk menyelesaikan studi ini. Karya tulis yang sederhana ini kupersembahkan untuk kedua orangtua tercinta. Terimakasih untuk segala perjuangan dan pengorbanan yang sangat luar biasa, semoga kelak aku dapat membahagiakan dan membanggakan kalian. Terimakasih untuk kakak dan adik yang selalu menghibur, memotivasi, dan memberi semangat selama pengerjaan skripsi ini. Untuk almamaterku Univeristas Lampung tempatku menimba ilmu. Terimakasih untuk pengalaman dan kenangan yang sangat berharga.

SANWACANA

Segala puji bagi Allah SWT., yang telah memberikan rahmat, karunia dan Ridha-Nya, shalawat serta salam senantiasa selalu tercurahkan kepada Nabi Muhammad SAW., sehingga penulis dapat menyelesaikan skripsi yang berjudul "Implementasi Klasifikasi Penyakit Ginjal Kronik Menggunakan Decision Tree Algoritma Classification And Regression Trees (CART)". Skripsi ini disusun sebagai salah satu syarat untuk mendapatkan gelar Sarjana Matematika pada jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Tidak sedikit kendala yang dihadapi oleh penulis dalam pelaksanaan serta penyusunan skripsi ini, namun berkat ridha Allah SWT dan bantuan dari berbagai pihak penulis dapat menyelesaikan skripsi ini. Dalam penyusunan skripsi, penulis menyadari adanya keterbatasan pengetahuan dan kemampuan yang dimiliki, sehingga penulis mendapatkan banyak bimbingan, arahan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terimakasih kepada:

- 1. Bapak Dr. Ahmad Faisol, S.Si., M.Sc. selaku dosen pembimbing utama yang senantiasa memberikan bimbingan, bantuan dan saran kepada penulis dalam menyelesaikan skripsi.
- 2. Alm. Bapak Amanto, S.Si., M.Si. selaku dosen pembimbing kedua yang telah membimbing dan membantu penulis dalam menyelesaikan skripsi.
- 3. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan

- pembimbing kedua yang telah memberikan arahan dan saran kepada penulis dalam menyelesaikan skripsi.
- 4. Bapak Drs. Eri Setiawan, M.Si. selaku dosen pembahas yang telah memberikan saran, pengarahan, nasihat serta bantuan kepada penulis pada seminar hasil penelitian.
- 5. Ibu Dr. Fitriani, S.Si., M.Sc. selaku Dosen Pembimbing Akademik yang selalu memberikan motivasi, bimbingan dan perhatian penuh kepada penulis selama perkuliahan.
- 6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si selaku Plt. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
- 7. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
- 8. Kedua orang tua penulis, Bapak Hartadi dan Ibu Wasilah yang selalu berusaha memberikan yang terbaik kepada penulis dan selalu memberikan semangat, dukungan, serta doa kepada penulis.
- 9. Kakak dan adik penulis, Anisa Harsimaya, Ardhia Cahyani dan M. Nurur Rachman, yang senantiasa memberikan motivasi dan doa kepada penulis.
- 10. Sahabat-sahabat semasa perkuliahan, Citra Puspa Tria, Dinda Ayu Rachmadina, Nabilla Yolanda Paramitha, Putri Aisyah Kusuma Wardhani, Sund Grace Yuni Hutagalung dan Zidny Ilma Zain yang senantiasa memberikan bantuan, semangat dan menghibur penulis hingga penyusunan skripsi.
- 11. Sahabat-sahabat yang selalu menemani penulis hingga saat ini, Tharisa Cahya Andini, Sherin Dwi Noviadiningrum, dan Aziza Nur Fatiha. Serta Taufiq Nugroho yang selalu memotivasi dan mendoakan penulis.
- Teman-teman Matematika 2019 atas kebersamaan dan kenangan yang telah diberikan kepada penulis selama menempuh Pendidikan di Universitas Lampung.
- 13. Seluruh pihak terkait yang telah membantu penulis dalam penyusunan skripsi.

Penulis menyadari ketidaksempunaan skripsi ini dan masih terdapat kekurangan baik dalam penyajian maupun teknik penulisan. Oleh karena itu, saran dan kritik yang membangun senantiasa penulis harapkan demi menyempurnakan skripsi ini. Semoga atas segala dukungan dan bantuan yang telah diberikan kepada penulis mendapatkan balasan dari Allah SWT serta dapat bermanfaat untuk semua pihak baik bagi diri penulis sendiri maupun bagi pembaca.

Bandar Lampung, 11 April 2023 Penulis

Fadhilah Gustriandini

DAFTAR ISI

		Halaman
DA	AFTAR TABEL	vii
DA	AFTAR GAMBAR	ix
I.	PENDAHULUAN	1
	1.1 Latar Belakang dan Masalah	1
	1.2 Tujuan Penelitian	5
	1.3 Manfaat Penelitian	5
II.	TINJAUAN PUSTAKA	6
	2.1 Data Mining	6
	2.2 Machine Learning	7
	2.3 Klasifikasi	8
	2.4 Decision Tree	9
	2.5 Classification and Regression Trees (CART)	10
	2.6 Preprocessing data	12
	2.7 Confusion matrix	14
	2.8. Penyakit Ginjal Kronik	16
III	. METODE PENELITIAN	17
	3.1 Tempat dan Waktu Penelitian	17
	3.2 Data Penelitian	17
	3.3 Metode Penelitian	20

IV.	HASIL DAN PEMBAHASAN	23
	4.1 Visualisasi Data	23
	4.2 Preprocessing Data	28
	4.2.1 Pembersihan Data	29
	4.2.2 Seleksi Atribut	31
	4.2.3 Encoding Label	36
	4.2.4 Transformasi Data	37
	4.2.5 Pembagian Data	40
	4.3 Membangun Model <i>Decision Tree</i> Algoritma CART	42
	4.3.1 Menentukan kandidat split	43
	4.3.2 Menentukan simpul terminal	60
	4.3.3 Menentukan label kelas	61
	4.4 Evaluasi Model	62
	4.5 Visualisasi decision tree algoritma CART	67
V.	KESIMPULAN	73
	5.1 Kesimpulan	73
DA	FTAR PUSTAKA	74
LA	MPIRAN	77

DAFTAR TABEL

Tabel	Halaman
1. Confusion matrix	15
2. Data variabel independen penyakit ginjal kronik	18
3. Jumlah pasien yang terindikasi penyakit ginjal kronik	23
4. Pasien yang mengidap hipertensi	24
5. Pasien yang mengidap diabetes mellitus	26
6. Statistika deskriptif	27
7. Jumlah data hilang	29
8. Hasil perhitungan information gain	35
9. Data setelah dilakukan seleksi atribut	36
10. Encoding label pada variabel htn, dm, dan classification	37
11. Data setelah dilakukan <i>encoding label</i>	37
12. Kuartil 1, 2 dan 3 pada variabel numerik	38
13. Transformasi data	38
14. Data setelah dilakukan transformasi	40
15. Hasil perhitungan indeks gini dan gini splitting index iterasi 1	44
16. Hasil perhitungan indeks gini dan <i>gini splitting index</i> iterasi 2	48

17. Hasil perhitungan indeks gini dan <i>gini splitting index</i> iterasi 3	52
18. Hasil perhitungan indeks gini dan gini splitting index iterasi 4	55
19. Confusion matrix 20:80	62
20. Confusion matrix 30:70	63
21. Confusion matrix 40:60	65
22. Perbandingan nilai evaluasi model	66

DAFTAR GAMBAR

Gambar	Halaman
1. Visualisasi decision tree	9
2. Flowchart proses klasifikasi decision tree CART	22
3. Pie chart pasien terindikasi penyakit ginjal kronik	24
4. Diagram batang mengenai variabel hipertensi	25
5. Diagram batang mengenai variabel diabetes mellitus	26
6. Visualisasi jumlah data yang hilang	30
7. Diagram batang information gain	35
8. Simpul pertama	48
9. Simpul kedua	52
10. Simpul ketiga	55
11. Simpul keenam	59
12. Visualisasi simpul 1 sampai 7 decision tree algoritma CART	60
13. Simpul akar pada model proporsi 30:70	67
14. Visualisasi <i>decision tree</i> pada proporsi pembagian 30:70	68

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Matematika memiliki banyak cabang ilmu dalam pembelajaran maupun penerapannya. Salah satu cabang yang kini banyak digunakan karena memiliki banyak manfaat dalam berbagai bidang kehidupan adalah statistika. Statistika selalu berkembang dari waktu ke waktu. Pada saat statistika baru saja dikenalkan, statistika hanya dimanfaatkan untuk mengilustrasikan kondisi dan memberikan solusi dari permasalahan kenegaraan. Namun, seiring berkembangnya zaman dan teknologi yang semakin canggih, statistika seringkali digunakan sebagai alat bantu dalam mengatasi berbagai masalah dan pengambilan keputusan. Statistika merupakan ilmu pengetahuan yang berkaitan dengan macam-macam cara untuk menghimpun data, pengolahan data maupun cara menganalisa agar mampu menarik kesimpulan berdasarkan data yang telah diperoleh (Jaya & Ardat, 2013).

Sebelum melakukan pengolahan data, perlu dilakukan suatu proses untuk menemukan informasi yang bermanfaat dari himpunan data dengan menggunakan teknik pengenalan pola. Proses tersebut dinamakan sebagai *data mining* (Zyen & Mulyo, 2022). Proses menemukan pola atau informasi-informasi yang menarik dari sekumpulan data terpilih ini dilakukan dengan menggunakan teknik, metode, maupun algoritma dalam *data mining* yang sangat bervariasi. Oleh karena itu, tujuan yang ingin dicapai menjadi faktor yang berpengaruh dalam pemilihan metode atau

algoritma yang tepat, karena informasi-informasi ini dibutuhkan dalam menyelesaikan berbagai bidang kehidupan. Salah satu teknik pengenalan pola dari *data mining* yang dapat diimplementasikan adalah dengan menggunakan *machine learning* (Asroni et al., 2018).

Machine learning merupakan sebuah metode atau algoritma yang menggunakan bantuan dari komputer untuk menemukan pola-pola dari data yang telah dimasukkan. Bantuan dari komputer sangat penting dalam penggunaan machine learning karena dapat membantu manusia yang berinteraksi dengan komputer untuk mendapatkan suatu nilai yang bisa dipelajari maupun diteliti. Proses ini dilakukan dengan tujuan agar pola yang telah diperoleh tersebut dapat digunakan untuk melakukan peramalan atau klasifikasi data baru (Najar et al., 2022). Berdasarkan pengawasan manusia, machine learning terbagi menjadi 3 jenis (Dey, 2016), yaitu supervised learning, unsupervised learning dan reinforcement learning. Supervised learning merupakan jenis machine learning yang membutuhkan pengawasan eksternal. Seluruh algoritma akan mempelajari pola-pola dari data training yang akan digunakan pada data testing untuk diprediksi atau diklasifikasi.

Klasifikasi merupakan salah satu teknik dari *data mining* dan menggunakan jenis *supervised learning* dalam pengolahannya. Klasifikasi merupakan hasil dari pembentukan model yang akan digunakan untuk melakukan prediksi suatu label kelas serta mengelompokkan data yang disusun secara otomatis (Han et al., 2012).

Dalam mengaplikasikan klasifikasi, seringkali ditemukan data yang tidak memenuhi asumsi klasik metode klasifikasi, contohnya, asumsi kenormalan data dan multikolinearitas. Salah satu metode klasifikasi yang dapat menjadi solusi dari permasalahan tersebut adalah dengan menggunakan metode *decision tree*, karena *decision tree* bersifat non-parametrik yang tidak perlu dilakukan normalisasi data. *Decision tree* adalah salah satu metode klasifikasi yang menggunakan teori graf

untuk melakukan pembagian dari sekumpulan data menjadi beberapa himpunan data yang akan memiliki visualisasi berbentuk pohon (Hastuti, 2016).

Terdapat beberapa algoritma dalam metode klasifikasi *decision tree*, salah satunya adalah CART (*Classification and Regression Trees*). CART digunakan untuk mendapatkan pemaparan keterkaitan antara variabel independen yang berupa atribut dengan satu atau lebih variabel dependen yang menjadi variabel target. Jika variabel dependen memiliki tipe data numerik maka model pohon yang diperoleh adalah *regression trees*. Kemudian, jika variabel dependen memiliki tipe data kategorik maka model pohon yang diperoleh adalah *classification trees* (F. E. Pratiwi & Zain, 2014).

Penyakit ginjal kronik dapat mengakibatkan terjadinya fungsi ginjal yang menurun secara bertahap, dan pada umumnya akan berakhir dengan gagal ginjal. Penyakit gagal ginjal kronik cenderung tidak dapat pulih kembali pada kondisi lanjut, oleh karena itu perlu dilakukan prediksi sejak dini dan penanganan lebih lanjut (Anggraini & Fadila, 2022).

Beberapa penelitian yang sebelumnya telah menggunakan metode *decision tree* algoritma CART yaitu, Pratiwi dan Zain pada tahun 2014 yang melakukan klasifikasi dari permasalahan pengangguran tingkat terbuka menggunakan CART di Sulawesi Utara. Hasil yang diperoleh dari penelitian ini adalah nilai akurasi sebesar 78,90% (Pratiwi & Zain, 2014). Selanjutnya, Sauddin dkk. melakukan penelitian pada tahun 2019 mengenai klasifikasi tingkat partisipasi angakatan kerja Kota Makassar menggunakan metode CART. Nilai akurasi yang diperoleh dari penelitian ini adalah sebesar 69,68% (Sauddin et al., 2019).

Selanjutnya, Tou dan Endraswati pada tahun 2022 melakukan penelitian mengenai klasifikasi hasil diagnosa pasien BPJS berdasarkan hubungan gejala dan jenis

penyakit menggunakan metode *decision tree* algoritma CART. Hasil klasifikasi dari penelitian ini mendapatkan nilai akurasi sebesar 69% (Tou & Endraswari, 2022). Kemudian, penelitian yang dilakukan oleh Fajriati dan Syafriandi pada tahun 2022, membuat klasifikasi pada masalah status kerja pada angkatan kerja masyarakat Kabupaten Tanah Datar dengan menggunakan metode *decision tree* algoritma CART dan algoritma CHAID. Hasil dengan akurasi terbaik diperoleh dari penggunaan algoritma CART yang mendapatkan nilai akurasi sebesar 73,9% (Fajriati & Syafriandi, 2022). Selanjutnya, penelitian yang dilakukan oleh Munawwaroh dan Primandari pada tahun 2022, membuat klasifikasi dari masalah identifikasi awal LILA ibu hamil menggunakan metode *decision tree* algoritma CART dan mendapatkan hasil klasifikasi dengan tingkat akurasi sebesar 90% (Munawwaroh & Primandari, 2022).

Sedangkan, beberapa penelitian yang sebelumnya telah membahas mengenai penyakit ginjal kronik diantaranya adalah penelitian yang dilakukan oleh Yunus pada tahun 2018 yang melakukan implementasi algoritma *K-Nearest Neighbor* berbasis *Particle Swarm Optimization*. Penelitian ini memperoleh tingkat akurasi sebesar 78.75% (Yunus, 2018). Kemudian, pada tahun 2020, Safuan. menggunakan algoritma ID3. Nilai akurasi yang diperoleh dari penelitian ini mencapai 96.25% (Safuan, 2020). Penyakit pada organ ginjal merupakan kelainan yang muncul karena disebabkan oleh berbagai faktor. Oleh karena itu, pada penelitian ini akan dilakukan perancangan pola klasifikasi dari beberapa faktor tersebut untuk membantu tenaga kesehatan dalam memprediksi pasien terindikasi penyakit ginjal kronik dengan menggunakan metode *decision tree* algoritma CART.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- 1. Membuat klasifikasi untuk menentukan pasien yang terindikasi penyakit ginjal kronik menggunakan metode *decision tree* algoritma CART
- 2. Memperoleh nilai akurasi model berdasarkan klasifikasi *decision tree* algoritma CART yang telah dibuat.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- Menambah wawasan dan acuan bagi peneliti yang ingin mengembangkan ilmu pengetahuan di bidang matematika dalam membuat klasifikasi menggunakan metode *decision tree* algoritma CART
- 2. Menjadi bahan pertimbangan tenaga kesehatan dalam memprediksi pasien yang terindikasi penyakit ginjal kronik menggunakan klasifikasi *decision tree* algoritma CART

II. TINJAUAN PUSTAKA

2.1 Data Mining

Proses untuk menemukan informasi-informasi yang bermanfaat dari sekumpulan data dapat disebut sebagai *data mining*. *Data mining* merupakan proses berulang yang membutuhkan interaksi manusia untuk mendapatkan pola atau model baru yang dapat digeneralisasi di masa yang akan datang. Sehingga, proses ini akan bermanfaat ketika digunakan dalam melakukan suatu tindakan seperti dalam pengambilan keputusan (Nurzahputra et al., 2016).

Proses menemukan pola atau informasi-informasi yang menarik dari sekumpulan data terpilih ini dilakukan dengan menggunakan teknik, metode, maupun algoritma dalam *data mining* yang sangat bervariasi. Oleh karena itu, keseluruhan tujuan yang ingin dicapai menjadi faktor yang berpengaruh dalam pemilihan metode atau algoritma yang tepat (Asroni et al., 2018). Pada *data mining*, terdapat beberapa teknik yang dapat digunakan, diantaranya prediktif atau klasifikasi, klastering dan asosiasi (Zyen & Mulyo, 2022).

2.2 Machine Learning

Machine learning merupakan sebuah metode atau algoritma yang menggunakan bantuan dari komputer untuk menemukan pola-pola dari data yang telah dimasukkan, sehingga pola tersebut dapat digunakan untuk melakukan peramalan atau klasifikasi data baru (Najar et al., 2022). Machine learning adalah cabang dari artificial inteliligence, yang merupakan ilmu yang berkaitan dengan desain dan pengembangan algoritma yang dapat memungkinkan komputer untuk berkembang berdasarkan beberapa kondisi dari data sensor atau basis data (Jalal & Ezzedine, 2019). Terdapat beberapa algoritma machine learning, untuk itu diperlukan pemilihan algoritma yang terbaik.

Berdasarkan pengawasan manusia, *machine learning* terbagi menjadi 3 jenis, yaitu (Dey, 2016):

a. Supervised Learning

Supervised learning merupakan jenis machine learning yang membutuhkan pengawasan eksternal. Sekumpulan data yang telah dimasukkan akan dibagi menjadi data training dan data testing. Data training memiliki variabel dependen yang perlu diprediksi atau diklasifikasi. Kemudian, seluruh algoritma akan mempelajari polapola dari data training yang akan digunakan pada data testing untuk diprediksi atau diklasifikasi.

b. Unsupervised Learning

Unsupervised learning mempelajari atribut baru dari data. Ketika terdapat data baru maka akan digunakan atribut yang telah dipelajari sebelumnya untuk mengenali kelas data. Jenis *machine learning* ini umum digunakan pada klastering.

c. Reinforcement Learning

Reinforcement learning adalah jenis machine learning yang membuat keputusan berdasarkan tindakan yang akan memberikan hasil positif, akan tetapi, pengguna machine learning ini tidak mengetahui tindakan yang akan diambil sampai akhirnya

sebuah kondisi diberikan. *Reinforcement learning* bergantung pada dua kriteria, yaitu pencarian *trial* dan *error* serta hasil yang tertunda.

2.3 Klasifikasi

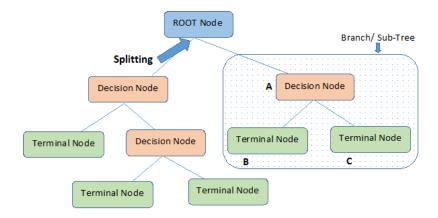
Klasifikasi merupakan salah satu teknik dari *data mining*. Klasifikasi merupakan hasil dari pembentukan model yang akan digunakan untuk melakukan prediksi suatu label kelas. Dalam klasifikasi, kelas adalah atribut yang merupakan variabel independen (Han et al., 2012). Klasifikasi juga dapat digunakan untuk mengelompokkan suatu data yang disusun secara sistematis. Penggunaan metode klasifikasi dapat dilakukan dengan menggunakan dua pendekatan, yaitu pendekatan parametrik dan non-parametrik. Pendekatan parametrik digunakan berdasarkan adanya asumsi bahwa pengembalian memiliki distribusi normal serta terdapat suatu hubungan linear antara perubahan nilai instrumen dengan dampak dari perubahan faktor resiko. Sedangkan pendekatan non-parametrik digunakan berdasarkan data historis yang tidak perlu melakukan normalisasi data, dengan mempertimbangkan pengembalian masa lalu yang sebenarnya (Astuti & Gunarsih, 2021).

Masalah klasifikasi seringkali terjadi dalam kehidupan sehari-hari dalam berbagai bidang, oleh karena itu, masalah ini harus diselesaikan dengan memilih metode klasifikasi yang tepat. Terdapat dua tahap dalam proses klasifikasi. Tahap pertama yaitu tahap *learning* untuk membuat pola model klasifikasi. Sedangkan, tahap kedua yaitu tahap pengklasifikasian yang merupakan tahap menggunakan model klasifikasi untuk melakukan prediksi label kelas dari suatu data. (Utomo, 2022).

2.4 Decision Tree

Dalam mengaplikasikan klasifikasi, seringkali ditemukan data yang tidak memenuhi asumsi klasik metode klasifikasi, contohnya, asumsi kenormalan data dan multikolinearitas. Salah satu metode klasifikasi yang dapat menjadi solusi dari permasalahan tersebut adalah dengan menggunakan metode *decision tree*, karena *decision tree* bersifat non-parametrik. *Decision tree* adalah salah satu metode klasifikasi yang menggunakan teori graf untuk melakukan pembagian dari sekumpulan data menjadi beberapa himpunan data. Visualisasi bentuk pohon merupakan hasil yang diperoleh setelah mengaplikasikan metode *decision tree* (Hastuti, 2016).

Visualisasi decision tree yang berbentuk pohon, terdiri dari simpul akar (root node), simpul keputusan (decision node), cabang (branches), dan simpul terminal (terminal nodes). Simpul akar (root node) merupakan simpul yang menjadi induk dari seluruh simpul pada decision tree. Setiap simpul pada decision tree mewakilkan variabel (attribute), cabang merupakan suatu keputusan, sementara simpul terminal adalah hasil (Patel & Prajapati, 2018). Visualisasi decision tree dapat disajikan seperti gambar berikut:



Gambar 1. Visualisasi decision tree

2.5 Classification and Regression Trees (CART)

Terdapat beberapa algoritma dalam metode klasifikasi *decision tree*, salah satunya adalah CART (*Classification and Regression Trees*). CART digunakan untuk mendapatkan pemaparan keterkaitan antara variabel independen yang berupa atribut dengan satu atau lebih variabel dependen yang menjadi variabel target. Model pohon yang diperoleh akan terbentuk berdasarkan tipe data dari variabel dependen. Jika variabel dependen memiliki tipe data numerik maka model pohon yang diperoleh adalah *regression trees*. Kemudian, jika variabel dependen memiliki tipe data kategorik maka model pohon yang diperoleh adalah *classification trees* (F. E. Pratiwi & Zain, 2014).

Algoritma CART memiliki tiga tahapan dalam menyelesaikan proses klasifikasi secara berurutan, sebagai berikut (Tou & Endraswari, 2022):

a. Menentukan kandidat split

Parameter yang akan digunakan dalam metode CART dapat ditentukan dengan cara melakukan pembagian kandidat split menjadi dua bagian, yaitu kandidat split kanan dan kandidat split kiri berdasarkan variabel independen. Parameter yang telah terpilih akan terbentuk menjadi suatu himpunan kelas yang disebut sebagai simpul. Simpul yang telah diperoleh akan melakukan proses pemilihan secara iterasi hingga akan didapat iterasi terakhir yang menghasilkan simpul terminal. Teknik untuk menentukan kandidat split pada algoritma CART dilakukan dengan menggunakan indeks gini i(t). Indeks gini merupakan suatu pengukuran dari variasi antara probabilitas nilai-nilai variabel independen (Sauddin et al., 2019). Indeks gini digunakan untuk mendapatkan kandidat split yang paling optimal. Persamaan indeks gini yaitu:

$$i(t) = 1 - \sum_{j=1}^{\infty} P^{2}(j|t)$$
 (2.1)

Keterangan:

i(t) = Indeks gini

P(j|t) = Jumlah kelas j terhadap simpul t dimana j = 1,2,3,...,n, dengan

$$P(j|t) = \frac{n_j(t)}{n_{(t)}}$$

 $n_i(t)$ = Jumlah data kelas j pada simpul t

 $n_{(t)}$ = Jumlah data pada simpul t

Tahap berikutnya yaitu melakukan perhitungan *gini splitting index* yang akan berperan sebagai tahap evaluasi pemilihan kandidat *split* s pada t menggunakan persamaan dibawah ini:

$$GINI_{split} = \sum_{j=1}^{\infty} \frac{n_{(t)}}{n} i(t)$$
(2.2)

Keterangan:

i(t) = Indeks gini

 $n_{(t)}$ = Jumlah data pada simpul t

n = Jumlah data

b. Menentukan simpul terminal

Proses untuk menentukan pola *decision tree* akan dihentikan ketika sampai pada jumlah iterasi dimana sebaran sampel telah berada pada satu kelas, dimana simpul terakhir disebut sebagai simpul terminal.

c. Menentukan label kelas

Suatu proses untuk menentukan label kelas dilakukan pada tiga macam simpul, yaitu simpul akar, simpul keputusan dan juga simpul terminal. Proses menentukan label kelas ini sangat penting untuk dilakukan pada simpul terminal, karena pada simpul terminal menggunakan label untuk melakukan prediksi objek pada kelas yang

terdapat pada simpul terminal tersebut (Sauddin et al., 2019). Dibawah ini merupakan persamaan yang digunakan untuk menentukan label kelas

$$P(j|t) = \max_{j} \frac{N_j(t)}{N(t)}$$
(2.3)

Keterangan:

P(j|t) = Jumlah kelas j pada simpul t

 $N_i(t)$ = Jumlah kelas pada simpul t

 $N_{(t)}$ = Jumlah pengamatan di simpul t

2.6 Preprocessing data

Preprocessing data adalah suatu proses dalam mempersiapkan data yang telah diperoleh sehingga data tersebut siap untuk menjadi sebuah model. Terdapat beberapa tahap yang dapat dilakukan dalam proses preprocessing data, diantaranya:

a. Pembersihan data

Tahap ini dilaksanakan untuk mempermudah proses pengklasifikasian serta memastikan bahwa data tersebut sudah benar dan siap untuk diolah menjadi suatu model (Laila et al., 2022). Pada tahap pembersihan data, akan dilakukan imputasi pada data yang hilang dengan menggunakan *mean* untuk data yang bertipe numerik dan modus untuk data yang bertipe kategorik.

b. Seleksi atribut

Seleksi atribut merupakan teknik untuk memilih atribut terbaik dari sekumpulan atribut data yang diperoleh. Seleksi atribut dapat dilaksanakan menggunakan *information gain*. Pada proses *information gain*, seluruh atribut akan diurutkan, kemudian atribut yang memiliki nilai terbesar dikatakan memiliki relevansi yang besar dengan data terkait, sehingga atribut tersebut akan dipilih. *Entropy* adalah

ukuran ketidakpastian kelas dengan menggunakan peluang kejadian atau atribut tertentu. Tahap pertama yang dilakukan pada proses *information gain* adalah mencari nilai *entropy* dengan menggunakan persamaan dibawah ini (Beniwal & Arora, 2012).

Entropy (S) =
$$\sum_{i}^{c} -P_{i} \log_{2}(P_{i})$$
 (2.4)

Keterangan:

c = Akumulasi nilai dari kelas klasifikasi

 P_i = Akumulasi sampel dari kelas i

Lalu, ketika nilai *entropy* telah diperoleh, maka akan dilakukan penghitungan *information gain* dengan menggunakan rumus dibawah ini.

$$Gain(S,A) = Entropy(S) - \sum_{Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$
 (2.5)

Keterangan:

Gain(S, A) = Nilai information gain dari atribut

A = Atribut

v = Peluang nilai atribut A

Values(A) = Peluang nilai himpunan A

 S_v = Jumlah sampel nilai dari v

S = Jumlah seluruh sampel data

 $Entropy(S_v)$ = Entropy sampel nilai v

c. Encoding label

Encoding label merupakan proses untuk memberikan label kelas pada atribut yang memiliki tipe data kategorik, yaitu dari 0 sampai n kelas (Latifah et al., 2019).

d. Transformasi data

Pada tahap ini, akan dilakukan transformasi data ke dalam format yang sesuai dengan format proses *data mining*. Atribut data yang digunakan merupakan atribut yang bertipe numerik. Sehingga, setelah dilakukan transformasi, maka data pada atribut tersebut akan berbentuk interval kelas (Tou & Endraswari, 2022). Data-data yang bertipe numerik ini akan ditransformasikan menjadi 4 kelas, yaitu kelas 0, 1, 2, dan 3 dengan menggunakan kuartil 1 (Q_1) sampai kuartil 3 (Q_3).

e. Pembagian data

Dalam klasifikasi, terdapat tahap pembagian data, yaitu tahap dimana sekumpulan data yang telah disiapkan akan dibagi menjadi *data training* dan *data testing*. Pembagian data menjadi 2 bagian ini dilakukan agar memperoleh model yang baik. *Data training* digunakan untuk melakukan estimasi terhadap model. Sedangkan, *data testing* digunakan untuk melakukan pengujian terhadap model yang sebelumnya telah diestimasi oleh *data training* (U. M. Pratiwi & Ibad, 2022). Rumus untuk menghitung jumlah dari *data testing* dan *data training* dari persentase pembagian data yang telah ditentukan adalah sebagai berikut:

Jumlah
$$data \ testing = Persentase \ data \ testing \times Jumlah \ data$$

Jumlah $data \ training = Persentase \ data \ training \times Jumlah \ data$

(2.6)

2.7 Confusion matrix

Confusion matrix digunakan pada tahap pengujian untuk melihat performa dari suatu model klasifikasi yang telah diperoleh setelah digunakan untuk mengolah data. Nilai True Positive (TP) dapat dikatakan ketika hasil nilai prediksi dan nilai yang sebenarnya sama, yang berarti bahwa hasil tersebut tepat. Sementara, jika hasil nilai prediksi dinyatakan tidak tepat namun nilai yang sebenarnya tepat, maka dikatakan

sebagai *False Negative* (FN). Kemudian, jika hasil nilai prediksi dan nilai yang sebenarnya tidak tepat, maka kondisi ini disebut sebagai *True Negative* (TN). Selanjutnya, jika hasil nilai prediksi tepat, namun nilai yang sebenarnya tidak tepat maka disebut sebagai *False Positive* (FP) (Laila et al., 2022).

Tabel 1. Confusion matrix

		Nilai prediksi		
		False	True	
Nilai yang	False	True Negative	False Positive	
sebenarnya	True	False Negative	True Positive	

Tahap pengujian menggunakan *confusion matrix* dapat dilakukan dengan empat evaluasi, yaitu (Utomo, 2022):

a. Accuracy

Accuracy merupakan evaluasi yang dihitung berdasarkan tingkat kedekatan antara hasil nilai prediksi dengan nilai yang sebenarnya. Berikut ini adalah rumus perhitungan nilai *accuracy*:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%$$
 (2.7)

b. Precision

Precision merupakan evaluasi yang dihitung dengan membandingkan data benar dengan seluruh data yang digunakan sistem baik benar maupun salah. Berikut ini adalah rumus perhitungan nilai *precision*:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$
 (2.8)

c. Recall

Recall merupakan evaluasi yang dihitung dengan membandingkan data benar dengan seluruh data benar yang digunakan maupun tidak digunakan sistem. Berikut ini adalah rumus perhitungan nilai *recall*:

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{2.9}$$

d. F1-score

F1-score adalah rata-rata harmonis dari nilai precision dan recall. Jika nilai dari f1-score baik, maka dapat dikatakan bahwa model klasifikasi yang telah diperoleh memiliki nilai precision dan recall yang baik pula. Berikut ini adalah rumus perhitungan nilai f1-score:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$
 (2.10)

2.8. Penyakit Ginjal Kronik

Penyakit ginjal kronik dapat mengakibatkan terjadinya fungsi ginjal yang menurun secara bertahap, dan pada umumnya akan berakhir dengan gagal ginjal. Gagal ginjal kronik merupakan suatu kondisi dimana organ ginjal mengalami kelainan struktural atau gangguan dari fungsi organ ginjal yang sudah terjadi selama lebih dari 3 bulan. Penyakit gagal ginjal kronik cenderung tidak dapat pulih kembali pada kondisi lanjut, oleh karena itu perlu prediksi sejak dini dan penanganan lebih lanjut. Penyakit pada organ ginjal merupakan kelainan yang muncul karena disebabkan oleh berbagai faktor. Contoh faktor yang dapat mempengaruhi penyakit ginjal ini, seperti infeksi, tumor, kelainan bawaan, penyakit metabolisme atau degeneratif, dan lain-lain (Anggraini & Fadila, 2022).

III. METODE PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2022/2023, yang bertempatan di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan untuk melakukan penelitian ini adalah data sekunder mengenai data penyakit ginjal kronik yang diunduh langsung dari situs web *UCI*: https://archive.ics.uci.edu/ml/datasets/Chronic Kidney Disease. Data tersebut merupakan data yang telah dikumpulkan dari pasien – pasien sebuah rumah sakit di India dengan periode kurang lebih 2 bulan. Jumlah data ini sebanyak 400 dan memiliki 26 variabel, dimana salah satu variabelnya adalah variabel dependen yang berperan sebagai variabel target. Data penelitian ini memiliki 25 jenis variabel independen yang berperan sebagai atribut. Dari 25 variabel tersebut, terdapat 11 variabel yang memiliki tipe data numerik dan 14 variabel yang memiliki tipe data kategorik. 25 variabel independen dari data yang digunakan pada penelitian ini dapat dilihat pada tabel berikut:

Tabel 2. Data variabel independen penyakit ginjal kronik

No.	Kode Variabel	Nama Variabel	Kategori	Keterangan
1.	age	Age	Numerik	Umur pasien.
2.	bp	Blood pressure	Numerik	Tekanan darah dalam satuan mm/Hg.
3.	sg	Specific gravity	Kategorik	Specific gravity merupakan kepekatan pada urin.
4.	al	albumin	Kategorik	Albumin adalah jenis protein yang dapat mengindikasi tahap awal kerusakan ginjal.
5.	su	sugar	Kategorik	Gula dalam darah.
6.	rbc	Red blood cells	Kategorik	Sel darah merah.
7.	рс	Pus cell	Kategorik	Pus cell kumpulan cairan, sel darah putih, mikrooorganisme dan bahan selular yang menunjukkan adanya luka yang terinfeksi.
8.	рсс	Pus cell clumps	Kategorik	Pus cell clumps adalah sekumpulan sel darah putih dalam pemeriksaan urin, bisa menjadi pertanda adanya infeksi.
9.	ba	Bacteria	Kategorik	Bakteri yang melalui uretra dan berjalan ke kandung kemih.
10.	bgr	Blood glucose random	Numerik	Kadar glukosa darah dalam satuan mgs/dl.

Tabel 2. Lanjutan

11.	bu	Blood urea	Numerik	Kadar urea darah dalam satuan
				mgs/dl. Urea darah adalah
				produk sisa metabolisme protein
				dalam tubuh.
12.	sc	Serum creatinine	Numerik	Kadar serum kreatinin dalam
				satuan mgs/dl. Serum kreatinin
				merupakan sampah hasil
				metabolisme otot dalam darah.
13.	sod	Sodium	Numerik	Kadar sodium dalam satuan
				mEq/L. Sodium merupakan zat
				elektrolit berfungsi untuk
				menjaga keseimbangan cairan
				tubuh.
14.	pot	Potassium	Numerik	Kadar potasium dalam satuan
				mEq/L. Potasium merupakan
				salah satu mineral dalam tubuh.
15.	hemo	Hemoglobin	Numerik	Kadar hemoglobin dalam satuan
				gms. Hemoglobin adalah protein
				dalam sel darah merah.
16.	pcv	Packed cell	Numerik	Packed cell volume
		volume		menunjukkan jumlah sel darah
				merah dalam satu liter darah.
17.	wc	White blood cell	Numerik	Jumlah sel darah putih dalam
		count		satuan sel/cmm
18.	rc	Red blood cell	Numerik	Jumlah sel darah merah dalam
		count		satuan juta/cmm
19.	htn	Hypertension	Kategorik	Hipertensi
20.	dm	Diabetes mellitus	Kategorik	Penyakit diabetes mellitus

Tabel 2. Lanjutan

21.	cad	Coronary artery	Kategorik	Penyakit jantung koroner
		disease		
22.	appet	Appetite	Kategorik	Nafsu makan pasien
23.	pe	Pedal edema	Kategorik	Pedal edema merupakan
				penyakit dimana terdapat koleksi
				abnormal cairan di ruang antara
				sel-sel.
24.	ane	Anemia	Kategorik	Anemia
25.	id	id	Kategorik	Nomor id pasien

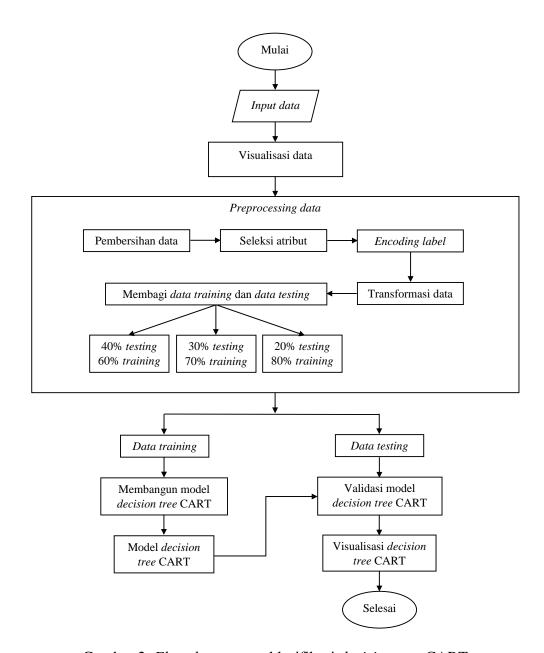
3.3 Metode Penelitian

Penelitian ini menggunakan model *Decision Tree* khususnya algoritma CART dengan tujuan mengklasifikasikan pasien yang terindikasi penyakit ginjal kronik. Berikut ini disajikan langkah-langkah yang akan dilakukan dalam penelitian:

- 1. Memasukkan data mengenai penyakit ginjal kronik yang telah diunduh dari *website UCI* ke dalam bahasa pemrograman *Python*.
- 2. Melakukan visualisasi data penyakit ginjal kronik dalam bentuk plot.
- 3. Melakukan *preprocessing* data, sebagai berikut:
 - a. Melakukan pemeriksaan data untuk mengetahui apakah terdapat data yang hilang atau memiliki duplikat.
 - b. Melakukan imputasi pada data yang hilang menggunakan *mean* pada data numerik dan menggunakan modus pada data kategorik.
 - c. Melakukan seleksi atribut menggunakan *information gain* dengan menggunakan persamaan (2.5).
 - d. Memilih variabel yang akan digunakan dalam penelitian dengan cara memilih atribut yang memiliki nilai tinggi pada perhitungan *information gain*.

- e. Melakukan *encoding label* pada data kategorik.
- f. Melakukan transformasi data pada tipe data numerik.
- g. Melakukan pembagian data untuk *data training* dan *data testing* sebanyak 3 bagian, dengan perbandingan 60:40, 70:30, dan 80:20 menggunakan persamaan (2.6)
- 4. Membangun model decision tree menggunakan algoritma CART:
 - a. Menentukan kandidat split kanan dan kandidat split kiri.
 - b. Menghitung nilai indeks gini menggunakan persamaan (2.1).
 - c. Menghitung nilai *gini splitting index* menggunakan persamaan (2.2).
 - d. Memilih nilai *gini splitting index* terbaik, dengan melihat nilai yang paling kecil.
 - e. Melakukan langkah b-d hingga sampai pada iterasi dimana sebaran sampel telah berada pada satu kelas.
 - f. Melakukan pemilihan simpul terminal.
 - g. Memberikan label kelas pada simpul terminal dengan menggunakan persamaan (2.3).
- 5. Melakukan validasi model *decision tree* algoritma CART menggunakan *confusion matrix*.
- 6. Menampilkan visualisasi dari *decision tree* algoritma CART yang diperoleh dari proses klasifikasi yang telah dilakukan.

Proses klasifikasi metode *decision tree* algoritma CART dengan Bahasa pemrograman *python* dapat disajikan secara singkat menggunakan *flowchart* dibawah ini:



Gambar 2. Flowchart proses klasifikasi decision tree CART

.

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan yang telah dilakukan mengenai pembangunan model *decision tree* algoritma CART dalam mengklasifikasikan pasien yang terindikasi penyakit ginjal kronik, dapat diambil berbagai kesimpulan sebagai berikut:

- 1. Terdapat 11 variabel independen atau atribut terbaik yang terpilih melalui seleksi atribut dengan mengurutkan nilai *information gain*, yaitu *blood glucose random* (bgr), *blood urea* (bu), *serum creatinine* (sc), *sodium* (sod), *potassium* (pot), *hemoglobin* (hemo), *packed cell volume* (pcv), *white blood cell count* (wc), *red blood cell count* (rc), *hypertension* (htn), dan *diabetes mellitus* (dm)
- Setelah melakukan evaluasi dari 3 proporsi pembagian data, model terbaik dari klasifikasi decision tree algoritma CART dengan data penyakit ginjal kronik diperoleh pada proporsi 30% data testing dan 70% data training dengan nilai accuracy, precision, dan f1-score berturut-turut sebesar 98.33%, 100%, dan 96.78%.

DAFTAR PUSTAKA

- Anggraini, S., & Fadila, Z. 2022. Kualitas Hidup Pasien Gagal Ginjal Kronik dengan Dialisis di Asia Tenggara: A Systematic Review. *Jurnal Kesehatan Masyarakat.* **11**(1): 77–83.
- Asroni, Fitri, H., & Prasetyo, E. 2018. Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*. **21**(1): 60–64.
- Astuti, P. E., & Gunarsih, T. 2021. Value-at-Risk Analysis in Risk Measurement and Formation of Optimal Portfolio in Banking Share. *JBTI : Jurnal Bisnis : Teori Dan Implementasi*. **12**(2): 103–114.
- Beniwal, S., & Arora, J. 2012. Classification and Feature Selection Techniques in Data Mining. *International Journal of Engineering Research & Technology* (*IJERT*). **1**(6): 1–6.
- Dey, A. 2016. Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*. **7**(3): 1174–1179.
- Fajriati, Y. R., & Syafriandi. 2022. Pengklasifikasian Status Kerja pada Angkatan Kerja di Kabupaten Tanah Datar Menggunakan Metode CART dan Metode CHAID. *Journal of Mathematics UNP*. **7**(3): 25–33.
- Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts and Techniques*. Elsevier Inc, Waltham.
- Hastuti, Y. 2016. Klasifikasi Karakteristik Mahasiswa Universitas Cokroaminoto Palopo Menggunakan Metode Naive Bayes dan Decision Tree. *Jurnal Dinamika*. **07**(2): 34–41.

- Jalal, D., & Ezzedine, T. 2019. Performance Analysis of Machine Learning Algorithms for Water Quality Monitoring System. 2019 International Conference on Internet of Things, Embedded Systems and Communications, IINTEC 2019 - Proceedings: 86–89.
- Jaya, I., & Ardat. 2013. *Penerapan statistik untuk pendidikan*. Citapustaka Media Perintis, Bandung
- Laila, W., Widiarto, W., Wijayanto, A., & Suryani, E. 2022. Rekomendasi Makanan Bagi Pasien Hiperlipidiemia Berdasarkan Hasil Klasifikasi Menggunakan Metode Naive Bayes dan Decision Tree. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*. **8**(2): 328–336.
- Latifah, R., Wulandari, E. S., & Kreshna, P. E. 2019. Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn. *Seminar Nasional Sains Dan Teknologi* 2019. **9**: 1–6.
- Munawwaroh, D. A., & Primandari, A. H. 2022. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree Dengan Algoritma CART Untuk Prediksi LILA Ibu Hamil Berpotensi Gizi Kurang. *Delta: Jurnal Ilmiah Pendidikan Matematika*. **10**(2), 367–380.
- Najar, A. M., Sudarsana, I. W., Albab, M. U., & Andhika, S. 2022. Machine Learning untuk Identifikasi Jenis Kanker Darah (Leukemia). *Vygotsky: Jurnal Pendidikan Matematika Dan Matematika*. **4**(1): 47.
- Nurzahputra, A., Safitri, A. R., & Muslim, M. A. 2016. Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree. *Prosiding Seminar Nasional Matematika X Universitas Negeri Semarang 2016*: 717–722.
- Patel, H., & Prajapati, P. 2018. Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Science and Engineering*. **6**(10): 74–78.
- Pratiwi, F. E., & Zain, I. 2014. Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains Dan Seni Pomits*. **3**(1): D54–D59.
- Pratiwi, U. M., & Ibad, M. 2022. Klasifikasi Faktor Yang Berpengaruh Dalam Kehamilan Tidak Diinginkan Menggunakan Metode Algoritma Decision Tree. *Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*. **3**(2): 406–416.
- Safuan. 2020. Deteksi Penyakit Gagal Ginjal Kronik Menggunakan Algoritma ID3. *Urnal Elektronika Dan Komputer.* **13**(1): 8–17.

- Sauddin, A., Alwi, W., & AN, A. Ningsih. 2019. Klasifikasi Tingkat Partisipasi Angkatan Kerja Kota Makassar Menggunakan Metode CART. *Jurnal Matematika Dan Statistika Serta Aplikasinya*. **7**(2): 20–36.
- Tou, N., & Endraswari, P. M. 2022. Implementasi Data Mining Dalam Klasifikasi Hasil Diagnosa Pasien Bpjs Menggunakan Algoritma Cart. *JIKA (Jurnal Informatika) Universitas Muhammadiyah Tangerang*. **6**(2): 170–179.
- Utomo, K. S. 2022. Perbandingan Algotitma Machine Learning Untuk Penentuan Klasifikasi Kemiskinan Multidimensi di Provinsi Nusa Tenggara Timur. *BPS Provinsi NTT: Jurnal Statistika Terapan.* **2**(1): 36–46.
- Yunus, W. 2018. Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik. *Jurnal Teknik Elektro CosPhi*. **2**(2): 51–55.
- Zyen, A. K., & Mulyo, H. 2022. Implementasi Algoritma Decision Tree C5. 0 untuk Klasifikasi Pasien Demam Berdarah di Kabupaten Rembang. *AMRI (Analisa, Metode, Rekayasa, Informatika)*. **1**(1): 42–51.