

**ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK
DIAGNOSIS PENYAKIT GINJAL KRONIS**

Skripsi

Oleh

**OLIVIA DESTI RIANA
NPM 1917051044**



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

ABSTRAK

ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK DIAGNOSIS PENYAKIT GINJAL KRONIS

Oleh

OLIVIA DESTI RIANA

Gagal ginjal kronis adalah suatu kondisi di mana fungsi ginjal secara perlahan mengalami penurunan akibat kerusakan pada jaringan ginjal. Akibatnya, dapat terjadi penumpukan cairan, elektrolit, dan limbah di dalam tubuh, yang dapat mengganggu keseimbangan tubuh. Tingginya tingkat kematian dan biaya perawatan yang diperlukan membuat pentingnya deteksi dan pencegahan dini gagal ginjal kronis melalui gaya hidup sehat dan pengendalian penyakit yang dapat meningkatkan risiko kondisi ini. Penelitian ini bertujuan untuk menerapkan algoritme *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* dalam proses diagnosa penyakit ginjal kronis. Dalam upaya mengevaluasi kinerja ketiga metode tersebut, digunakan *dataset* yang berasal dari *platform* Kaggle, terdiri dari 400 data dengan 24 atribut. *Dataset* ini terdiri dari 250 data yang mewakili pasien dengan penyakit ginjal kronis dan 150 data yang mewakili pasien tanpa penyakit ginjal kronis. Penggunaan algoritme-algoritme tersebut menghasilkan tingkat akurasi masing-masing sebesar 0,97 (*Decision Tree*), 0,98 (*Support Vector Machine*), dan 0,96 (*Naïve Bayes*). Hasil ini menunjukkan bahwa metode *Support Vector Machine* memiliki performa terbaik dengan akurasi tertinggi, yaitu 0,98. Dengan demikian, penelitian ini memberikan sumbangan berharga untuk pemahaman lebih lanjut dalam mendiagnosis penyakit ginjal kronis menggunakan teknik *data mining*.

Kata kunci : *Machine Learning*, *Decision Tree*, *Support Vector Machine*, *Naïve Bayes*, Penyakit Ginjal Kronis.

ABSTRACT

PERFORMANCE ANALYSIS OF VARIOUS CLASSIFICATION METHODS FOR CHRONIC KIDNEY DISEASE DIAGNOSIS

By

OLIVIA DESTI RIANA

Chronic kidney failure is a condition in which kidney function gradually declines due to damage to the kidney tissue. As a result, there can be an accumulation of fluids, electrolytes, and waste within the body, which can disrupt the body's balance. The high rate of mortality and required treatment costs underscore the importance of early detection and prevention of chronic kidney failure through a healthy lifestyle and control of diseases that can increase the risk of this condition. This study aims to implement the Decision Tree, Support Vector Machine, and Naïve Bayes algorithms in the process of diagnosing chronic kidney disease. In an effort to evaluate the performance of these three methods, a dataset sourced from the Kaggle platform was used, consisting of 400 data points with 24 attributes. The dataset comprises 250 data points representing patients with chronic kidney disease and 150 data points representing patients without chronic kidney disease. The use of these algorithms resulted in accuracy rates of 0.97 (Decision Tree), 0.98 (Support Vector Machine), and 0.96 (Naïve Bayes) respectively. These results indicate that the Support Vector Machine method demonstrates the best performance with the highest accuracy of 0.98. Thus, this research provides a valuable contribution to further understanding the diagnosis of chronic kidney disease using data mining techniques.

Key words : Machine Learning, Decision Tree, Support Vector Machine, Naïve Bayes, Chronic Kidney Disease.

**ANALISIS KINERJA BERBAGAI METODE KLASIFIKASI UNTUK
DIAGNOSIS PENYAKIT GINJAL KRONIS**

Oleh

OLIVIA DESTI RIANA

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



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

Judul Skripsi : **ANALISIS KINERJA BERBAGAI METODE
KLASIFIKASI UNTUK DIAGNOSIS
PENYAKIT GINJAL KRONIS**

Nama Mahasiswa : **Olivia Desti Riana**

Nomor Pokok Mahasiswa : 1917051044

Program Studi : S1 Ilmu Komputer

Jurusan : Ilmu Komputer

Fakultas : Matematika Dan Ilmu Pengetahuan Alam

MENYETUJUI

1. **Komisi Pembimbing**



Prof. Admi Syarif, Ph.D.
NIP. 19670103 199203 1 003



Dewi Asiah Shofiana, S.Komp., M.Kom.
NIP. 19950929 202012 2 030

2. **Ketua Jurusan Ilmu Komputer**




Didik Kurniawan, S.Si., M.T.
NIP. 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : Prof. Admi Syarif, Ph.D.



.....

Sekretaris : Dewi Asiah Shofiana, S.Komp., M.Kom.



.....

Penguji Bukan Pembimbing : Dr. rer. Nat Akmal Junaidi, M.Sc.



.....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: 31 Juli 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Olivia Desti Riana

NPM : 1917051044

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Analisis Kinerja Berbagai Metode Klasifikasi Untuk Diagnosis Penyakit Ginjal Kronis” merupakan karya saya sendiri, bukan karya orang lain. Semua tulisan yang tertulis dalam skripsi ini telah mengikuti kaidah penulisan ilmiah Universitas Lampung. Jika di kemudian hari terbukti bahwa karya tulis ilmiah saya terbukti hasil menjiplak karya orang lain, maka saya siap menerima sanksi berupa pencabutan gelar yang saya peroleh.

Bandar Lampung, 31 Juli 2023



Olivia Desti Riana

NPM. 1917051044

RIWAYAT HIDUP



Penulis dilahirkan di Lampung Barat, pada tanggal 22 Desember 2000, sebagai anak ketiga dari tiga bersaudara. Penulis menyelesaikan pendidikan formal di SD N 02 Way Petai Lampung Barat dan selesai pada tahun 2013. Kemudian pendidikan menengah pertama di SMP N 02 Sumber Jaya Lampung Barat yang selesai pada tahun 2016, lalu melanjutkan ke pendidikan menengah atas di SMA Sumber Jaya Lampung Barat yang diselesaikan pada tahun 2019.

Pada tahun 2019 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Penerima Mahasiswa Perluasan Akses Pendidikan (PMPAP). Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain:

1. Mengikuti Karya Wisata Ilmiah (KWI) di Desa Tambah Dadi, Kecamatan Purbolinggo, Kabupaten Lampung Timur pada bulan Desember tahun 2019.
2. Menjadi anggota kaderisasi Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2020.
3. Menjadi Panitia Penerimaan Mahasiswa baru Jurusan Ilmu Komputer (PRINTER) pada divisi Komisi Disiplin (KOMDIS) pada tahun 2021.
4. Menjadi Panitia divisi Photography pada Pekan Raya Jurusan (PRJ) HIMAKOM 2021.
5. Melaksanakan Kerja Praktik periode 2021/2022 di PT Jasa Raharja Persero Cabang Lampung.
6. Melaksanakan Kuliah Kerja Nyata periode 2 pada tahun 2022 di Desa Tebing, Kecamatan Melinting, Kabupaten Lampung Timur.

MOTTO

“Kamu adalah pencipta masa depanmu sendiri. Mulailah hari ini.”

-Catherine Pulsifer-

“Setiap orang punya waktunya masing-masing, jangan pernah iri dengan pencapaian orang lain.”

-Anonim-

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur kepada Allah Subhanahu Wa Ta'alaah atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini.

Sholawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Mama dan Bapakku tercinta, terima kasih atas segala pengorbanan, do'a, dan cinta yang kalian berikan. Tanpa dukungan kalian, skripsi ini tidak akan pernah terwujud. Kalian adalah sumber inspirasi sejati dalam hidupku. Kalian adalah dua sosok yang tak tergantikan dalam hidupku. Terima kasih atas kesabaran kalian, motivasi kalian ketika aku meragukan diri sendiri, dan kebanggaan kalian ketika aku mencapai setiap pencapaian. Segala hal yang kalian lakukan untukku, baik yang terlihat maupun yang tersembunyi, tidak pernah kubiarkan sia-sia. Skripsi ini adalah bukti cinta dan penghargaanku kepada kalian.

Kepada kakakku serta keluarga besar yang selalu memberikan do'a dan dukungan semangat kepadaku.

Sahabat-sahabatku, terima kasih telah menemaniku, mendukungku, dan selalu memberikan kebahagiaan dalam hidupku.

Seluruh Keluarga Besar Ilmu Komputer 2019.

Almamater Tercinta Universitas Lampung.

SANWACANA

Puji syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa, karena atas rahmat dan hidayah-Nya skripsi ini dapat diselesaikan. Skripsi dengan judul “*Analisis Kinerja berbagai Metode Klasifikasi untuk Diagnosis Penyakit Ginjal Kronis*” adalah salah satu syarat untuk memperoleh gelar sarjana Ilmu Komputer di Universitas Lampung. Dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Allah SWT yang telah memberikan limpahan berkah, rahmat, hidayah, dan karunia-Nya.
2. Kedua orang tua saya, kakak dan ayuk yang selalu memberikan doa, dukungan, semangat, motivasi, dan kasih sayang luar biasa tak terhingga. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan di akhirat, Aamiin.
3. Bapak Prof. Admi Syarif, Ph.D., selaku pembimbing utama atas kesediannya untuk memberikan bimbingan, saran dan kritik dalam proses penyelesaian skripsi ini.
4. Ibu Dewi Asiah Shofiana, S.Komp., M.Kom., selaku pembimbing kedua atas kesediannya memberikan bimbingan, saran dan kritik dalam proses penyelesaian skripsi ini.
5. Dr. rer. nat. Akmal Junaidi, M.Sc., selaku dosen pembahas sekaligus Sekretaris Jurusan Ilmu Komputer yang telah memberikan masukan dan saran yang bermanfaat guna menyempurnakan penulisan skripsi ini.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan FMIPA unila.
7. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer.
8. Ibu Yohana Tri Utami, S.Kom., M.Kom., selaku pembimbing akademik.
9. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan banyak ilmu dan pengalaman selama penulis menjalani perkuliahan.

10. Lidia Ade Putri, Mentari Mei Wulan, Ulfa Tania Putri, Sonia Maharani, Tina Maharani, Enisa Azhari, dan Eva Hardiyanti adalah sahabat yang selalu menemani dan memberi semangat untuk dapat menyelesaikan skripsi ini.
11. Salsabilla Julia Farhana, Sendy Hani Pramita, Finka Marisa Geananda sufie, Devi Rahmadia fitri, Vira Verina, Hani Cita Lestari, dan Zahara Liza Mulyani yang merupakan sahabat-sahabat saya di kampus yang sejak awal kuliah selalu ada dan saling membantu satu sama lain, tempatku menuangkan segala kebahagiaan dan kesedihan. Terima kasih sudah menjadi sahabat yang baik.
12. Keluarga Ilmu Komputer 2019 yang tidak bisa disebut satu persatu yang telah bersedia menjadi rekan kelompok, rekan diskusi, dan rekan bercanda. Terima kasih sudah memberi warna dan pengalaman selama masa perkuliahan.

Bandar Lampung, 31 Juli 2023

Olivia Desti Riana

DAFTAR ISI

	Halaman
DAFTAR ISI.....	i
DAFTAR TABEL	iii
DAFTAR GAMBAR.....	iv
I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terdahulu.....	5
2.2. Penyakit Ginjal Kronis	16
2.3. <i>Data Mining</i>	17
2.4. Klasifikasi.....	18
2.5. <i>Decision Tree</i>	19
2.5.1. Kelebihan Metode <i>Decision Tree</i>	21
2.5.2. Kekurangan Metode <i>Decision Tree</i>	22
2.6. <i>Support Vector Machine</i> (SVM)	22
2.6.1. Kelebihan Metode <i>Support Vector Machine</i> (SVM)	24
2.6.2. Kekurangan Metode <i>Support Vector Machine</i> (SVM)	24
2.7. Metode <i>Naïve Bayes</i>	24
2.7.1. Kelebihan Metode <i>Naïve Bayes</i>	26
2.7.2. Kekurangan Metode <i>Naïve Bayes</i>	26
2.8. Evaluasi Kinerja Klasifikasi	26

2.8.1.	<i>Accuracy</i>	27
2.8.2.	<i>Precision</i>	28
2.8.3.	<i>Recall</i> atau <i>Sensitivity</i>	28
2.8.4.	<i>F-1 Score</i>	28
2.8.5.	<i>K-Fold Cross Validation</i>	28
III.	METODOLOGI PENELITIAN	30
3.1.	Tempat dan Waktu Penelitian	30
3.2.	Data dan Alat.....	31
3.2.1.	Data	31
3.2.2.	Alat.....	33
3.2.3.	Studi Literatur	36
3.2.4.	Pengambilan Data	37
3.2.5.	<i>Preprocessing</i>	37
3.2.6.	Pembagian Data	37
3.2.7.	Pemodelan.....	37
3.2.8.	Evaluasi.....	38
3.2.9.	Hasil	38
IV.	HASIL DAN PEMBAHASAN	39
4.1.	Hasil Penelitian.....	39
4.1.1.	<i>Preprocessing</i>	39
4.1.2.	Pemodelan.....	60
4.2.	Pembahasan	66
V.	SIMPULAN DAN SARAN	71
5.1.	Simpulan.....	71
5.2.	Saran.....	72
	DAFTAR PUSTAKA	73

DAFTAR TABEL

Tabel	Halaman
1. Penelitian terdahulu	5
2. <i>Confusion matrix</i> (Desiani, 2022).....	27
3. Prediksi	27
4. <i>K-Fold Cross Validation</i>	29
5. Alur waktu weneelitian.....	30
6. Penjelasan atribut (Kumar et al., 2022)	31
7. Transformasi atribut.....	40
8. Jumlah nilai nol pada setiap atribut numerik.....	57
9. Jumlah nilai nol pada setiap atribut kategorik	57
10. Kategori unik	59
11. <i>Iterasi k-fold cross-validation</i> pada <i>Decision Tree</i>	61
12. <i>Confusion matrix Decision tree</i>	61
13. Metrik evaluasi <i>decision tree</i>	62
14. <i>Iterasi k-fold cross-validation</i> pada <i>Support Vector Machine</i>	63
15. <i>Confusion matrix Support Vector Machine</i>	63
16. Metrik evaluasi <i>Support Vector Machine</i>	63
17. <i>Iterasi k-fold cross-validation</i> pada Naïve Bayes.....	65
18. <i>Confusion matrix</i> Naïve Bayes	65
19. Matrik evaluasi Naïve Bayes.....	66
20. Perbandingan 3 metode.....	67
21. Perbandingan dengan beberapa penelitian sebelumnya	68

DAFTAR GAMBAR

Gambar	Halaman
1. Faktor pengaruh penyakit ginjal kronis (Senan et al., 2021).....	16
2. Model susunan <i>Decision Tree</i>	19
3. SVM linear (Trivusi, 2022)	22
4. SVM non-linear	23
5. Usaha menemukan <i>hyperplane</i> terbaik (Wibawa et al., 2018).....	23
6. Alur kerja penelitian.	36
7. Distribusi atribut numerik.....	40
8. Distribusi atribut kategorik	41
9. <i>Heatmap</i> dari <i>dataset</i> penyakit ginjal kronis	42
10. Plot <i>violin</i> pada atribut <i>red blood cell count</i>	43
11. Plot KDE atribut <i>red blood cell count</i>	44
12. Plot <i>violin</i> atribut <i>white blood cell count</i>	44
13. Plot KDE atribut <i>white blood cell count</i>	45
14. Plot <i>violin</i> atribut <i>packed cell volume</i>	46
15. Plot KDE atribut <i>packed cell volume</i>	47
16. Plot <i>violin</i> atribut hemoglobin	47
17. Plot KDE atribut hemoglobin	48
18. Plot <i>violin</i> atribut <i>blood glucose random</i>	49
19. Plot KDE atribut <i>blood glucose random</i>	50
20. Plot <i>violin</i> atribut <i>sodium</i>	50
21. Plot KDE atribut <i>sodium</i>	51
22. Plot <i>violin</i> atribut <i>blood urea</i>	52
23. Plot KDE atribut <i>blood urea</i>	53

24.	Plot <i>violin</i> atribut <i>specific gravity</i>	53
25.	Plot KDE atribut <i>specific gravity</i>	54
26.	Plot <i>scatter</i> hubungan hemoglobin dan <i>packed cell volume</i>	55
27.	Plot <i>scatter</i> hubungan <i>red blood cell count</i> dan <i>packed cell volume</i>	56
28.	Plot <i>scatter</i> hubungan <i>sugar</i> dan <i>blood glucose random</i>	56
29.	Plot <i>scatter</i> hubungan <i>packed cell volume</i> dan <i>blood urea</i>	57
30.	Atribut kategorik setelah dilakukan <i>encoding label</i>	60

I. PENDAHULUAN

1.1. Latar Belakang

Gagal ginjal kronis adalah kondisi dimana ginjal mengalami penurunan fungsi secara bertahap akibat adanya kerusakan pada jaringan ginjal. Fungsi ginjal yaitu membuang sisa metabolisme dan mengatur keseimbangan cairan dalam tubuh manusia. Jika ginjal mengalami penurunan fungsi dalam mempertahankan metabolisme dan keseimbangan cairan, maka dapat menyebabkan cairan, elektrolit, dan limbah menumpuk di dalam tubuh dan akan menimbulkan banyak gangguan (Kusuma et al., 2019). Gejala penderita penyakit gagal ginjal kronis stadium 1 – 3 biasanya tidak begitu terlihat. Gejala gagal ginjal kronis akan terasa ketika sudah mencapai stadium 4 dan 5. Kondisi ini mengharuskan penderita melakukan cuci darah. Penyakit gagal ginjal kronis bersifat progresif dan tidak bisa disembuhkan kembali, dimana tingkat mortalitas yang tinggi, dan memakan biaya yang mahal (dr. Meva Nareza, 2021).

Masalah yang dihadapi penderita penyakit gagal ginjal kronis adalah mahalnya biaya perawatan dan belum ditemukannya obat untuk penyakit gagal ginjal kronis. Sepuluh persen populasi dunia menderita penyakit ginjal kronis, dan jutaan orang meninggal setiap tahun karena kurangnya akses pengobatan (Aulia, 2017a). Menurut Riskesdas 2013, kejadian penyakit tidak menular (PTM) semakin meningkat dan menjadi penyebab utama kematian di Indonesia. Penyakit tersebut antara lain hipertensi (25,8%), obesitas (15,4%), stroke (12,1%), diabetes melitus (2,3%), penyakit jantung koroner (1,5%), dan gagal ginjal kronis (0,2%)(Aulia, 2017b). Angka kejadian gagal ginjal kronik di

Indonesia berdasarkan data Riskesdas pada tahun 2018 yaitu sebesar 0,38% dari 252.124.458 penduduk Indonesia dimana 713.783 orang menderita gagal ginjal kronik.

Untuk itu penting bagi setiap orang untuk memperhatikan kesehatan ginjal. sebagai langkah pencegahan diperlukan deteksi dini pada penyakit ginjal agar dapat dicegah dengan menjalani pola hidup sehat dan mengontrol penyakit yang dapat meningkatkan risiko terjadinya gagal ginjal kronis. Penyebab penyakit ginjal yang paling sering adalah hipertensi, diabetes, dan radang ginjal (dr. Meva Nareza, 2021).

Data mining adalah metode yang banyak digunakan untuk membuat penemuan ilmiah, dengan cara mengelola *dataset* yang selama ini hanya disimpan tanpa dilakukan pengelolaan lebih lanjut (Amalia, 2018). Penggunaan teknik *data mining* sudah tersebar luas di dunia medis untuk membantu membuat prediksi tentang masalah kesehatan.

Beberapa penelitian terdahulu yang telah dilakukan yaitu *Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques* oleh (Senan et al., 2021), menggunakan Metode *k-Nearest Neighbour* dan *Random Forest*. Dari hasil penelitian ini didapatkan *accuracy* terbaik yaitu Metode *Random Forest* dengan *accuracy* 100%. Selanjutnya penelitian dengan judul *Comparative analysis of classification algorithms for chronic kidney disease diagnosis* oleh (Saringat et al., 2019), menggunakan metode *Support Vector Machine*, *Decision Tree*, *k-Nearest neighbor*, dan *Regression*. Dari hasil penelitian ini didapatkan *accuracy* terbaik yaitu Metode *Regression* dengan *accuracy* 98.25%. Selanjutnya pada penelitian yang berjudul *Machine Learning Algoritma Selection for Chronic Kidney Disease Diagnosis and Classification* oleh (Gokiladevi et al., 2022), dengan metode *K-NN*, *SVM*, *Random Forest*, dan *Logistic Regression*. Dari hasil

penelitian ini didapatkan *accuracy* terbaik yaitu Metode *Random Forest* memperoleh *accuracy* 99%.

Selanjutnya akan dilakukan penelitian dengan metode *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes*. *Decision Tree* adalah suatu sistem yang dirancang untuk membantu menemukan masalah dan mengambil keputusan dengan mempertimbangkan berbagai faktor yang ada dalam kerangka masalah tersebut (Azmi & Dahria, 2013). *Support Vector Machine* adalah Sebuah sistem pembelajaran yang menggunakan ruang hipotesis dalam bentuk fungsi linier dalam fitur berdimensi tinggi menggunakan algoritme berdasarkan teori optimasi (Susilowati et al., 2015). *Naïve Bayes* adalah teknik prediksi dengan probabilistik sederhana yang merupakan dasar dari penerapan teorema *Bayes* (aturan *Bayes*), dan asumsi independensi (ketidaktergantungan) yang kuat, teorema ini menyatakan seberapa besar tingkat keyakinan subjektif harus berubah secara wajar dengan adanya isyarat baru (Prasetyo et al., 2013).

Berdasarkan kondisi tersebut, maka akan dilakukan penelitian dengan judul sebagai berikut “**Analisis Kinerja Berbagai Metode Klasifikasi untuk Diagnosis Penyakit Ginjal Kronis**”. Diharapkan penelitian ini dapat digunakan sebagai informasi dan dapat memberikan sumbangan pemikiran untuk penelitian selanjutnya dalam mendiagnosis penyakit ginjal kronis.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dirumuskan masalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritme *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* untuk memprediksi seseorang apakah terkena penyakit ginjal kronis?
2. Bagaimana *performance* metode *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* pada klasifikasi diagnosis penyakit ginjal kronis?

1.3. Batasan Masalah

Prediksi penyakit ginjal kronis harus fokus pada masalah agar masalah yang diteliti tidak terlalu luas dan sesuai dengan tujuan yang ingin dicapai. Berikut batasannya:

1. Penelitian ini menggunakan *Dataset* dari *Kaggle*.
2. Metode yang digunakan adalah *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes*.
3. Bahasa pemrograman yang digunakan yaitu python.

1.4. Tujuan Penelitian

1. Mengimplementasikan algoritme *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* untuk diagnosis penyakit ginjal kronis.
2. Mengetahui *performance* dari metode *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* untuk diagnosis penyakit ginjal kronis.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini yaitu dapat mengetahui kinerja dari beberapa metode klasifikasi pada diagnosis penyakit ginjal kronis yang diharapkan dapat memberikan sumbangan pemikiran dan mampu dikembangkan lebih baik lagi pada penelitian selanjutnya.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini tidak terlepas dari penelitian-penelitian sebelumnya yang telah dilakukan. Rangkuman penelitian sebelumnya disajikan pada Tabel 1 berikut:

Tabel 1. Penelitian terdahulu

Penelitian	Data	Metode	Hasil
(Senan et al., 2021)	Nama data: penyakit Ginjal Kronis Jumlah data: 400 pasien Atribut: 24 Sumber: <i>University of California.</i>	<i>Support Vector Machine, k-Nearest Neighbour, Random Forest.</i>	Metode <i>Support Vector Machine</i> : accuracy 96.67%. Metode <i>k-Nearest Neighbour</i> : accuracy 98.33%. Metode <i>Random Forest</i>: accuracy 100%.
(Saringat et al., 2019)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400	<i>Support Vector Machine, Decision Tree,</i>	Metode <i>Support Vector Machine</i> : accuracy 90.25%.

Penelitian	Data	Metode	Hasil
	<p>Atribut: 25</p> <p>Sumber: <i>UCI Machine Learning Repository website.</i></p>	<p><i>k-Nearest neighbor, Regression.</i></p>	<p>Metode <i>Decision Tree</i>: accuracy 95.50%.</p> <p>Metode <i>k-Nearest neighbor</i>: accuracy 94.75%.</p> <p>Metode Regression: accuracy 98.25%.</p>
(Gokiladevi et al., 2022)	<p>Nama data: Penyakit Ginjal Kronis</p> <p>Jumlah data: 400</p> <p>Atribut: 24</p> <p>Sumber: <i>UCI, benchmark CKD.</i></p>	<p><i>K-NN, SVM, Random Forest, Decision tree, Logistic Regression.</i></p>	<p>Metode <i>K-NN</i>: accuracy 67.50%.</p> <p>Metode <i>SVM</i>: accuracy 73.75%.</p> <p>Metode <i>Random Forest</i>: accuracy 98.75%.</p> <p>Metode <i>Decision tree</i>: accuracy 96.25%.</p> <p>Metode <i>Logistic Regression</i>: accuracy 94.68%.</p>

Penelitian	Data	Metode	Hasil
(Kumar et al., 2022)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 pasien Atribut: 24 Sumber: <i>University of California.</i>	<i>AdaBoost,</i> <i>Bagging,</i> <i>CatBoost,</i> <i>Decision Tree,</i> <i>Extra Trees,</i> <i>Gaussian</i> <i>Naive Bayes,</i> <i>Gradient</i> <i>Boosting,</i> <i>K-nearest</i> <i>neighbors,</i> <i>LightGBM,</i> <i>Multi-Layer</i> <i>Perceptron,</i> <i>Random</i> <i>Forest,</i> <i>Stochastic</i> <i>Gradient,</i> <i>Boosting,</i> <i>Support vector</i> <i>machines,</i> <i>XGBoost.</i>	Metode <i>AdaBoost</i> : <i>accuracy 97%.</i> Metode <i>Bagging</i> : <i>accuracy 60%.</i> Metode <i>CatBoost</i> : <i>accuracy 96%.</i> Metode <i>Decision</i> <i>Tree</i> : <i>accuracy</i> <i>94.00%.</i> Metode <i>Extra</i> <i>Trees</i>: <i>accuracy</i> <i>98%.</i> Metode <i>Gaussian</i> <i>Naive Bayes</i> : <i>accuracy 93.00%.</i> Metode <i>Gradient</i> <i>Boosting</i> : <i>accuracy 93.00%.</i> Metode <i>K-nearest</i> <i>neighbors</i> : <i>accuracy 67.00%.</i>

Penelitian	Data	Metode	Hasil
			<p>Metode <i>LightGBM</i>: <i>accuracy</i> 96%.</p> <p>Metode <i>Multi-Layer Perceptron</i>: <i>accuracy</i> 89%.</p> <p>Metode <i>K-nearest neighbors</i>: <i>accuracy</i> 67.00%.</p> <p>Metode <i>Stochastic Gradient Boosting</i>: <i>accuracy</i> 93%.</p> <p>Metode SVM: <i>accuracy</i> 97%.</p> <p>Metode <i>XGBoost</i>: <i>accuracy</i> 95%.</p>
(Krishna et al., 2022)	<p>Nama data: gagal jantung</p> <p>Jumlah data: 299</p> <p>Atribut: 13</p>	<p><i>Decision Tree</i>, <i>Gaussian Naïve Bayes</i>.</p>	<p>Metode <i>Decision Tree</i>: <i>accuracy</i> 82%.</p>

Penelitian	Data	Metode	Hasil
	Sumber: Kaggle.		Metode Gaussian Naïve Bayes: accuracy 86 %.
(Tekale et al., 2022)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 Atribut: 25 Sumber: <i>uci repository</i> .	<i>Decision Tree, Support Vector Machine (SVM).</i>	Metode <i>Decision Tree</i> : accuracy 92%. Metode Support Vector Machine (SVM): accuracy 97 %.
(Zeynu, 2018)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 Atribut: 24 Sumber: <i>uci repository</i> .	<i>KNN, ANN, NB, Ensemble model.</i>	Metode <i>KNN</i> : accuracy 98.5%. Metode <i>ANN</i> : accuracy 97.75%. Metode <i>NB</i> : accuracy 94.5%. Metode Ensemble model: accuracy 99%.
(Faddillah et al., 2019)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 pasien	<i>Naïve Bayes</i>	Metode naïve bayes: accuracy 91.25%.

Penelitian	Data	Metode	Hasil
	Atribut: 24 Sumber: <i>Indians Chronic Kidney Disease (CKD)</i> .		
(Amalia, 2018)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 pasien Atribut: 24 Sumber: <i>uci repository</i> .	<i>Support Vector Machine (SVM)</i> <i>Neural Network (NN)</i>	Metode <i>Support Vector Machine (SVM)</i>: accuracy 95.16%. Metode <i>Neural Network</i> : accuracy 93.36%.
(Ariani & Samsuryadi, 2019)	Nama data: Penyakit Ginjal Kronis Jumlah data: 400 pasien Atribut: 24 Sumber: <i>Repository University of California (UCI)</i>	<i>K-Nearest Neighbor (k-NN)</i>	Metode <i>K-Nearest Neighbor (KNN)</i>: accuracy 85.83%.

Penelitian	Data	Metode	Hasil
	<i>Repository Machine Learning Benchmark</i>).		

Pada penelitian yang berjudul “*Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques*” yang dilakukan oleh (Senan et al., 2021). Pada penelitian ini *dataset* dikumpulkan dari 400 pasien dari *University of California*. *Dataset* memiliki 24 atribut yang terbagi menjadi 11 atribut numerik dan 13 atribut nominal, atribut tersebut yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, and anemia*, serta atribut kelas seperti “*ckd*” dan “*not ckd*” untuk klasifikasi. *Dataset* merupakan 250 kasus kelas “*ckd*” oleh 62.5% dan 150 kasus “*not ckd*” sebesar 37.5%. Metode yang digunakan pada penelitian ini yaitu SVM, *k-Nearest Neighbour*, dan *Random Forest*. Dari penelitian ini diperoleh hasil yaitu dengan Metode SVM memperoleh *accuracy* 96.67, *precision* 92.00, *recall* 94.74, *F1-score* 97.30. Metode *k-Nearest Neighbour* memperoleh *accuracy* 98.33, *precision* 100, *recall* 97.37, *F1-score* 98.67. Metode *Random Forest* memperoleh *accuracy* 100, *precision* 100, *recall* 100, *F1-score* 100.

Pada penelitian yang berjudul “*Comparative analysis of classification algorithms for chronic kidney disease diagnosis*” oleh (Saringat et al., 2019). Telah dilakukan penelitian untuk melakukan perbandingan beberapa algoritme klasifikasi yaitu *Support Vector Machine, Decision Tree, k-Nearest neighbor, Regression* dalam memprediksi penyakit ginjal kronis. *Dataset* penyakit ginjal kronis pada penelitian ini diperoleh dari *UCI Machine Learning Repository website*. Atribut dalam *dataset* terdiri dari 2 tipe yaitu nominal dan numerik, antara lain *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatine, sodium, potassium, hemoglobin, packed cell volume, white blood*

cell, red blood cells, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal, edema, anemia, serta dengan 2 *class* yaitu *ckd* dan *not ckd*. Dimana terdapat 400 data dan 25 atribut. Data dikumpulkan dalam waktu 2 bulan. Hasil dari penelitian ini yaitu Metode *Support Vector Machine* memperoleh *accuracy* 90.25, *precision* 44.79, *recall* 46.37. Metode *Decision Tree* memperoleh *Accuracy* 95.50, *precision* 63.41, *recall* 63.78. Metode *k-Nearest neighbor* memperoleh *accuracy* 94.75, *precision* 62.73, *recall* 63.91. Metode *Regression* memperoleh *Accuracy* 98.25, *precision* 99.60, *recall* 97.60.

Pada penelitian dengan judul “*Machine Learning Algorithm Selection For Chronic Kidney Disease Diagnosis and Classification*” oleh (Gokiladevi et al., 2022). Penelitian ini bertujuan menyelidiki kinerja diagnosis *ckd* dari berbagai model *machine learning*, model tersebut yaitu *K-NN, SVM, Random Forest, Logistic Regression*. Pada penelitian ini data yang digunakan diperoleh dari situs penyedia data untuk keperluan penelitian pada set data *benchmark CKD* yang diterapkan dari UCI. Sebanyak 400 data yang terdiri dari 24 atribut dan 2 kelas. Dari 400 data terdapat 62.50% kasus kelas positif *ckd* dan 37.50% kasus kelas negative *ckd*. Hasil dari penelitian ini yaitu dengan metode *K-NN* diperoleh *precision* 0.79, *recall* 0.68, dan *F-Score* 0.68. Metode *SVM* memperoleh *precision* 0.85, *recall* 0.74, *F-Score* 0.74. Metode *Random Forest* memperoleh *precision* 0.99, *recall* 0.99, *F-Score* 0.99. Metode *Logistic Regression* memperoleh *precision* 0.79, *recall* 0.96, *F-Score* 0.96. Dapat disimpulkan bahwa Metode *Random Forest* yang paling unggul dibandingkan model lainnya.

Penelitian terdahulu berjudul “*Chi2 -MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease*” oleh (Kumar et al., 2022). Keunikan dari penelitian ini adalah mengembangkan sistem diagnostik untuk mendeteksi penyakit ginjal kronis menggunakan berbagai algoritme *machine learning* (ML). Ini mendukung pendekatan pemilihan fitur hybrid. Penelitian ini menggunakan 400 data klinis dari pasien *CKD*, dimana data terdiri dari 25 atribut yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium*,

hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, dan anemia, serta atribut kelas. Data diperoleh dari kumpulan data yang disediakan oleh *University of California, Irvine (UCI)*. Berbagai teknik persiapan data seperti pengkodean fitur kategorikal, nilai yang hilang, imputasi, penghapusan outlier, mengatasi ketidakseimbangan data, menskalakan data ke tingkat yang sama, dan seleksi fitur yang relevan diambil untuk mempersiapkan *dataset* untuk model prediktif. Uji Hibrida Chi-Square (Chi²) Pendekatan pemilihan fitur berbasis *mutual information (MI)* telah diusulkan untuk menghapus fitur yang berlebihan.

Metode yang digunakan yaitu *Ada Boost, Bagging, CatBoost, Extra Trees, Gaussian Naive Bayes, Gradient Boosting, K-nearest neighbors, LightGBM, Multi-Layer Perceptron, Random Forest, Stochastic Gradient Boosting, Support vector machines, XGBoost*. Dari hasil penelitian ini diperoleh hasil yaitu dengan metode *Ada Boost* diperoleh *precision 98%, recall 96 %, F1-score 97%*. Metode *Bagging* memperoleh *precision 30.00%, recall 50.00%, F1-score 37.00%*. Metode *CatBoost* memperoleh *precision 96%, recall 95%, F1-score 96%*. Metode *Extra Trees* memperoleh *precision 99%, recall 98%, F1-score 98%*. Metode *Gaussian Naive Bayes* memperoleh *precision 93.00%, recall 93.00%, F1-score 93.00%*. Metode *Gradient Boosting* memperoleh *precision 95.00%, recall 88.00%, F1-score 93.00%*. Metode *K-nearest neighbors* memperoleh *precision 66.00%, recall 67.00%, F1-score %*. Metode *LightGBM* memperoleh *precision 96%, recall 95%, F1-score 96%*. Metode *Multi-Layer Perceptron* memperoleh *precision 89.00%, recall 89.00%, F1-score 89.00%*. Metode *Random Forest* memperoleh *precision 97.00%, recall 96.00%, F1-score 96.00%*. Metode *Stochastic Gradient Boosting* memperoleh *precision 94%, recall 92%, F1-score 93%*. Metode *Support vector machines (SVM)* memperoleh *precision 98.00%, recall 97.00%, F1-score 97.00%*. Metode *XGBoost* memperoleh *precision 95%, recall 94%, F1-score 95%*.

Penelitian terdahulu berjudul “*Prediction on Cardiovascular disease using Decision tree and Naïve Bayes*” oleh (Krishna et al., 2022). Penelitian ini dilakukan untuk

membantu memprediksi penyakit kardiovaskular dengan akurasi lebih baik yaitu dengan menerapkan teknik ML seperti *Decision Tree* dan *Naive Bayes*. *Dataset* yang digunakan pada penelitian ini berasal dari Kaggle, yaitu *dataset* gagal jantung. Sebanyak 299 spesimen dan 13 atribut. Dari 13 atribut, 6 adalah atribut biner dan tujuh atribut yang tersisa adalah non-biner. Atribut hasil adalah DEATH_EVENT. Potensi gagal jantung pada manusia. Hasil dari penelitian ini yaitu dengan Metode *Decision Tree* memperoleh *accuracy* 82% dan Metode *Gaussian Naïve Bayes* memperoleh *accuracy* 86 %.

Penelitian terdahulu berjudul “*Prediction of Chronic Kidney Disease Using Machine Learning Algorithm*” oleh (Tekale et al., 2022). Penelitian ini menggunakan *dataset* dari *uci repository*. *Dataset* berisi 400 catatan pasien yang mengandung 25 atribut, tetapi pada penelitian ini hanya 14 atribut yang digunakan untuk membangun model. Metode yang digunakan yaitu *Decision Tree* dan *Support Vector Machine* (SVM). Pada penelitian ini menjelajahi berbagai algoritme pembelajaran mesin. Menganalisis 14 atribut berbeda yang terkait, keakuratan prediktif pasien CKD dan berbagai algoritme pembelajaran mesin seperti *Decision Tree* dan *Support Vector Machine* (SVM). Dari hasil analisis algoritme *Decision Tree* diperoleh *Precision* 85%, *Recall* 95%, *F1 Measure* 89% dan Metode *Support Vector Machine* (SVM) memperoleh *Precision* 93%, *Recall* 98%, *F1 Measure* 95%.

Penelitian terdahulu yang berjudul “*Prediction of Chronic Kidney Disease Using Data Mining Feature Selection and Ensemble Method*” oleh (Zeynu, 2018). Dalam penelitian ini metode yang digunakan yaitu *K-Nearest Neighbor*, *Artificial Neural Network*, *Naïve Bayes* and *Support Vector Machine* untuk mendiagnosis penyakit ginjal dengan membangun dua model penting untuk memprediksi penyakit ginjal kronis yaitu, bagaimana memilih fitur dan ansambel model. *Dataset* pada penelitian ini dikumpulkan dari *uci repository*. Kumpulan data berisi 400 data dari 24 atribut dan satu atribut kelas. *Dataset* berisi 400 data (250 CKD, 150 not CKD) dan jumlah atribut. 24 + kelas = 25 (Angka 11, Nominal 14). Hasil klasifikasi dengan metode *KNN* memperoleh *accuracy* 98.5%. Metode *ANN* memperoleh *accuracy* 97.75. Metode *NB*

memperoleh *accuracy* 94.5%. Metode SVM memperoleh *accuracy* 97.75%. Metode *Ensemble model* memperoleh *accuracy* 99%.

Penelitian terdahulu berjudul “Penerapan Algoritme *Naïve Bayes* Untuk Diagnosis Penyakit Gagal Ginjal Kronis” dilakukan oleh (Faddillah et al., 2019). Penelitian ini menjelaskan penerapan metode algoritme *Naïve Bayes* untuk mendiagnosis atau memprediksi gagal ginjal kronis. Sistem ini didasarkan pada *dataset Indians Chronic Kidney Disease (CKD)* yang diperoleh dari situs web *UCI Machine Learning Repository* yang dibuat pada tahun 2015. Hasil penelitian ini mencapai akurasi 91.25% dari 400 *dataset* yang diuji. 400 *dataset* tersebut terdiri dengan 24 atribut yang digunakan sebagai referensi. Atribut yang digunakan yaitu *age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatine, sodium, potassium, hemoglobin, packed cell volume, white blood cell, red blood cells, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal, edema* serta *anemia*.

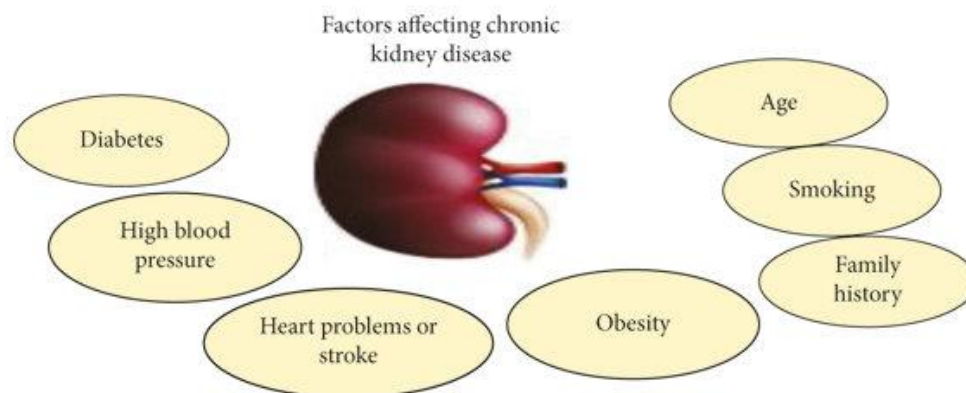
Penelitian selanjutnya berjudul “Perbandingan Metode *Data Mining* SVM dan NN Untuk Klasifikasi Penyakit Ginjal Kronis” dilakukan oleh (Amalia, 2018). Penelitian ini melakukan pengolahan *dataset* penyakit ginjal kronis dari *UCI repository*. Data yang digunakan yaitu data sekunder dan data yang digunakan sebanyak 400 *record* dan terdiri 24 atribut dan satu label. Para peneliti memproses data untuk menemukan model dengan skor akurasi yang lebih tinggi untuk *dataset* penyakit ginjal kronis. Hasil penelitian menunjukkan bahwa metode *Neural Network* mencapai nilai akurasi sebesar 93.36% dan metode *Support Vector Machine (SVM)* mencapai nilai akurasi sebesar 95.16%.

Selanjutnya penelitian yang dilakukan oleh (Ariani & Samsuryadi, 2019) dengan judul “Klasifikasi Penyakit Ginjal Kronis Menggunakan *K-Nearest Neighbor*”. Penelitian ini menggunakan *dataset* sekunder, *dataset* penyakit ginjal kronis yang terdiri dari 400 *record* dengan 24 fitur, dan memiliki 2 kelas yaitu penyakit ginjal kronis dan penyakit ginjal non-kronis. *Dataset* penyakit ginjal kronis diperoleh dari *Repository University*

of California (UCI Repository Machine Learning Benchmark). Dari hasil klasifikasi mencapai akurasi yang cukup tinggi yaitu 85,83%.

2.2. Penyakit Ginjal Kronis

Ginjal merupakan sepasang organ yang memiliki peran penting. Untuk Setiap ginjal berukuran sekitar 10-15 cm, beratnya 150 gram dan terletak di sisi kanan dan kiri perut bagian belakang. Ginjal memiliki beberapa fungsi yaitu menyaring darah dan hasil metabolisme dengan membuang kelebihan cairan dalam tubuh manusia, penyesuaian keseimbangan cairan dan elektrolit seperti natrium dan kalium, merangsang pembentukan sel darah merah serta mengatur tekanan darah. Ginjal menyaring 190 liter darah setiap harinya (Kusuma et al., 2019).



Gambar 1. Faktor pengaruh penyakit ginjal kronis (Senan et al., 2021).

Pada Gambar 1 merupakan faktor yang mempengaruhi penyakit ginjal kronis yaitu diabetes merupakan salah satu faktor utama yang berkontribusi terhadap penyakit ginjal kronis. Jika kadar gula darah tinggi dalam jangka waktu yang lama, dapat menyebabkan kerusakan pada pembuluh darah kecil di ginjal, yang pada akhirnya dapat mengakibatkan kerusakan ginjal. Tekanan darah tinggi atau hipertensi dapat

menyebabkan kerusakan pada pembuluh darah ginjal. Jika tekanan darah tidak terkontrol dengan baik, ginjal akan bekerja keras dan menyebabkan kerusakan pada fungsi ginjal. Gangguan jantung seperti gagal jantung atau stroke dapat mempengaruhi aliran darah ke ginjal. Jika pasokan darah ke ginjal tidak mencukupi, maka ginjal dapat mengalami kerusakan yang pada akhirnya berujung pada penyakit ginjal kronis. Kelebihan berat badan atau obesitas juga menjadi faktor risiko dalam perkembangan penyakit ginjal kronis. Obesitas dapat meningkatkan tekanan darah dan menyebabkan resistensi insulin, yang pada gilirannya dapat merusak ginjal. Faktor keturunan juga dapat memainkan peran dalam meningkatkan risiko seseorang terkena penyakit ginjal kronis. Jika ada keturunan keluarga yang memiliki penyakit ginjal kronis, maka risiko seseorang untuk mengembangkan penyakit ini juga meningkat. Kebiasaan merokok merupakan faktor risiko tambahan yang dapat merusak pembuluh darah, termasuk pembuluh darah di ginjal. Merokok dapat mempercepat kerusakan ginjal dan memperburuk kondisi penyakit ginjal kronis. Peningkatan usia juga menjadi faktor risiko dalam perkembangan penyakit ginjal kronis. Semakin tua seseorang, semakin besar kemungkinan mereka mengalami penurunan fungsi ginjal. Semua faktor ini dapat berinteraksi satu sama lain dan meningkatkan risiko seseorang terkena penyakit ginjal kronis. Penting untuk memahami faktor-faktor ini dan mengambil langkah-langkah pencegahan yang tepat untuk menjaga kesehatan ginjal.

Penyakit ginjal kronis didefinisikan sebagai kerusakan ginjal yang berkelanjutan jangka panjang (menahun/kronis), ditandai dengan penurunan kinerja ginjal menyaring darah (Laju Filtrasi Glomerulus/LFG). Pasien PGK Seringkali tidak ada gejala atau tanda sampai fungsi ginjal terjaga kurang dari 15%. (Kusuma et al., 2019).

2.3. Data Mining

Data Mining, juga dikenal sebagai *knowledge discovery in database (KDD)*, adalah aktivitas yang melibatkan pengumpulan dan penggunaan data historis untuk menemukan keteraturan, pola, atau hubungan dalam kumpulan data besar. Hasil dari *data mining* ini dapat digunakan untuk meningkatkan pengambilan keputusan di masa

mendatang (Santosa, 2007). *Data mining* melakukan proses menganalisis dan mengekstraksi pengetahuan secara otomatis menggunakan satu atau lebih teknik pembelajaran komputer (*Machine Learning*) (Hermawati, 2013).

Metode belajar (*Learning*) secara garis besar dibagi menjadi dua pendekatan yaitu *supervised learning* dan *unsupervised learning*. *Supervised Learning* adalah metode pembelajaran dengan pelatihan dan pelatih. Dalam *supervised learning* fungsi atau model di uji dengan set data validasi. Dalam pengujian beberapa membagi data menjadi dua kelompok, set *training* dan set *testing*. Ada juga yang membagi data menjadi tiga bagian, set *training*, validasi, dan *testing*. Fungsi validasi dan *testing* sebenarnya sama. Dengan kata lain, ini menguji apakah suatu fungsi atau model cukup baik bila digunakan untuk prediksi. Jika kinerja selama validasi atau pengujian tidak memenuhi harapan, atur ulang dan sesuaikan nilai parameter model untuk meningkatkan kinerja model. Setelah fungsi-fungsi tersebut ditemukan dan kinerjanya cukup bagus, barulah bisa melakukan tugas mengelompokkan objek tertentu yang belum diketahui spesifikasinya atau membuat prediksi. Metode *unsupervised learning* metode diterapkan tanpa adanya latihan (*training*) dan tanpa guru (*teacher*) atau label. *Unsupervised learning* membutuhkan data untuk dikelompokkan ke dalam kelas. Hal ini karena data belum memiliki label. Label menunjukkan di mana data akan dikelompokkan. Kelompok metode *unsupervised learning* ini termasuk *clustering* dan *self-organizing* (SOM)(Santosa, 2007).

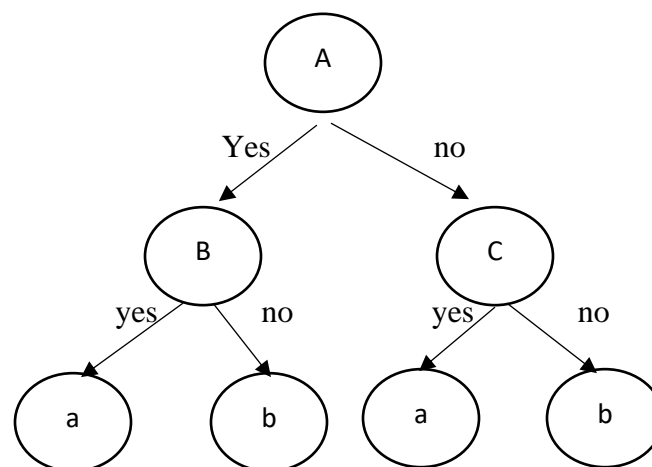
2.4. Klasifikasi

Penggunaan teknologi tidak dapat dipisahkan dari aktivitas manusia sehari-hari. Komputer adalah salah satu teknologi yang digunakan untuk mengolah atau menghitung data. Ada banyak metode teknologi dalam memproses atau menghitung data. Metode ini dirancang untuk memfasilitasi pemrosesan data, tergantung pada jenis data yang diproses. Salah satu metode pengolahan data adalah klasifikasi (Wibawa et al., 2018).

Klasifikasi merupakan pengelompokan *record* data baru ke salah satu dari beberapa kelas yang sebelumnya telah didefinisikan (Hermawati, 2013). Pengelompokan data didasarkan sepenuhnya pada ciri-ciri yang dimiliki sesuai dengan kelasnya. Dalam prosesnya, klasifikasi dapat diselesaikan dengan banyak cara baik secara manual maupun dengan bantuan teknologi. Klasifikasi yang diselesaikan secara manual adalah klasifikasi yang diselesaikan dengan bantuan manusia tanpa bantuan algoritme komputer. Klasifikasi dengan bantuan teknologi, memiliki banyak algoritme, termasuk *naïve bayes*, *support vector machine*, *decision tree*, *fuzzy* dan jaringan syaraf sintetik (Wibawa et al., 2018).

2.5. Decision Tree

Metode *decision tree* merupakan salah satu metode teknik klasifikasi yang ada pada *data mining*. *Decision tree* adalah algoritme yang paling banyak digunakan untuk masalah klasifikasi. Metode *decision Tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Pohon keputusan juga berguna untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sekumpulan kandidat variabel input dan variabel target. *Decision tree* terdiri dari beberapa simpul *tree's root*, *internal node* dan *leafs*.



Gambar 2. Model susunan *decision tree*.

Pada Gambar 2 Konsep *entropi* digunakan untuk menentukan atribut mana yang dipisahkan oleh pohon. Dalam pohon keputusan, setiap simpul internal membagi ruang menjadi dua bagian atau lebih sesuai dengan fungsi diskrit dari nilai atribut input. Dalam kasus yang paling sederhana dan umum, setiap pengujian menganggap semua atribut tunggal, sehingga ruang di partisi kosong disesuaikan dengan nilai atribut. Klasifikasi menggunakan pohon keputusan dilakukan dengan melakukan *routing* dari simpul akar hingga mencapai simpul daun. Algoritme pohon keputusan meliputi ID3, C4.5, dan CART (Wibawa et al., 2018).

Algoritme C4.5 diperkenalkan oleh Quinlan sebagai versi perbaikan dari ID3. Dalam ID3, induksi *decision tree* hanya dapat dilakukan pada jalur bertipe kategorikal (nominal/ordinal), sedangkan tipe numerik (*internal/rasio*) tidak dapat menangani fitur yang melibatkan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (deriving) *rule set*. Algoritme C4.5 juga menggunakan kriteria gain dalam menemukan fitur yang menjadi pemecah simpul untuk pohon yang diinduksi (Muslim et al., 2019).

Ada beberapa langkah dalam algoritme untuk membangun pohon keputusan, Larose yaitu (Bahri & Lubis, 2020):

1. Menyiapkan data training. Data *training* diambil dari data histori yang pernah terjadi sebelumnya dan telah dikelompokkan pada kelas-kelas tertentu.
2. Hitung akar dari pohon. Akar diambil dari atribut yang terpilih, dengan menghitung nilai gain dari setiap atribut, nilai gain tertinggi menjadi akar pertama. Untuk menghitung nilai gain dari atribut, hitung terlebih dahulu nilai *entropy*.

Perhitungan Nilai *Entropy* menggunakan Persamaan 1 berikut:

$$entropy(s) = \sum_{i=1}^n -p_i * \log_2 p_i \dots \dots \dots (1)$$

Keterangan :

S : Himpunan Kasus

A : Atribut

n : Jumlah Partisi S

Pi : Proporsi dari Si terhadap S

3. Menghitung nilai Gain

Perhitungan Gain menggunakan Persamaan 2 berikut:

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

Keterangan:

S = Himpunan Kasus

A = Atribut/Fitur

N = Jumlah Partisi atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

4. Ulangi langkah 2 dan 3 hingga semua terpartisi.

5. Proses partisi pohon keputusan berakhir ketika:

- a) Semua *record* dari node N diberi kelas yang sama
- b) Tidak ada atribut lain dalam *record* yang dipartisi lagi. Dalam cabang kosong tidak ada *record* di dalamnya.

2.5.1. Kelebihan Metode *Decision Tree*

Adapun kelebihan Metode *Decision Tree* menurut (Wibawa et al., 2018) sebagai berikut:

1. Data yang lebih akurat.
2. Meningkatkan efisiensi perhitungan.
3. Menghindari kehilangan atribut terus menerus.

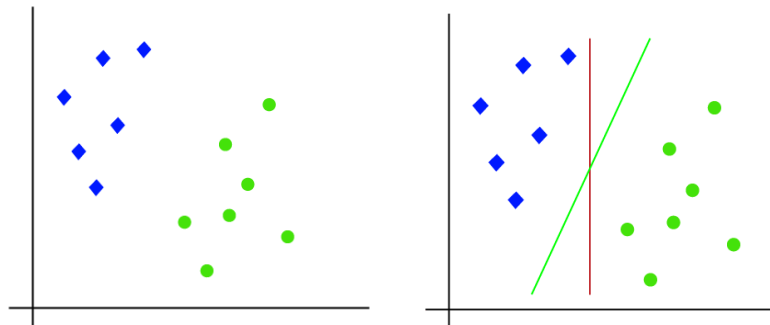
2.5.2. Kekurangan Metode *Decision Tree*

Adapun kekurangan Metode *Decision Tree* menurut (Wibawa et al., 2018) sebagai berikut:

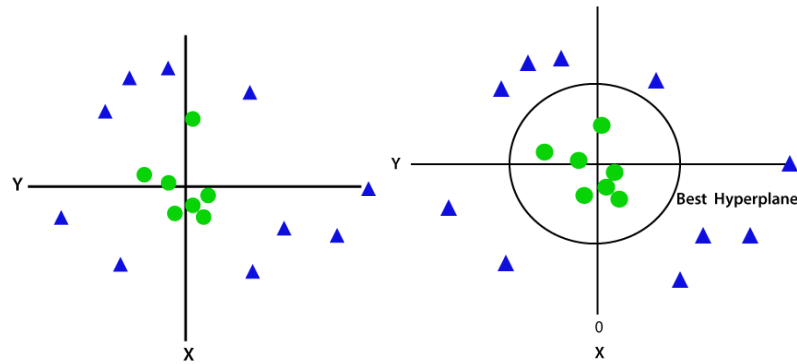
1. Percabangan bisa kosong.
2. Percabangan tidak signifikan.

2.6. *Support Vector Machine (SVM)*

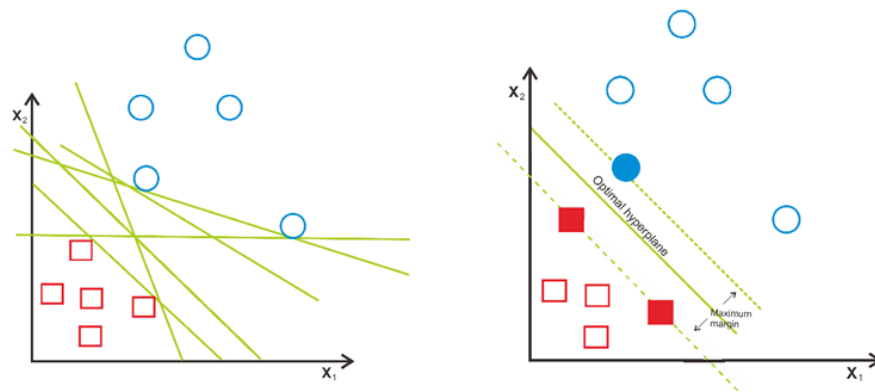
Metode *Support Vector Machine (SVM)* pertama kali dikenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian dari konsep-konsep unggulan bidang *pattern recognition*. *Support Vector Machine (SVM)* adalah sistem pembelajaran dengan ruang hipotesis berdasarkan fungsi linier pada sebuah fitur berdimensi tinggi dan dilatih dengan algoritme pembelajaran berdasarkan teori optimasi. Tingkat akurasi dalam model dihasilkan oleh proses peralihan dengan SVM sangat tergantung pada fungsi kernel dan parameter yang digunakan. Berdasarkan karakteristiknya, metode SVM dibagi menjadi dua yaitu Linear SVM dan SVM non-linear. SVM linier adalah data dipisahkan secara linier dengan memisahkan kelas pada *hyperplane* dengan *soft margin*. Sedangkan SVM Non-linier adalah mengimplementasikan fungsi kernel trick dengan ruang berdimensi tinggi (Puspitasari et al., 2018). Dapat dilihat pada Gambar 3 dan 4 contoh SVM linear dan non-linear.



Gambar 3. SVM linear (Trivusi, 2022)



Gambar 4. SVM non-linear



Gambar 5. Usaha menemukan *hyperplane* terbaik (Wibawa et al., 2018).

Gambar 5 adalah konsep SVM adalah menemukan pemisah yang optimal pada *hyperplane*. *Hyperplane* pemisah terbaik ditemukan dengan mencari nilai $f(x)$ pada *margin hyperplane* (Kurnianto et al., 2018). Menghitung nilai $f(x)$ menggunakan Persamaan 3 berikut:

$$f(x) = \text{sign} (\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b) \dots \dots \dots (3)$$

Mencari nilai σ menggunakan metode *sequential training*. Kemudian menghitung nilai matriks *hessian* dengan disimbolkan dengan $[D]$ dan melakukan iterasi σ (Kurnianto et al., 2018). Menghitung matriks *hessian* menggunakan Persamaan 4 berikut:

$$[D]_{IJ} = y_i y_j (K(x_i, x_j) + \lambda^2) \dots \dots \dots (4)$$

Iterasi pelatihan α dihitung dengan Persamaan 5, 6, 7 berikut:

$$[D]_{IJ} = \sum_{j=1}^n \alpha_j D_{ij} \dots \dots \dots (5)$$

$$\delta \alpha_i = \min(\max[y(1 - E_i), -\alpha_i], C - \alpha_i) \dots \dots \dots (6)$$

$$\alpha_i = \alpha_i + \delta \alpha_i \dots \dots \dots (7)$$

2.6.1. Kelebihan Metode *Support Vector Machine* (SVM)

Adapun kelebihan Metode *Support Vector Machine* menurut (Wibawa et al., 2018) sebagai berikut:

1. Pengklasifikasi SVM memberikan akurasi tinggi dan bekerja dengan baik. Di ruang berdimensi tinggi.
2. Mudah diimplementasikan (Wibawa et al., 2018).

2.6.2. Kekurangan Metode *Support Vector Machine* (SVM)

Adapun kekurangan Metode *Support Vector Machine* menurut (Wibawa et al., 2018) sebagai berikut:

1. Sulit dipakai dalam problem berskala besar, dalam skala besar yang dimaksudkan adalah jumlah sampel yang diolah.
2. Jenis kernel svm berpengaruh pada akurasi sistem.

2.7. Metode *Naïve Bayes*

Naïve Bayes dikenal sebagai teorema Bayes karena merupakan metode probabilistik dan statistik dari ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang masa

depan berdasarkan pengalaman masa lalu. Dalam interpretasi *Bayes*, teorema ini menyatakan seberapa besar tingkat keyakinan subjektif harus berubah secara wajar dengan adanya isyarat baru. *Bayes* adalah teknik prediksi dengan probabilistik sederhana yang merupakan dasar dari penerapan teorema Bayes (aturan *Bayes*), dan asumsi independensi (ketidaktergantungan) yang kuat. Model yang digunakan pada *naïve bayes* adalah model fitur independent. Hubungkan *naïve bayes* dengan klasifikasi, *korelasi* hipotesis dan *evidence* yaitu hipotesis dalam teorema bayes adalah label kelas untuk menjadi target pemetaan dalam klasifikasi, sedangkan *evidence* merupakan fitur-fitur untuk menjadi masukan dalam sebuah model klasifikasi. Jika X merupakan *vector* masukan yang berisi fitur, dan Y adalah label kelas, jadi *naïve bayes* dituliskan dengan notasi $P(Y|X)$ yang berarti probabilitas label kelas Y didapatkan setelah fitur -fitur X diamati, notasi ini disebut probabilitas akhir (*posterior probability*) untuk Y. Probabilitas awal $P(Y)$ (*prior probability*) Y. Pembelajaran probabilitas akhir ($P(Y|X)$) harus dilakukan saat proses pelatihan pada model untuk setiap kombinasi X dan Y berdasarkan informasi dari data latih (Prasetyo et al., 2013).

Perhitungan *Naïve Bayes* pada Persamaan 8 (Zeynu, 2018):

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)} \dots \dots \dots (8)$$

Y : kelas spesifik

X : Data yang belum diketahui

$P(Y|X)$: Probabilitas hipotesis Y berdasarkan kondisi X

$P(Y)$: Hipotesis Y

$P(X|Y)$: Probabilitas X berdasarkan kondisi hipotesis Y

$P(X)$: Probabilitas X

2.7.1. Kelebihan Metode *Naïve Bayes*

Adapun kelebihan Metode *Naïve Bayes* menurut (Wibawa et al., 2018) sebagai berikut:

1. *Naïve Bayes* masih bekerja dengan baik saat menjalankan pengujian pada tipe data kategorikal.
2. Semua atribut bersifat independen satu sama lain tergantung dari nilai variabel kelasnya.
3. Performa baik.

2.7.2. Kekurangan Metode *Naïve Bayes*

Adapun kekurangan Metode *Naïve Bayes* (Wibawa et al., 2018) sebagai berikut:

1. Sangat sensitif dan kurang akurat pada fitur yang terlalu banyak.
2. Ukuran vektor fitur yang dihasilkan sangat besar, sehingga membutuhkan teknik untuk memperkecil ukuran vektor tersebut.

2.8. Evaluasi Kinerja Klasifikasi

Machine learning adalah cabang kecerdasan buatan (*artificial intelligence*) yang membangun sistem berdasarkan data. *Machine learning* adalah proses pembelajaran komputer dari data (*learn from data*). Tanpa data, komputer tidak dapat belajar. Salah satu aplikasi *machine learning* adalah *supervised learning*. Klasifikasi adalah *supervised learning*, model prediksi dengan hasil prediksi diskrit. Cara mengukur kinerja model klasifikasi adalah membandingkan nilai aktual dan prediksi. *Confusion matrix* adalah pengukuran kinerja untuk masalah klasifikasi *machine learning* yang keluarannya dua kelas atau lebih. *Confusion matrix* adalah tabel dengan empat kombinasi nilai prediksi dan aktual yang berbeda. *Confusion matrix* memiliki empat istilah yang menggambarkan hasil dari proses klasifikasi yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative* (Anggreany, 2022).

Tabel 2. *Confusion matrix* (Desiani, 2022).

Kelas	Prediksi <i>Positive</i>	Prediksi <i>Negative</i>
Aktual <i>Positive</i>	TP	FN
Aktual <i>Negative</i>	FP	TN

Tabel 3. Prediksi.

Hasil Prediksi	Keterangan
<i>True Positive (TP)</i>	Prediksi <i>positive</i> dan itu benar.
<i>True Negative (TN)</i>	Prediksi <i>negative</i> dan itu benar.
<i>False Positive (FP)</i>	Prediksi <i>positive</i> dan itu salah.
<i>False Negative (FN)</i>	Prediksi <i>negative</i> dan itu salah.

Nilai prediksi adalah output dari program, nilainya *positive* dan *negative*. Nilai aktual adalah nilai sebenarnya yaitu *True* dan *False*.

Confusion matrix terdapat beberapa perhitungan di dalamnya yaitu *Accuracy*, *Precision*, *Recall* atau *Sensitivity*, dan *F-1 Score*.

2.8.1. *Accuracy*

Accuracy menjelaskan seberapa akurat model dalam mengklasifikasikan dengan benar (Anggreany, 2022). Perhitungan *Accuracy* menggunakan Persamaan 10:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100\% \dots\dots\dots(10)$$

2.8.2. Precision

Precision menggambarkan keakuratan antara data yang diminta dan hasil prediksi yang diberikan oleh model (Anggreany, 2022). Perhitungan *Precision* menggunakan Persamaan 11:

$$Precision = \frac{(TP)}{(TP+FP)} \times 100\% \dots \dots \dots (11)$$

2.8.3. Recall atau Sensitivity

Recall atau *Sensitivity* menggambarkan keberhasilan model dalam mengambil informasi (Anggreany, 2022). Perhitungan *Recall* atau *Sensitivity* menggunakan Persamaan 12:

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \dots \dots \dots (12)$$

2.8.4. F-1 Score

F-1 Score mewakili perbandingan rata-rata *precision* dan *recall* yang dibobotkan. *Accuracy* tepat digunakan sebagai acuan kinerja algoritme jika *dataset* memiliki jumlah data *false negative* dan *false positive*. Tapi jika jumlahnya tidak mendekati, maka menggunakan *F-1 Score* sebagai acuan (Anggreany, 2022). Perhitungan *F-1 Score* menggunakan Persamaan 13:

$$F - 1 \text{ Score} = 2 \times \frac{(Recall \times Precision)}{(Recall+Precision)} \times 100\% \dots \dots \dots (13)$$

2.8.5. K-Fold Cross Validation

K-fold cross validation melibatkan eksperimen validasi di mana pelatihan, validasi, dan pengujian dilakukan. Pada eksperimen pertama, 90% *dataset* yang dipilih secara acak

digunakan untuk pelatihan, sementara 10% sisanya digunakan untuk pengujian. Pada eksperimen kedua, set 90% *dataset* yang benar-benar berbeda secara acak dipilih untuk pelatihan, sementara 10% sisanya digunakan untuk pengujian. Proses ini diulang dengan *set* yang berbeda dari 90% *dataset* pelatihan dan 10% *dataset* pengujian, seperti yang ditunjukkan dalam Tabel 4, di mana total 5 eksperimen dilakukan secara berurutan. Dengan asumsi bahwa pemilihan *dataset* pelatihan dan pengujian benar-benar acak dan proses *k-fold cross validation* adalah *ergodik*, *output* yang benar didapatkan dengan mengambil rata-rata dari *output* semua eksperimen (Ahmed et al., 2018).

Tabel 4. *K-Fold Cross Validation*

Percobaan 1	<i>Test</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>
Percobaan 2	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>
Percobaan 3	<i>Train</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Train</i>
Percobaan 4	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>
Percobaan 5	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Train</i>	<i>Test</i>

3.2. Data dan Alat

Adapun data dan alat yang digunakan untuk penelitian ini sebagai berikut:

3.2.1. Data

Data pada penelitian ini yaitu didapatkan dari situs *Kaggle* pada laman: <https://www.kaggle.com/code/niteshyadav3103/chronic-kidney-disease-prediction-98-accuracy>. *Dataset* ini diambil dari akun *Kaggle* yang dikembangkan oleh Nitesh Yadav, seorang Data Science Intern di Celebal Technologies dari India. Data berupa file CSV yang di dalamnya terdapat 400 *dataset* yang terdiri dari 24 atribut. Berikut dapat dilihat pada Tabel 6 penjelasan secara detail atribut dari penyakit ginjal kronis.

Tabel 6. Penjelasan atribut (Kumar et al., 2022)

No	Atribut	Tipe Atribut	Penjelasan
1	Umur	Numerik	Usia individu dalam tahun.
2	Tekanan Darah	Numerik	Tekanan darah diukur dalam mm/Hg.
3	Gravitas	Nominal	Berat jenis.
4	Albumin	Nominal	Nilai Albumin Nominal dalam Urin.
5	<i>Sugar</i>	Nominal	Nilai Gula Nominal dalam darah.
6	Sel Darah Merah	Nominal	Sel Darah Merah: Normal atau Abnormal.
7	Pussel	Nominal	Sel nanah: Normal atau abnormal.
8	Pus cells	Nominal	Gumpalan sel nanah: ada atau tidak ada.
9	Bakteri	Nominal	Bakteri: ada atau tidak ada.

No	Atribut	Tipe Atribut	Penjelasan
10	GDS	Numerik	Glukosa Darah Acak diukur dalam mg/dl.
11	<i>Ureum</i>	Numerik	Urea darah (mg/dl).
12	<i>Kreatinin</i>	Numerik	Nilai Serum Kreatinin (mgs/dl) dalam darah.
13	<i>Natrium</i>	Numerik	Natrium (mEq/L).
14	<i>Kalium</i>	Numerik	Kalium (mEq/L).
15	Hemoglobin	Numerik	Hemoglobin (gm).
16	MCV	Numerik	Volume Sel yang Dikemas.
17	Jumlah Sel Darah Putih	Numerik	Jumlah Sel darah putih (Cells/cumm).
18	Jumlah Sel Darah Merah	Numerik	Jumlah Sel Darah Merah (jutaan/cumm).
19	<i>Hipertensi</i>	Nominal	Hipertensi: Ya atau Tidak.
20	Diabetes	Nominal	Diabetes Mellitus: Ya atau Tidak.
21	cad	Nominal	Penyakit Arteri Koroner: Ya atau Tidak.
22	Nafsu Makan	Nominal	Nafsu makan: Baik atau Buruk.
23	Edema	Nominal	Pedal Edema: Ya atau Tidak.
24	Anemia	Nominal	Anemia: Ya atau Tidak.
25	Klasifikasi	Nominal	Kolom Target: Diklasifikasikan sebagai ckd atau not ckd.

3.2.2. Alat

Adapun alat pendukung dalam penelitian ini terdiri dari perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai berikut:

3.2.2.1. Perangkat Keras (*Hardware*)

Perangkat keras (*hardware*) yang digunakan pada penelitian ini yaitu sebuah laptop dengan spesifikasi sebagai berikut:

- a. *System Manufacture* : *Lenovo*
- b. *Processor* : *Intel Core i5-4210U*
- c. *CPU* : *1.70GHz 2.39 GHz*
- d. *Installed RAM* : *8,00 GB*
- e. *System type* : *64-bit operating system*

3.2.2.2. Perangkat Lunak (*Software*)

Perangkat lunak (*software*) pada penelitian ini sebagai berikut:

- a. Sistem Operasi : Windows 10 64 bit.
- b. Python 3.10.6
Python adalah bahasa pemrograman komputer interpretatif berorientasi objek, tingkat tinggi, yang ditemukan oleh Guido van Rossum pada tahun 1989 dan pertama kali dirilis ke publik pada tahun 1991 (Chen & Liu, 2022).
- c. *Anaconda Navigator*
Anaconda adalah platform untuk meningkatkan aset, berkolaborasi, dan meluncurkan proyek ilmiah. *Anaconda Navigator* adalah antarmuka pengguna grafis (GUI) yang dapat digunakan untuk mengelola paket yang menjalankan aplikasi dan menggunakan *library* dalam kode program yang membutuhkan data.

d. *Jupyter Notebook*

Jupyter adalah singkatan dari tiga bahasa pemrograman yaitu Julia (Ju), *Python* (Py), dan R. *Jupyter* adalah aplikasi web gratis untuk membuat dan berbagi dokumen yang berisi kode, komputasi, visualisasi, dan teks (Medina, 2022).

e. *Excel*

Microsoft Excel, atau biasa disebut dengan *Excel*, merupakan aplikasi perangkat lunak yang biasa digunakan untuk mengolah kumpulan data secara otomatis dalam berbagai format, mulai dari membuat rumus, perhitungan dasar dan tabel hingga pengelolaan data. *Excel* pertama kali muncul pada tahun 1985 dan terus berkembang hingga saat ini (Eugenia, 2021).

f. *Web Browser*

Web Browser disebut juga perambah merupakan perangkat lunak dengan fungsi melihat dan berinteraksi dengan dokumen yang disediakan oleh *server web*. *Google chrome* adalah aplikasi peramban yang digunakan untuk jelajah dunia maya seperti *Firefox*, *Opera*, dan *Microsoft Edge*. *Google Chrome* dibuat dan dirancang oleh *Google*. *Web Browser* digunakan untuk menampilkan hasil *website* yang dibuat. *Mozilla Firefox*, *Google Chrome* dan safari adalah *Web Browser* yang paling sering digunakan (Abdulghani & Gozali, 2020).

3.2.2.3. *Library*

Adapun *library* yang digunakan untuk penelitian ini sebagai berikut:

a. *Library Numpy*

Numpy, singkatan dari *Numerical Python*, merupakan *library* yang banyak digunakan dalam proses analisis data. *Numpy* dapat digunakan untuk operasi vektor dan matriks dan dapat mengelola *array*. Salah satu keunggulan *numpy* adalah dapat diintegrasikan dengan beberapa Bahasa pemrograman lain (Khasanah, 2021).

b. *Matplotlib*

Matplotlib adalah *library* dalam *python* yang digunakan untuk visualisasi data. Visualisasi data adalah salah satu langkah kunci dalam membuat hasil analisis lebih mudah dipahami. *Matplotlib* berisi fungsi yang memungkinkan menyesuaikan plot gambar dan menambahkan label ke plot. *Matplotlib* dapat digunakan untuk visualisasi data dalam 2D dan 3D (Khasanah, 2021).

c. *Library Pandas*

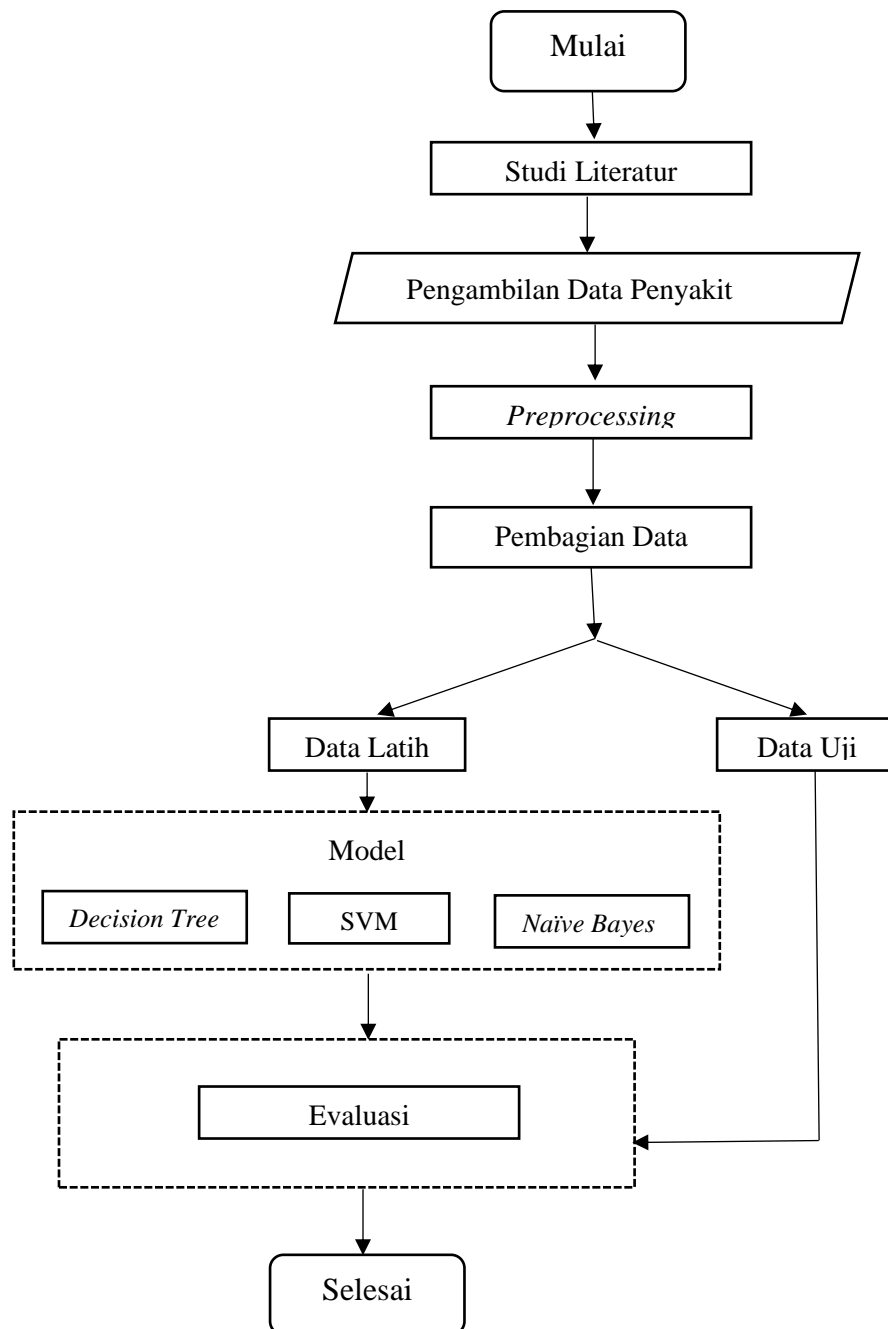
Pandas merupakan *library* dalam *python* yang *open source*. *Pandas* digunakan untuk memfasilitasi pemrosesan dan analisis data terstruktur. Membuat tabel, menyesuaikan ukuran data, memvalidasi data, dan lain lain. *Pandas* juga memproses data menggunakan teknik seperti penggabungan, pemisahan, pengelompokan, dan agregasi. Selain itu, *pandas* dapat membaca *file* dalam format seperti *.txt*, *.csv*, dan *.tsv* (Khasanah, 2021).

d. *Scikit-Learn*

Scikit-Learn adalah *library* dalam *python* untuk pembelajaran mesin. *Scikit-Learn* kompatibel dengan *library Numpy* dan *Scipy*. *Library* ini memudahkan untuk menganalisis data menggunakan algoritme pembelajaran mesin seperti algoritme regresi, *Naïve Bayes*, *Clustering*, *Decision Tree*, *parameter Turning*, *Data Processing Tool*, *Ekspor/Impor Model*, dan *Machine Learning Pipeline*. Algoritme lanjutan yaitu *Ensemble*, manipulasi fitur, deteksi *outlier*, serta pemilihan dan validasi model (Khasanah, 2021).

3.2.2.4. Alur Kerja Penelitian

Adapun alur kerja penelitian dapat dilihat pada Gambar 6 berikut:



Gambar 6. Alur kerja penelitian.

3.2.3. Studi Literatur

Studi literatur adalah menemukan sebuah masalah untuk diteliti, mencari informasi terkait dengan masalah yang akan diteliti, kemudian mengkaji sejumlah teori yang

mendasari penelitian ini, seperti jurnal dan buku yang relevan dengan masalah yang akan diteliti.

3.2.4. Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data sekunder dan diperoleh dari *Kaggle*, website penyedia data survei. Data yang digunakan sebanyak 400 data, terdiri dari 24 atribut. Dari total tersebut, 250 data mewakili pasien yang terkena penyakit ginjal kronis, sementara 150 data mewakili pasien yang tidak terkena penyakit ginjal kronis. *Dataset* diperoleh dari *Kaggle* yaitu *website* penyedia data pada laman:

<https://www.kaggle.com/code/niteshyadav3103/chronic-kidney-disease-prediction-98-accuracy>

3.2.5. Preprocessing

pada tahap *preprocessing* ini merupakan tahap untuk mengeliminasi atau menghilangkan beberapa masalah yang dapat mengganggu proses pengolahan data.

3.2.6. Pembagian Data

Pada tahap ini, data dipecah menjadi data latih dan data uji. Data latih adalah data yang digunakan untuk melatih algoritme untuk mengenali pola, dan data uji adalah data yang nantinya digunakan untuk menguji hasil dari pelatihan yang dilakukan oleh algoritme. Pembagian data ini menggunakan *k-fold cross validation* dengan menggunakan *5 fold*.

3.2.7. Pemodelan

Tahap selanjutnya adalah tahap klasifikasi, pada penelitian ini digunakan tiga metode yaitu metode *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes*. Pemilihan tiga metode ini berdasarkan saran dari penelitian sebelumnya untuk mencoba metode lain dalam menerapkan metode klasifikasi untuk mendiagnosis penyakit ginjal kronis.

3.2.8. Evaluasi

Pada tahap ini evaluasi dilakukan untuk mengukur kinerja metode klasifikasi. Pada penelitian ini menggunakan *Confusion matrix*.

3.2.9. Hasil

Hasil merupakan simpulan akhir dari penelitian yang sudah akurat karena telah dilakukan tahap evaluasi dimana tahap ini merupakan tahap pengesahan dan pengujian kebenaran dari algoritme yang menyatakan seseorang terkena penyakit ginjal kronis dan tidak terkena penyakit ginjal kronis.

V. SIMPULAN DAN SARAN

5.1. Simpulan

Penelitian analisis kinerja berbagai metode klasifikasi untuk diagnosis penyakit ginjal kronis, menggunakan metode *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* dapat ditarik kesimpulan sebagai berikut:

1. Algoritme *Decision Tree*, *Support Vector Machine*, dan *Naïve Bayes* berhasil diimplementasikan untuk memprediksi penyakit ginjal kronis. Data yang digunakan merupakan data penyakit ginjal kronis yang diperoleh dari Kaggle, dengan jumlah data 400 yang terdiri dari 24 atribut. Proposisi pembagian data 80:20, dengan 80 data *training* dan 20 data *testing*.
2. *Performance* metode *Decision Tree* diperoleh *precision* 0.97, *recall* 0.97, *F1-score* 0.97, dan akurasi 0.97. Metode *Support Vector Machine* diperoleh *precision* 0.98, *recall* 0.98, *F1-score* 0.98, dan akurasi 0.98. Metode *Naïve Bayes* diperoleh *precision* 0.96, *recall* 0.96, *F1-score* 0.96, dan akurasi 0.96. Berdasarkan hasil tersebut, dapat disimpulkan bahwa ketiga metode klasifikasi yang digunakan dapat memberikan prediksi yang cukup baik untuk diagnosis penyakit ginjal kronik. Metode *support vector machine* mencapai hasil terbaik dengan mencapai akurasi sempurna 0.98. *Decision Tree* dan *Naive Bayes* juga bekerja dengan baik. Hal ini menunjukkan bahwa metode *support vector machine* dapat digunakan sebagai alat yang efektif untuk mendiagnosis penyakit ginjal kronis.

5.2. Saran

Saran yang diberikan dalam penelitian ini sebagai berikut:

1. Dalam penelitian selanjutnya, disarankan untuk menggunakan metode klasifikasi yang lebih baru seperti *deep learning* atau *ensemble learning*.
2. Penelitian selanjutnya dapat menggunakan *dataset* yang lebih besar dan mencakup variasi atribut yang lebih luas. Hal ini akan membantu dalam menganalisis kinerja metode klasifikasi secara lebih baik.
3. Disarankan untuk mengimplementasikan hasil penelitian dalam bentuk sistem web atau aplikasi. Sistem ini akan menjadi alat bantu bagi dokter atau tenaga medis dalam mendiagnosa penyakit ginjal kronis dengan lebih efektif dan efisien.

DAFTAR PUSTAKA

- Abdulghani, T., & Gozali, M. M. H. (2020). Sistem Konsultasi dan Bimbingan Online Berbasis Web Menggunakan Webrtc (Studi Kasus : Fakultas Teknik Universitas Suryakencana). *Media Jurnal Informatika*, 11(2), 42. <https://doi.org/10.35194/mji.v11i2.1037>
- Ahmed, F. Y. H., Ali, Y. H., & Shamsuddin, S. M. (2018). Using K-fold cross validation proposed models for SpikeProp learning enhancements. *International Journal of Engineering and Technology(UAE)*, 7(4), 145–151. <https://doi.org/10.14419/ijet.v7i4.11.20790>
- Amalia, H. (2018). *Perbandingan Metode Data Mining SVM dan NN untuk Klasifikasi Penyakit Ginjal Kronis*. 14(1), 1–6.
- Anggreany, M. S. (2022). *Confusion Matrix*. Socs.Binus.Ac.Id. <https://socs.binus.ac.id/2020/11/01/confusion-matrix/>
- Ariani, A., & Samsuryadi. (2019). Klasifikasi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor. *Prosiding Annual Research Seminar 2019*, 5(1), 148–151.
- Aulia. (2017a). *Ginjal Kronis*. P2ptm.Kemkes.Go.Id. <https://p2ptm.kemkes.go.id/kegiatan-p2ptm/subdit-penyakit-jantung-dan-pembuluh-darah/ginjal-kronis>
- Aulia. (2017b). *Rutin Aktivitas Fisik, Keluarga Terhindar PTM*. P2ptm.Kemkes.Go.Id. <https://p2ptm.kemkes.go.id/kegiatan-p2ptm/subdit-penyakit-jantung-dan-pembuluh-darah/rutin-aktivitas-fisik-keluarga-terhindar-ptm>
- Azmi, Z., & Dahria, M. (2013). Decision Tree Berbasis Algoritma Untuk. In *Jurnal SAINTIKOM* (Vol. 12, pp. 157–164).
- Bahri, S., & Lubis, A. (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. *Jurnal Sintaksis*, 2(1), 63–70.
- Chen, X., & Liu, W. (2022). The Value of Python Programming in General Education and Comprehensive Quality Improvement of Medical Students Based on a Retrospective Cohort Study. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/4043992>
- Desiani, A. (2022). Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest

- Neighbor Pada Klasifikasi Penyakit Hati. *Simkom*, 7(2), 104–110. <https://doi.org/10.51717/simkom.v7i2.96>
- dr. Meva Nareza. (2021). *Gagal Ginjal Kronis*. Alodokter.Com. <https://www.alodokter.com/gagal-ginjal-kronis>
- Eugenia, C. (2021). *Seputar Fungsi Microsoft Excel yang Perlu Kalian Ketahui*. Dqlab.Id. <https://www.dqlab.id/seputar-fungsi-microsoft-excel-yang-perlu-kalian-ketahui>
- Faddillah, A. N., Wijaya, J., & Hidayat, R. (2019). Penerapan Algoritma Naive Bayes Untuk Diagnosa Penyakit Gagal Ginjal Kronis. *Jurnal Information ...*, 18(2), 102–106. <https://doi.org/10.36054/jict-ikmi.v18i2.69>
- Gokiladevi, M., Santhoshkumar, S., & Varadarajan, V. (2022). Machine Learning Algorithm Selection for Chronic Kidney Disease Diagnosis and Classification. *Malaysian Journal of Computer Science*, 2022(Special Issue 1), 102–115. <https://doi.org/10.22452/mjcs.sp2022no1.8>
- Hermawati, P. A. (2013). *Data Mining*. Andi.
- Khasanah, L. U. (2021). *Kenali 4 Library Python yang Populer untuk Data Science*. Dqlab.Id. <https://www.dqlab.id/kenali-4-library-python-yang-populer-untuk-data-science>
- Krishna, V. S., Meghana, R. P., Reddy, N. V. S., & Rao, B. A. (2022). *Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers*. <https://doi.org/10.1088/1742-6596/2161/1/012015>
- Kumar, S., Mohammad, K., Uddin, M., Hasan, H., Rahman, M., Howlader, A., & Uddin, K. M. A. (2022). Intelligent Systems with Applications Chi 2 -MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease. *Intelligent Systems with Applications*, 16(October), 200144. <https://doi.org/10.1016/j.iswa.2022.200144>
- Kurnianto, E. A., Cholissodin, I., & Santoso, E. (2018). Klasifikasi Penderita Penyakit Ginjal Kronis Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(12), 6597–6602.
- Kusuma, H., Suhartini, Ropyanto, C. B., Hastuti, Y. D., Hidayati, W., Sujianto, U., Widyaningsih, S., Lazuardi, N., Yuwono, I. H., Husain, F., Z.N, E. G., Selvia, A., & Benita, M. Y. (2019). *Mengenal Penyakit Ginjal Kronis dan Perawatannya* (H. Kusuma (ed.)). Fakultas Kedokteran Universitas Diponegoro.
- Medina, M. I. (2022). *Penting untuk Data Scientist, Ketahui Apa Saja Fungsi dan Fitur Jupyter*. Glints.Com. <https://glints.com/id/lowongan/jupyter-adalah/#.Y53otHZBzIU>

- Muslim, M. A., Prasetyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). *data mining algoritma c4.5.pdf* (E. Listiana & N. Cahyani (eds.)).
- Prasetyo, E., Rahajoe, R. A. D., & Arizal, A. (2013). Perbandingan K-Support Vector Nearest Neighbour Terhadap Decision Tree dan Naive Bayes. *Seminar Nasional Teknik Informatika*, 1–6.
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *J-Ptiik*, 2(2), 802–810. <http://j-ptiik.ub.ac.id>
- Santosa, B. (2007). *DATA MINING: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu.
- Saringat, Z., Mustapha, A., Saedudin, R. D. R., & Samsudin, N. A. (2019). Comparative analysis of classification algorithms for chronic kidney disease diagnosis. *Bulletin of Electrical Engineering and Informatics*, 8(4), 1496–1501. <https://doi.org/10.11591/eei.v8i4.1621>
- Senan, E. M., Al-Adhaileh, M. H., Alsaade, F. W., Aldhyani, T. H. H., Alqarni, A. A., Alsharif, N., Uddin, M. I., Alahmadi, A. H., Jadhav, M. E., & Alzahrani, M. Y. (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/1004767>
- Susilowati, E., Sabariah, M. K., & Gozali, A. A. (2015). Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter. *E-Proceeding of Engineering*, 2(1), 1478–1484.
- Tekale, S., Shingavi, P., Wandhekar, S., & Chatorikar, A. (2022). *Prediction of Chronic Kidney Disease Using Machine Learning Algorithm. October 2018*. <https://doi.org/10.17148/IJARCCCE.2018.71021>
- Trivusi. (2022). *Penjelasan Lengkap Algoritma Support Vector Machine (SVM)*. Trivusi.Web.Id. <https://www.trivusi.web.id/2022/04/algoritma-svm.html>
- Wibawa, A. P., Akbar, M. G. A., Fathony, P. M., & Dwiyanto, F. A. (2018). metode metode klasifikasi.pdf. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 5.
- Zeynu, S. (2018). *Prediction of Chronic Kidney Disease Using Data Mining Feature Selection and Ensemble Method*. 15, 168–176.