

BAB II

LANDASAN TEORI

2.1 PENYAKIT GAGAL JANTUNG

Gagal jantung merupakan keadaan dimana jantung tidak lagi mampu memompa darah dalam jumlah yang memadai ke jaringan untuk memenuhi kebutuhan metabolisme tubuh atau kemampuan tersebut hanya dapat terjadi dengan tekanan pengisian jantung yang tinggi bahkan dapat terjadi pula keduanya. Terjadinya gagal jantung diawali dengan adanya kerusakan pada jantung atau *myocardium*, diikuti penurunan curah jantung. Bila kebutuhan metabolisme tidak terpenuhi, maka jantung akan memberikan respon mekanisme kompensasi untuk mempertahankan fungsi jantung agar tetap dapat memompa darah. Terapi bagi penderita gagal jantung berupa terapi *non-farmakologis* yang artinya pada penderita gagal jantung merupakan terapi perawatan mandiri yang dilakukan oleh pasien, dan terapi *farmakologis* adalah terapi untuk mengatasi gejala, memperlambat pemburukan kondisi jantung dan mengatasi terjadinya akut akibat respon kompensasi jantung. Ada beberapa macam obat-obatan yang dikonsumsi oleh penderita penyakit gagal jantung yaitu: *diuretik, antagonis aldosterone, angiotensin converting enzyme inhibitor, angiotensin receptor blocker, beta blocker, glikosida jantung, vasodilator, agonis beta, bipyridine, dan natriuretic peptide* [8].

Gagal jantung merupakan masalah kesehatan yang progresif dengan angka kematian dan penyakit dalam suatu populasi yang tinggi di negara maju maupun

negara berkembang termasuk Indonesia. Di Indonesia, usia pasien gagal jantung relatif lebih muda dibanding Eropa dan Amerika disertai dengan tampilan klinis yang lebih berat. Prevalensi dari gagal jantung sendiri semakin meningkat karena pasien yang mengalami kerusakan jantung yang bersifat akut dapat berlanjut menjadi gagal jantung kronik. *World Health Organization* (WHO) menggambarkan bahwa meningkatnya jumlah penyakit gagal jantung di dunia, termasuk Asia diakibatkan oleh meningkatnya angka perokok, tingkat obesitas, *dyslipidemia*, dan diabetes. Angka kejadian gagal jantung meningkat juga seiring dengan bertambahnya usia [9].

2.2 DATA MINING

Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengembalian informasi dari *database* yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan *database*. Tujuan dari data mining adalah untuk mengekstrak informasi dengan metode cerdas dari kumpulan data kemudian mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut [10].

Ada beberapa jenis dan tahapan dalam data mining yang dijelaskan sebagai berikut:

2.2.1 Teknik Data Mining

Sebagaimana yang dijelaskan pada [11] ada beberapa macam teknik data mining yang dapat dibagi menjadi beberapa kelompok berdasarkan tugas masing-masing yaitu:

1. *Description* (Deskripsi)

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. *Estimation* (Estimasi)

Hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

3. *Prediction* (Prediksi)

Hampir sama dengan klasifikasi dan estimasi, dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi dalam bisnis dan penelitian:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi tingkat pengangguran lima tahun akan datang.

4. *Classification* (Klasifikasi)

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- b. Mendiagnosis penyakit seorang pasien untuk mendapatkan informasi termasuk penyakit apa yang diderita oleh pasien.

5. *Clustering* (Pengkusteran)

Pengkusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak samaan dengan *record-record* dalam kluster lain. Contoh pengkusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Melakukan pengkusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. *Association* (Asosiasi)

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons positif terhadap penawaran upgrade layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

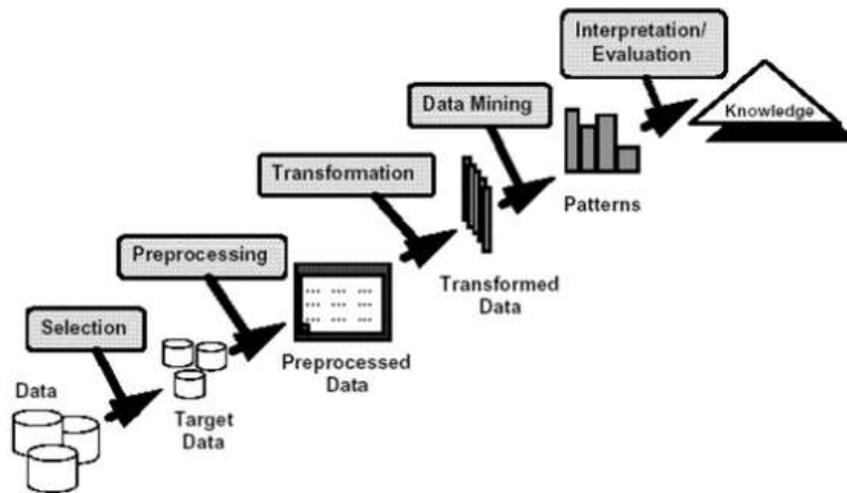
2.2.2 Proses Tahapan Data Mining

Data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data [12].

Serangkaian proses tahapan data mining tersebut adalah sebagai berikut:

1. Pembersih data (untuk membuang data yang tidak konsisten dan noise).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining).
4. Aplikasi teknik data mining, proses ekstraksi pola dari data yang ada.
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik visualisasi).

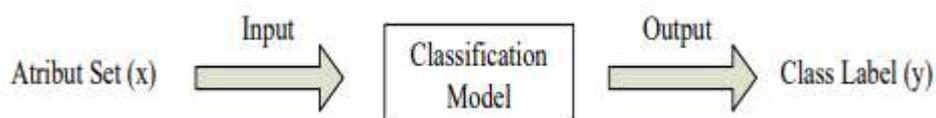
Berikut gambar dari proses tahapan data mining:



Gambar 2. 1 Proses Tahapan Data Mining (KDD) [12]

2.3 KLASIFIKASI

Klasifikasi merupakan teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Klasifikasi merupakan suatu teknik menemukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan [13].



Gambar 2. 2 Blok Diagram Model Klasifikasi [13]

Berikut penjelasan beberapa algoritma klasifikasi yang sering digunakan antara lain:

1. *Support Vector Machine (SVM)*

Support Vector Machine merupakan sistem pembelajaran yang menggunakan ruang hipotesis yang berupa fungsi-fungsi linear di dalam sebuah fitur yang memiliki dimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran berdasarkan teori Optimasi [14]. Pada penelitian yang dilakukan oleh Ichwan, Irma dan Zeni [15] menyatakan bahwa kelemahan metode SVM berdasarkan penelitian mereka yaitu sulit mengaplikasikan untuk himpunan dengan jumlah dimensi yang sangat besar.

2. *K-Nearest Neighbor (KNN)*

K-Nearest Neighbor adalah algoritma *supervised learning* dimana hasil dari *instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat, Algoritma ini menggunakan *Neighborhood Class cation* sebagai nilai prediksi dari nilai *instance* yang baru [16]. Sama halnya dengan penelitian yang dilakukan oleh [15] menyatakan bahwa kelemahan metode KNN adalah sulit mengklasifikasikan pola-pola yang tersebar acak.

3. *Neural Network (NN)*

Metode *Neural Network* memiliki kinerja yang baik dalam mengatasi masalah data *nonlinear*, namun *Neural Network* memiliki keterbatasan dalam mengatasi data *noise* yang tinggi. Untuk menyelesaikan masalah tersebut diperlukan metode *bagging* untuk mengurangi data *noise* pada

metode *Neural Network* [17]. Pada penelitian yang dilakukan oleh Aji Prasetya dkk [18] menyatakan bahwa dalam model ini menggunakan analisis yang sangat kompleks sehingga sangat susah untuk diinterpretasikan.

2.4 ALGORITMA C4.5

Algoritma *C4.5* merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki keunggulan utama dibanding algoritma yang lainnya. Kelebihan dari algoritma *C4.5* dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani dua atribut yang bertipe diskrit dan numerik. Dalam membangun pohon, algoritma *C4.5* membaca semua sampel data pelatihan dari memori dan memuatnya ke dalam memori secara keseluruhan pada waktu yang bersamaan [19]. Secara umum algoritma *C4.5* untuk membangun pohon keputusan sebagai berikut:

1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi atribut terpilih dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua atribut terpilih pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 2.1 berikut:

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots (2.1)$$

Keterangan:

S: Himpunan kasus

A: Atribut

N: Jumlah partisi atribut A

| Si |: Jumlah kasus pada partisi ke-i

| S |: Jumlah kasus dalam S

Dari keterangan di atas akan diperoleh nilai *gain* dan atribut yang paling tertinggi.

Gain adalah salah satu atribut *selection measure* digunakan untuk memilih test atribut tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai test atribut dari suatu node, untuk itu penghitungan nilai entropi dapat dilihat pada persamaan 2.2 berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots (2.2)$$

Keterangan:

S: Himpunan kasus

A: Atribut

N: Jumlah partisi S

Pi: Proporsi dari Si terhadap S

Kemudian lakukan pengulangan dalam perhitungan *gain* dan *entropy* sampai semua *record* ter partisi. Proses partisi pohon keputusan akan berhenti saat:

- 1) Semua *record* dalam simpul N mendapat kelas yang sama.
- 2) Tidak ada atribut di dalam *record* yang di partisi lagi dan tidak ada yang kosong.

2.5 ALGORITMA NAÏVE BAYES

Naïve Bayes adalah suatu metode yang digunakan untuk dapat memperkirakan atau memprediksi suatu *class* dari suatu objek yang kelas nya tidak diketahui dari masing-masing kelompok atribut yang ada, dan menentukan *class* mana yang paling optimal berdasarkan pengaruh yang didapat dari hasil pengamatan. Klasifikasi *Naïve Bayes* adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [20]. *Naïve Bayes* ditemukan oleh seseorang ilmuwan dari Inggris yaitu Thomas Bayes. Algoritma ini dapat menghasilkan prediksi peluang yang akan datang berdasarkan suatu pengalaman sebelumnya, teorema Bayes memiliki bentuk umum dapat dilihat pada persamaan 2.3 berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots(2.3)$$

Keterangan:

X = Kelas data yang belum diketahui.

H = Hipotesa data X adalah kelas spesifik.

P (X|H) = Probabilitas hipotesa H berdasarkan keadaan X (posteriori probability).

P (H) = Probabilitas hipotesa H (posteriori probability).

P (X|H) = Probabilitas X berdasarkan keadaan tersebut.

P (X) = Probabilitas dari X.

Adapun alur dari metode Naïve Bayes adalah sebagai berikut:

1. Baca data training.
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:

- a. Cari nilai mean dan standar deviasi dari masing – masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel mean, standar deviasi dan probabilitas.

2.6 DATA PREPARATION

Data preparation adalah proses menyiapkan data mentah sehingga layak untuk diproses dan dianalisis lebih lanjut. Semua algoritma *machine learning* dapat menemukan pola dalam data, namun apabila algoritma tersebut diberikan data yang “kacau” maka model yang dihasilkan akan memberikan luaran yang kacau juga. Jadi syarat utamanya adalah data yang bersih, bukan tentang algoritma yang dipakai. Algoritma yang sederhana tetapi dengan data yang bersih dapat mengalahkan algoritma yang rumit dengan data yang kacau. Secara umum proses persiapan data terbagi atas tiga kegiatan utama yaitu pembersihan data (*data cleaning*), rekayasa fitur (*feature engineering*) dan pembagian data (*data Split*) [21].

a. Pembersih Data

Pembersihan data atau *Data Cleaning* adalah proses mendeteksi dan memperbaiki (atau menghapus) data yang rusak atau tidak akurat dari dataset, tabel, atau basis data untuk meningkatkan kualitas data. Pada tahap ini juga dilakukan identifikasi terhadap data yang tidak lengkap, salah, tidak akurat atau tidak relevan dari dataset.

b. Rekayasa Fitur

Rekayasa fitur atau *Feature Engineering* adalah proses perubahan dataset dimana dilakukan proses ekstraksi sebuah kolom menjadi kolom-kolom lainnya atau proses mereduksi kolom.

c. Pembagian Data

Setelah mendapatkan data yang bersih, selanjutnya adalah melakukan proses *data splitting* (pemecahan data) dimana kita membagi data tersebut menjadi beberapa bagian diantaranya data latih (*training data*), data uji (*testing data*), dan data validasi (*validation data*). Metode ini dikenal dengan nama *holdout* dimana kita membagi dataset menjadi 2 bagian yaitu *training set* dan *holdout set*. *Training set* berisi data latih yang digunakan untuk membangun model dan biasanya ukurannya lebih besar dibandingkan dengan *holdout set*. Sedangkan *holdout set* dapat terdiri atas data uji (*testing data*), dan data validasi (*validation data*) atau hanya data uji saja.

2.7 EDA (Exploratory Data Analysis)

Exploratory Data Analysis adalah suatu proses uji investigasi awal yang bertujuan untuk mengidentifikasi pola, menemukan anomali, menguji hipotesis dan memeriksa asumsi. Dengan melakukan EDA, pengguna akan sangat terbantu dalam mendeteksi kesalahan dari awal, dapat mengidentifikasi *outlier*, mengetahui hubungan antar data serta dapat menggali faktor-faktor penting dari data. Proses EDA ini sangat bermanfaat dalam proses analisis statistik. Pada proses pengolahan data, dalam melakukan *exploratory data analysis* dapat menggunakan beberapa

teknik diantaranya Statistik deskriptif, *Univariate analysis*, dan *Multivariate analysis* [22].

- a. Statistik deskriptif, yaitu mendeskripsikan data sehingga menghasilkan informasi secara umum tanpa bertujuan untuk menarik kesimpulan. Statistik deskriptif dapat menampilkan beberapa informasi penting seperti nilai rata-rata, median, modus, standar deviasi, variasi dan kecekungan. Statistik deskriptif ini dapat ditampilkan dalam berbagai bentuk seperti tabel, diagram, grafik, dll.
- b. *Univariate analysis*, yaitu menganalisis kolom secara terpisah dan melihat distribusi datanya. *Univariate analysis* secara umum terbagi dua, yaitu numerik dan kategori. Analisis ini digunakan juga dengan tujuan untuk mengambil kesimpulan dengan menggunakan beberapa analisis inferensial yang mungkin digunakan.
- c. *Multivariate analysis*, yaitu menggabungkan beberapa kolom dengan tujuan menemukan hubungan antara satu kolom dengan kolom lainnya. *Multivariate analysis* ini melibatkan variabel dalam jumlah lebih atau sama dengan tiga variabel.

2.8 EVALUASI MODEL

Model evaluasi merupakan suatu model dari *machine learning* dapat diukur atau dievaluasi menggunakan satu atau lebih metode. Terdapat beragam pendekatan dalam mengevaluasi model antara lain pengukuran akurasi, *error* estimasi, komparasi data, tingkat kepercayaan akurasi, dan sebagainya. Tentunya tidak

semua pendekatan tersebut digunakan untuk menguji kinerja suatu model, namun tergantung dari tujuan Pemodelan nya [23]. Ada beberapa metrik pada model evaluasi ini diantaranya *confusion matrix*, *accuracy*, *precision*, *recall* dan *F1-score*.

a. *Confusion matrix*

Confusion matrix adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi pada sekumpulan data uji yang nilai sebenarnya diketahui. Semua ukuran kecuali AUC dapat dihitung dengan menggunakan paling kiri empat parameter yaitu *True Positives* (TP), *True Negatives* (TN), *False Positive* (FP), dan *False Negatif* (FN).

b. *Accuracy*

Akurasi adalah ukuran kinerja yang paling intuitif dan itu hanyalah rasio pengamatan yang diperiksa dengan benar terhadap total pengamatan. Orang mungkin berpikir bahwa, jika kita memiliki akurasi tinggi maka model kita adalah yang terbaik.

c. *Presisi*

Presisi adalah rasio pengamatan positif yang diprediksi dengan benar dengan total pengamatan positif yang diprediksi.

d. *Recall*

Recall adalah rasio pengamatan positif yang diprediksi dengan benar dengan semua pengamatan di kelas yang sebenarnya.

e. *F1-score*

Skor F1 adalah rata-rata tertimbang dari *Precision* dan *Recall*. Oleh karena itu, skor ini memperhitungkan positif palsu dan negatif palsu. Secara intuitif

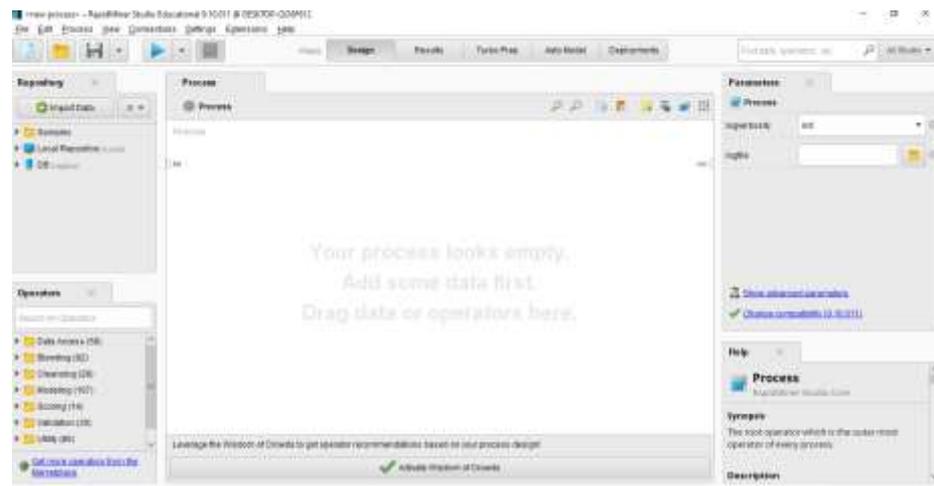
tidak semudah memahami akurasi, tetapi F1 biasanya lebih berguna daripada akurasi, terutama jika Anda memiliki distribusi kelas yang tidak merata.

2.9 ALAT BANTU PENGUJIAN

Ada banyak *tools* yang tersedia untuk data mining. Tujuan utama Data Mining *tools* adalah untuk menemukan data, mengekstrak data, menyaring data, dan mendistribusikan informasi. Pada penelitian ini penulis akan menggunakan *tools Rapid Miner* dan ada beberapa penjelasan data mining *tools* antara lain:

1. Rapid Miner

Penulis akan menggunakan *tools Rapid Miner* karena semua fasilitas sudah disediakan. Dengan menggunakan aplikasi ini tidak dibutuhkan kemampuan *coding* khusus. *Rapid Miner* dikhususkan untuk pengguna data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti *Model Bayes in, Modelling, Tree Induction, Neural Network* dan lain-lain. Banyak juga metode yang disediakan *Rapid Miner* seperti *Klasifikasi, Clustering* dan lainnya. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java di bawah lisensi *GNU Public License*. *Rapid Miner* dapat dijalankan pada sistem operasi apapun [24].



Gambar 2. 3 Tampilan Awal Rapid Miner

2. Weka

Weka merupakan singkatan dari “*Waikato Environment for Knowledge Analysis*” adalah perangkat lunak pembelajaran mesin populer yang ditulis dalam bahasa pemrograman java. Weka mencakup kumpulan algoritma dan visualisasi untuk analisis data [25].

3. Orange

Orange merupakan sebuah *tools open source* untuk pengolahan data mining. Orange memiliki widget yang berfungsi sebagai unit komputasi untuk membaca, memproses, melakukan visualisasi, melakukan analisis, mengeksplorasi data, dan lain-lain [26].

4. SPSS

SPSS merupakan salah satu sekian banyak *software* statistika yang telah dikenal luas di kalangan penggunaannya. Disamping masih banyak lagi software statistika lainnya seperti Minitab, Microstate dan masih banyak

lagi. SPSS sebagai sebuah *tools* mempunyai banyak kelebihan, terutama untuk aplikasi di bidang ilmu sosial [27].

2.10 PENELITIAN SEJENIS

Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik data mining untuk menggali berbagai informasi dari sebuah database. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian ini serta mengenai algoritma yang akan digunakan penulis dalam penelitian ini dapat dilihat pada tabel 2.1 berikut:

Tabel 2. 1 Penelitian Sejenis

No	Penulis dan Tahun	Judul	Masalah	Metode	Hasil
1	Abdul Rohman dan M. Rochcham, 2018.	Model Algoritma <i>C4.5</i> untuk Prediksi Penyakit Jantung	Memprediksi penyakit jantung sehingga terbentuk model nya, dengan mengoptimalkan atribut-atribut yang berasal dari dataset yang terpercaya untuk memprediksi penyakit jantung.	Metode yang digunakan adalah algoritma <i>C4.5</i>	Pengujian dengan menggunakan <i>C4.5</i> didapat nilai accuracy adalah 86,59 % dengan nilai AUC adalah 0.957, dan masuk kategori kelompok klasifikasi yang sangat baik, karena nilai AUC antara 0.90 sampai 1.00

2	Agil Langga dan Agung Nilogiri, 2019.	Perbandingan Algoritma <i>Naïve Bayes</i> dengan Algoritma <i>K-Nearest Neighbor</i> untuk prediksi penyakit jantung	Membandingkan perhitungan hasil prediksi dan menampilkan hasil akurasi, presisi, dan recall dari pengujian dataset penyakit jantung	Menggunakan metode <i>Naïve Bayes</i> dan <i>K-Nearest Neighbor</i> .	Algoritma <i>Naive Bayes</i> menghasilkan performa tertinggi pada <i>10-fold Cross Validation</i> , folder data pengujian ke-4 dengan akurasi yaitu 90.00%, dan nilai presisi sebesar 86,67%. Algoritma <i>K-Nearest Neighbor</i> juga menghasilkan akurasi tertinggi dengan jumlah keterangan sebesar 7 pada pengujian <i>10-fold Cross Validation</i> dengan akurasi tertinggi pada folder data pengujian ke-4 yaitu sebesar 80.00%, dan presisi sebesar 90.00%. Kesimpulan bahwa algoritma <i>Naive Bayes</i> lebih akurat dan lebih baik
---	---------------------------------------	--	---	---	---

					dalam klasifikasi penyakit jantung dibandingkan algoritma <i>K-Nearest Neighbor</i> .
3	Edric dan Saut Parsaoran Tamba, 2022.	Prediksi Penyakit Gagal Jantung dengan menggunakan <i>Random Forest</i>	Untuk memprediksi probabilitas penyakit gagal jantung. Sistem klasifikasi dengan <i>Random Forest</i> yang dibuat dapat diimplementasikan ke dalam rangkaian alat untuk menciptakan alat deteksi aritmia EKG otomatis portable.	Algoritma <i>Random Forest</i>	Algoritma <i>Random Forest</i> yang digunakan sukses dengan meraih tingkat akurasi sebesar 82,6087% yang kemudian di Optimasi dengan teknik <i>K - Fold</i> Dan <i>Grid Search CV</i> menjadi 85,058%.

4	Pandito Dewa dan Dian Palupi, 2019.	Prediksi Penyakit Jantung dengan Algoritma Klasifikasi	Melakukan sebuah komparasi terhadap algoritma-algoritma yang memiliki tingkat akurasi tertinggi tersebut dengan menggunakan objek <i>stat log heart disease</i> dataset. Sehingga berbagai perbandingan algoritma tersebut akan menghasilkan algoritma dengan akurasi paling baik.	Algoritma Klasifikasi menggunakan metode <i>Naïve Bayes, Support Vector Machine, C4.5, Logistic Regression</i> dan <i>Back Propagation</i> .	Berdasarkan <i>cross validation</i> dengan masing-masing algoritma yang ditetapkan, sehingga menghasilkan akurasi tertinggi didapat oleh algoritma <i>naïve Bayes</i> dengan akurasi 84.07%. Kemudian untuk presisi tertinggi didapat oleh algoritma <i>naïve Bayes</i> dengan presisi 86.16%. Selanjutnya untuk <i>recall</i> tertinggi didapat oleh algoritma <i>support vector machine</i> dengan <i>recall</i> 94.67%.
---	-------------------------------------	--	--	--	--

5	Socayo Adi dan Atik Wintarti, 2022.	Kompresi Metode <i>Support Vector Machine</i> (SVM), <i>K-Nearest Neighbors</i> (KNN), dan <i>Random Forest</i> (RF) untuk prediksi penyakit gagal jantung.	Membandingkan prediksi akurasi dari tiga metode yang dipakai oleh peneliti.	Metode <i>Support Vector Machine</i> (SVM), <i>K-Nearest Neighbors</i> (KNN), dan <i>Random Forest</i> (RF)	Akurasi paling tinggi yang dapat dihasilkan adalah menggunakan metode SVM dan RF dimana menghasilkan akurasi yang bernilai sama, yaitu 97% dengan rincian metode SVM menggunakan parameter $C = 1$, $\gamma = 0.01$, kernel = linear dan dalam total waktu running program selama 2.82 detik sedangkan metode RF menggunakan parameter estimators = 30, random state = 0, dan dalam total waktu running program selama 7.29 detik. Lalu untuk metode KNN menghasilkan akurasi yang bernilai
---	-------------------------------------	---	---	---	---

					93%, dengan menggunakan parameter neighbors = 20 dan dalam total waktu running program selama 0.60 detik.
6	Jefri Junifer, Hery Tanjaya dan Kenichi, 2021.	Mendeteksi penyakit jantung menggunakan <i>Machine Learning</i> dengan Algoritma <i>Logistic regression</i>	Mengimplementasikan penggunaan algoritma <i>logistic regresi</i> , dimana algoritma tersebut memakai fungsi logistik untuk menghasilkan binary atau nol dan satu sebagai penentuan klasifikasi.	Algoritma <i>Logistic regression</i>	Pada data <i>training</i> , metode logistik regresi mempunyai nilai <i>sensitivity</i> yang paling tinggi yaitu 88.54% dibanding metode lainnya. Pada data <i>testing</i> , metode logistik regresi mempunyai nilai kekhususan yang paling tinggi yaitu 87.50% dibanding metode lainnya

7	Wahyu Nugraha, 2021.	Prediksi penyakit jantung <i>Cardiovascular</i> menggunakan model Algoritma Klasifikasi.	Agar mengetahui hasil perbandingan prediksi Penyakit jantung menggunakan beberapa algoritma klasifikasi pada <i>machine learning</i> .	Menggunakan metode <i>Random Forest, Support Vector Machines, Gradient Boosting Machines, XGBOOST</i> , dan <i>Light GBM</i> .	Model klasifikasi menggunakan <i>XGBOOST</i> rata-rata memperoleh nilai tertinggi baik menggunakan pengukuran <i>accuracy, F1 score</i> maupun AUC. Model klasifikasi SVM memperoleh nilai rata-rata hampir mirip dengan <i>model Gradient Boosting</i> . Sedangkan untuk model pengukuran dengan nilai terendah diperoleh dengan model klasifikasi <i>LightGBM</i> .
8	Derisma, 2020.	Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung	Untuk mengembangkan sistem prediksi untuk menganalisis dan memprediksi	Algoritma <i>Naive Bayes, Random Forest</i> , dan <i>Neural Network</i>	Hasil perbandingan menunjukkan bahwa dalam penggunaan algoritma klasifikasi data mining yang digunakan yaitu Algoritma <i>Naive</i>

		dengan Teknik Data Mining.	kemungkinan penyakit jantung.		<i>Bayes, Random Forest, Neural Network</i> dapat kita lihat bahwa algoritma <i>Naive Bayes</i> adalah algoritma yang tepat dan akurat digunakan untuk dapat melakukan prediksi penderita penyakit jantung dengan <i>persentase</i> sebesar 83 %.
9	Quswatun Hasanah, Hardian Oktavianto, dan Yeni Dwi Rahayu, 2022	Analisis Algoritma <i>Gaussian Naive Bayes</i> Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung	Bagaimana menentukan kriteria hasil klasifikasi terhadap output data asli (nilai aktual) menggunakan confusion matrix dengan algoritma <i>Gaussian Naive Bayes</i>	Algoritma <i>Gaussian Naive Bayes</i>	Hasil penerapan algoritma <i>Gaussian Naive Bayes</i> dan dilakukan pengujian menggunakan <i>K-fold</i> , nilai akurasi tertinggi ada pada 10 -fold pada tahap 9 dengan skenario 4 sebesar 69%, presisi tertinggi ada pada 4 -fold pada tahap 4 skenario 2 sebesar 65,73% serta nilai recall tertinggi ada pada 10

					- fold pada tahap ke 4 skenario 4 yaitu 95,91%.
10	Baiq Andriska dan Intan Komala Dewi, 2021.	Komparasi Metode Klasifikasi Data Mining <i>Decision Tree</i> dan <i>Naïve Bayes</i> Untuk Prediksi Penyakit Diabetes	Diperlukan metode terbaik dalam melakukan klasifikasi penyakit diabetes untuk mengetahui secara dini seseorang mengalami diabetes.	Algoritma <i>Decision Tree</i> dan <i>Naïve Bayes</i>	Algoritma klasifikasi <i>decision tree</i> lebih baik dalam prediksi penyakit diabetes dengan nilai akurasi 95,58% dan nilai AUC 0,981 lebih tinggi dibandingkan <i>naïve bayes</i> dengan akurasi 87,69% dan nilai AUC 0,947

Dari beberapa penelitian sejenis pada tabel 2.1 diatas dapat dilihat bahwa algoritma *C4.5* dan *Naïve Bayes* merupakan salah satu metode data klasifikasi yang dapat menghasilkan akurasi tertinggi, namun kedua metode ini masih sama-sama memiliki kelebihan dan kekurangan masing-masing.

Berdasarkan kesimpulan penelitian sejenis, maka yang membedakan dengan penelitian ini adalah penulis akan memprediksi penyakit gagal jantung menggunakan dua metode yaitu metode *C4.5* dan *Naïve Bayes* serta mengukur tingkat akurasi nya. Data yang didapat dari situs web yaitu Kaggle.com yang di ambil pada laman <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>. Mempunyai 918 data dan mempunyai 12 atribut yang siap digunakan.

Atribut yang digunakan yaitu terdiri dari:

1. *Age* : Usia pasien [tahun]
2. *Gender* : Jenis kelamin pasien [M: Pria, F: Wanita]
3. *Chest pain type* : Tipe nyeri dada [TA: Angina Khas, ATA: Angina Atypical, NAP: Nyeri Non-Angina, ASY: Tanpa Gejala]
4. *Resting BP* : Tekanan darah maksimum[mm Hg]
5. *Cholesterol* : Kolesterol [mm/dl]
6. *Fasting BS* : Gula darah [1: jika Puasa BS > 120 mg/dl, 0: sebaliknya]
7. *Resting ECG* : Hasil elektrokardiogram [Normal: Normal, ST: memiliki kelainan gelombang ST-T (inversi gelombang T atau depresi ST > 0,05 mV), LVH: menunjukkan kemungkinan atau pasti hypertrophy ventrikel kiri menurut kriteria Estes]
8. *Max HR* : Detak jantung maksimum [Nilai numerik antara 60 - 202]

9. *Exercise Angina* : Nyeri Dada Akibat Olahraga [Y: Ya, N: Tidak]
10. *Old peak* : Gangguan Kompleks [Nilai numerik diukur dalam depresi]
11. *ST Slope* : Kemiringan Detak Jantung [Up: up sloping, Flat: flat,
Down: down sloping]
12. *Heart Disease* : Kelas keluaran [1: penyakit jantung, 0: Normal]