

Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung

Gaussian Naive Bayes Algorithm Analysis Of Data Classification Of Heart Failure Patiens

Quswatun Hasanah¹, Hardian Oktavianto^{2*}, Yeni Dwi Rahayu³

¹ Mahasiswa Program studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email: meiapril350@gmail.com

²Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author
email: hardian@unmuhjember.ac.id

³Fakultas Teknik, Universitas Muhammadiyah Jember
email: yenidwirahayu@unmuhjember.ac.id

Abstrak

Penelitian ini menjelaskan penggunaan algoritma Gaussian Naive Bayes terhadap klasifikasi data pasien penderita gagal jantung. dataset yang digunakan dalam penelitian ini berasal dari Kaggle dengan jumlah variabel sebanyak 30 yang diperoleh dari data pasien penderita gagal jantung sebanyak 1000 orang. Pada tahap skenario uji coba dalam penelitian ini menggunakan Cross Fold Validation dengan nilai $k=2, 4, 5,$ dan 10 . Penentuan kriteria hasil klasifikasi terhadap output data asli (nilai aktual) menggunakan confusion matrix. Dari implementasi yang dilakukan menggunakan algoritma Gaussian Naive Bayes terhadap data pasien penderita gagal jantung diperoleh hasil validasi terbaik pada fold 10 tahap 9 dengan skenario 4 yaitu akurasi sebesar 69%, untuk nilai presisi terbaik terdapat pada fold 4 tahap 4 dengan skenario 2 sebesar 65,73% dan nilai recall tertinggi pada fold 10 tahap 4 skenario 4 sebesar 95,91%.

Kata Kunci: Klasifikasi, *Gaussian Naive Bayes*, Gagal Jantung, *Cross Validation*.

Abstract

This study describes the use of the Gaussian Naive Bayes algorithm for data classification of patients with heart failure. The dataset used in this study came from Kaggle with a total of 30 variables obtained from data on 1000 patients with heart failure. At the stage of the trial scenario in this study using Cross Fold Validation with values of $k = 2, 4, 5,$ and 10 . Determination of the classification results criteria for the original data output (actual value) using a confusion matrix. From the implementation carried out using the Gaussian Naive Bayes algorithm on patient data with heart failure, the best validation results were obtained at fold 10 stage 9 with scenario 4, namely accuracy of 69%, for the best precision value was in fold 4 stage 4 with scenario 2 of 65.73 % and the highest recall value in fold 10 stage 4 scenario 4 was 95.91%.

Keywords: *Classification, Gaussian Naive Bayes, Heart Failure, Cross Validation.*

1. PENDAHULUAN

Gagal Jantung memiliki arti sebagai suatu keadaan dimana jantung tidak mampu memompa darah yang cukup pada seluruh tubuh dengan merasakan ciri-ciri seperti sesak nafas ketika melakukan kegiatan atau ketika tidur dengan posisi terlentang tanpa menggunakan bantal, serta membengkak pada bagian tungkai bawah (PERKI, 2015). Gagal jantung

berkontribusi terhadap 287.000 kematian per tahun. Sekitar setengah dari orang yang mengalami gagal jantung meninggal dalam waktu lima tahun setelah di diagnosis. Negara Indonesia menduduki peringkat keempat penderita gagal jantung kongestif terbanyak di Asia Tenggara setelah negara Filipina, Myanmar dan Laos (Maulida, 2018).

Menurut diagnosa dokter atau tanda-tanda, estimasi tingkat penderita penyakit gagal

jantung tertinggi berada di Provinsi Jawa Barat sebanyak 96.487 orang (0,3%), sementara tingkat penderita paling rendah berada di Provinsi Kepulauan Bangka Belitung, yaitu sebanyak 945 orang (0,1%) (Kemenkes RI, 2014). Jumlah kematian pasien penderita gagal jantung yang telah mendapat perawatan berkisar di angka 2%-17% (Maggioni et al., 2013). Meskipun gagal jantung tidak dapat disembuhkan, banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup. Berbagai cara pencegahan yang dapat dilakukan ialah dengan memberikan pemahaman dan kesadaran terhadap masyarakat tentang cara mengenali gejalanya dan memberikan pemahaman secara sederhana agar masyarakat dapat mengetahuinya, selain itu diperlukan sikap profesional dalam penanganan kesehatan agar dapat melakukan tindakan dengan cepat dan tepat (Ponikowski et al., 2014). Seiring perkembangan zaman dalam dunia kesehatan terjadi pula perkembangan yang cukup pesat. Salah satunya penggunaan Machine Learning dan Data Mining dalam dunia kesehatan.

Salah satu teknik yang dapat diterapkan ialah teknik klasifikasi. Klasifikasi mirip dengan estimasi, hanya saja variabel target bersifat kategorik dari pada numerik. Teknik klasifikasi mempunyai banyak algoritma yang dapat digunakan dalam menyelesaikan permasalahan tergantung kesesuaian data yang akan di teliti (Larose & Larose, 2014). Adapun beberapa algoritma yang dapat digunakan dalam melakukan sebuah klasifikasi salah satunya adalah Algoritma Gaussian Naive Bayes dimana algoritma tersebut hanya membutuhkan data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi serta algoritma tersebut dapat diterapkan pada sebuah data continue yang bersifat numerik.

Pada penelitian dengan judul Analisis Metode K-Nearest Neighbors Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung menggunakan algoritma KNN dan teknik SMOTE menghasilkan akurasi sebesar 63,33% (hidayah nur umi & dkk, 2021). Pada penelitian dengan judul Klasifikasi Penderita penyakit

Parkinson menggunakan Metode Gaussian Naive Bayes dengan menggunakan data continue yang didapat dari UCI hasil akurasi sebesar 64,91% dan presisi sebesar 72,72% (Purwatiningsih & Nilogiri, 2020). Pada penelitian dengan judul Cancer Classification Using Gaussian Naive Bayes Algorithm dengan data continue yang didapat dari UCI menghasilkan akurasi sebesar 98% pada kanker payudara dan 90% pada kanker paru-paru (Kamel, dkk, 2019). Pada penelitian dengan judul Perbandingan Kinerja Algoritma Gaussian Naive Bayes dan K-Nearest Neighbors(KNN) Untuk Mengklasifikasi Penyakit Hepatitis C Virus(HCV) berupa data continue menghasilkan nilai akurasi 90,98%, presisi 69,91%, Recall 61,57%, dan f1 score 63,24% pada algoritma Gaussian Naive Bayes dan akurasi sebesar 91,80%, presisi 68,96%, recall 51,85%, dan f1 score sebesar 54,3% pada algoritma knn (Raharja, dkk 2021).

Berdasarkan penelitian yang dilakukan sebelumnya yakni pada penyakit parkinson dan kanker dengan menggunakan metode gaussian naive bayes menghasilkan akurasi yang cukup tinggi. Oleh karena itu dari serangkaian fakta tentang algoritma Gaussian Naive Bayes dan kemiripan data pada penelitian sebelumnya, maka pada penelitian ini penulis tertarik untuk meneliti dengan judul Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung.

2. TINJAUAN PUSTAKA

a. GAGAL JANTUNG

Gagal jantung merupakan sekumpulan gejala yang kompleks dimana seorang pasien harus memiliki ciri-ciri seperti, nafas pendek yang tipikal ketika istirahat atau ketika melakukan kegiatan diiringi atau tidak kelelahan), mengalami tanda retensi cairan (kongesti paru atau edema pada pergelangan kaki), dan terdapat bukti nyata dari gangguan struktur atau fungsi jantung ketika istirahat (PERKI, 2015).

Di negara maju secara ekonomi, hingga satu dari lima diperkirakan mengalami gagal jantung di beberapa titik dalam hidup mereka. dan bahkan lebih banyak orang akan

terpengaruh sebagai anggota keluarga, teman, atau profesional perawatan kesehatan. Di seluruh dunia, 17% -45% pasien gagal jantung yang mendapat perawatan di rumah sakit meninggal dalam 1 tahun setelah masuk dan beberapa meninggal dalam 5 tahun setelah masuk (Ponikowski et al., 2014).

Angka kematian pasien penderita gagal jantung yang telah mendapat perawatan berkisar di angka 2%-17% (Maggioni et al., 2013). Meskipun gagal jantung tidak dapat disembuhkan, banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup (Ponikowski et al., 2014). Menurut diagnosa atau tanda-tanda, estimasi tingkat penderita penyakit gagal jantung tertinggi berada di Provinsi Jawa Barat sebanyak 96.487 orang (0,3%), sementara tingkat penderita paling rendah berada di Provinsi Kepulauan Bangka Belitung, yaitu sebanyak 945 orang (0,1%) (Kemenkes RI, 2014).

b. DATA MINING

Data mining merupakan suatu proses pengumpulan informasi penting dari sebuah data yang jumlahnya cukup besar. Data mining dapat dikatakan sebuah kumpulan proses yang bertujuan menemukan nilai tambah berupa pengetahuan yang sejauh ini belum dapat ditemukan secara manual dari suatu kumpulan data (Bustami, 2014). Data mining dikatakan sebagai pelajaran apa yang telah terjadi dimasa lampau dan diterapkan dimasa mendatang agar mendapatkan hasil yang lebih baik. Secara khusus, data mining dikenal dengan banyak metode dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model (Naparini, 2016). Berdasarkan tugasnya data mining dibagi ke dalam beberapa kelompok yaitu sebagai berikut (Larose, 2014).

1. Deskripsi

Deskripsi mengidentifikasi pola yang tersembunyi secara tersembunyi dan mengubah pola tersebut menjadi suatu aturan yang mudah dimengerti oleh para ahli.

2. Estimasi

Estimasi memiliki kesamaan dengan prediksi, kecuali pada variabel estimasi lebih ke arah numerik.

3. Prediksi

Prediksi melakukan klasifikasi berdasarkan perilaku yang akan di prediksi pada masa mendatang.

4. Klasifikasi

Klasifikasi mirip dengan estimasi, hanya saja variabel target bersifat kategorik daripada numerik. Bentuk data mining yang memeriksa sekumpulan besar record, masing-masing record terdapat informasi tentang variabel target serta sekumpulan variabel input atau prediktor.

5. Pengklasteran

Pengklasteran yaitu sebuah pengelompokan data tanpa latar belakang kelas tertentu terhadap objek tersebut.

6. Asosiasi

Asosiasi dapat menemukan suatu atribut yang muncul dalam suatu waktu.

c. KLASIFIKASI

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan sebuah kelas atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis (Bustami, 2014).

d. NAIVE BAYES

Naive Bayes termasuk dalam pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa yang akan datang berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana asumsi kondisi antar variabel yang saling bebas. Klasifikasi Naive Bayes mengasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada kaitannya dengan ciri dari kelas lainnya (Bustami, 2014). Persamaan dari teorema Bayes adalah sebagai berikut.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2.1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui.

H : Hipotesis data X adalah suatu kelas spesifik.

$P(H|X)$: Probabilitas hipotesis H sesuai kondisi X (posteriori probability).

$P(H)$: Probabilitas hipotesis H (prior probability).

$P(X|H)$: Probabilitas X sesuai kondisi terhadap hipotesis H .

$P(X)$: Probabilitas X

e. GAUSSIAN NAIVE BAYES

Apabila ditemukan data dengan kontinyu, maka akan digunakan pendistribusian Gaussian Naive Bayes. Berikut ini adalah bentuk pendistribusian dari Gaussian Naive Bayes:

$$P(X_i) = x_i | Y = y_j = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2.2)$$

1. Tentukan data latih
2. Menghitung jumlah dan probabilitas, jika terdapat data numerik, maka:
3. Tentukan nilai mean dan standar deviasi dari masing-masing parameter yang termasuk data numerik.
4. Tentukan nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut. Menggunakan rumus 2.3.

$$f(w) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(w-\mu)^2}{\sigma^2}} \quad (2.3)$$

5. Menemukan nilai dalam tabel mean, standar deviasi dan probabilitas

a. Mean

Merupakan rata-rata nilai yang dihasilkan dari penjumlahan seluruh nilai pada masing-masing data, kemudian dibagi dengan banyak data yang ada. Rumus mean dapat dilihat pada persamaan 2.4.

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2.4)$$

b. Standar deviasi

Merupakan simpangan baku, seperti varian dan standar deviasi juga merupakan suatu ukuran dispersi. Standar deviasi cukup banyak digunakan karena memiliki satuan

ukuran data asalnya. Rumus standar deviasi dapat dilihat pada persamaan 2.5.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (2.5)$$

f. CONFUSION MATRIX

Confusion Matrix merupakan alat ukur yang digunakan untuk menilai seberapa akurat hasil klasifikasi yang dilakukan pada sebuah penelitian. TP dan TN digunakan saat klasifikasi menghasilkan sesuatu yang benar, sedangkan FP dan FN digunakan ketika klasifikasi menghasilkan kesalahan (Han, et.al, 2012). Apa yang digunakan sebagai pelatihan model. Berikut contoh pengukuran evaluasi dua kelas pada Tabel 1.

Tabel 1. Confusion Matrix Dua Kelas

| Kelas prediksi | Kelas aktual | |
|----------------|--------------|-------------|
| | Positif (+) | Negatif (-) |
| Positif (+) | TP | FP |
| Negatif (-) | FN | TN |

Sumber : Han (2020)

Langkah – langkah dalam evaluasi dimulai dari perhitungan akurasi. Akurasi menyatakan presentase yang sama pada data uji yang diklasifikasikan benar dalam model klasifikasi. Kemudian evaluasi dengan presisi dan recall, presisi merupakan tolak ukur kepastian, yaitu berapa presentase yang sama mendapat label positif ialah benar pada kenyataannya, dan recall merupakan tolak ukur kelengkapan yaitu berapa presentase yang sama positif yang menghasilkan label positif. Rumus akurasi, presisi, dan recall ditunjukkan pada Tabel 2.

Tabel 2. Rumus Akurasi, Presisi, dan Recall

| | |
|---------|-------------------------------------|
| akurasi | $\frac{TP + TN}{TP + TN + FP + FN}$ |
| presisi | $\frac{TP}{TP + FP}$ |
| recall | $\frac{TP}{TP + FN}$ |

Sumber : Han (2020)

g. CROSS VALIDATION

Cross validation merupakan sebuah metode yang digunakan untuk mengevaluasi dan perbandingan terhadap pembelajaran dari algoritma (learning algorithms) dengan cara membagi data menjadi dua bagian, satu bagian digunakan untuk training dan bagian lainnya digunakan sebagai testing (Wibowo & Jumiaty, 2016).

h. JUPYTER NOTEBOOK

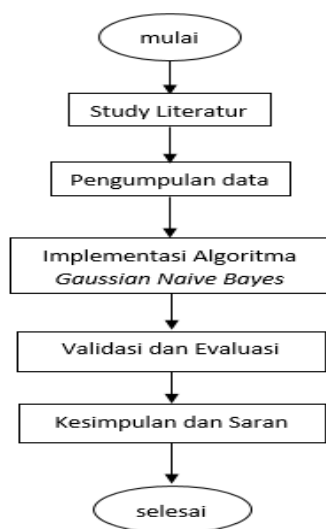
Menurut Avila dkk dalam situs <https://jupyter.org/>, “ Jupyter Notebook merupakan aplikasi terbuka(open source) yang berfungsi untuk membuat dan membagikan dokumen berisi persamaan, visualisasi data dan lainnya. Jupyter Notebook merupakan sebuah proyek nirlaba, sumber terbuka, yang lahir dari proyek Ipython pada tahun 2014 yang berevolusi untuk mendukung ilmu data interaktif dan komputasi ilmiah berbagai bahasa pemrograman.

3. METODELOGI PENELITIAN

a. Tahapan Penelitian

Dalam mengerjakan Tugas Akhir tersebut perlu adanya langkah-langkah penelitian yang mendukung dan maksimal dalam penyelesaian. Berikut ini adalah diagram metode penelitian yang berisi tahapan penelitian yang akan dilakukan sebagai berikut:

Gambar 1. Diagram alur penelitian



Sumber : hasil gambar menurut alur penelitian sendiri

b. Studi Literatur

Dalam penelitian penulis mencari beberapa referensi teori yang relevan terhadap kasus dan Mengetahui sebuah permasalahan. Studi literatur yang dilakukan dalam penelitian tersebut yaitu dengan melakukan pencarian jurnal-jurnal dari penelitian yang telah dilakukan sebelumnya yang dapat dijadikan sebagai pendukung dan dapat dijadikan rujukan atau referensi yang dapat memperkuat argumentasi-argumentasi yang ada.

c. Pengumpulan data

Pengumpulan data yang dilakukan dalam penelitian ini diperoleh melalui Kaggle yang terdiri dari data pasien penyakit gagal jantung pada tahun 2020 yang dapat diakses melalui link berikut ini :

<https://www.kaggle.com/jackleenrasmymbareh/heart-failure>

Dataset berupa jumlah kasus pasien penyakit gagal jantung pada tahun 2020 yang terdiri dari 1000 dan terbagi atas 508 pasien hidup dan 492 pasien dinyatakan meninggal.

4. HASIL DAN PEMBAHASAN

a. Hasil Klasifikasi

Pada proses klasifikasi dilakukan dengan menggunakan Algoritma *Gaussian Naive Bayes*. Pengujian dilakukan menggunakan *K-fold* dengan nilai K 2, 4, 5, dan 10. Klasifikasi ini dilakukan untuk mengetahui nilai akurasi, presisi, dan recall.

1. Pengujian menggunakan *kfold 2*

Dalam pengujian menggunakan *kfold 2* menggunakan algoritma *Gaussian Naive Bayes* mendapatkan hasil terbesar dengan akurasi 60,2%, presisi 57,10%, dan recall 88,23% pada percobaan ke 2.

Tabel 3. Confusion Matrix Algoritma GNB *Kfold 2*

| Kelas prediksi | Kelas Aktual | |
|---------------------|---------------------|---------------------|
| | <i>Positive (1)</i> | <i>Negative (0)</i> |
| <i>Positive(1)</i> | 76 | 169 |
| <i>Negative (0)</i> | 30 | 225 |

Sumber : hasil perhitungan *jupyter notebook*

2. Pengujian menggunakan *kfold 4*

Dalam pengujian menggunakan *kfold 4* menggunakan algoritma *Gaussian Naive Bayes* mendapatkan hasil terbesar dengan akurasi

67,6%, presisi 65,73%, dan recall 85,40% pada percobaan ke 4.

Tabel 4. Confusion Matrix Algoritma GNB Kfold 4

| Kelas prediksi | Kelas Aktual | |
|----------------|--------------|--------------|
| | Positive (1) | Negative (0) |
| Positive(1) | 52 | 61 |
| Negative (0) | 20 | 117 |

Sumber : hasil perhitungan jupyter notebook

3. Pengujian menggunakan kfold 5

Dalam pengujian menggunakan kfold 5 menggunakan algoritma Gaussian Naive Bayes mendapatkan hasil terbesar dengan akurasi 67%, presisi 65%, dan recall 84,25% pada percobaan ke 5.

Tabel 5. Confusion Matrix Algoritma GNB Kfold 5

| Kelas prediksi | Kelas Aktual | |
|----------------|--------------|--------------|
| | Positive (1) | Negative (0) |
| Positive(1) | 43 | 49 |
| Negative (0) | 17 | 91 |

Sumber : hasil perhitungan jupyter notebook

4. Pengujian menggunakan kfold 10

Dalam pengujian menggunakan kfold 10 menggunakan algoritma Gaussian Naive Bayes mendapatkan hasil terbesar dengan akurasi 69%, presisi 62,12%, dan recall 87,23% pada percobaan ke 9.

Tabel 6. Confusion Matrix Algoritma GNB Kfold 10

| Kelas prediksi | Kelas Aktual | |
|----------------|--------------|--------------|
| | Positive (1) | Negative (0) |
| Positive(1) | 28 | 25 |
| Negative (0) | 6 | 41 |

Sumber : hasil perhitungan jupyter notebook

b. Hasil Evaluasi dan Validasi

Proses Validasi menggunakan K-Fold Cross Validation yang membagi dataset menjadi beberapa bagian. Penentuan nilai K-fold yang digunakan pada penelitian ini bernilai 2, 4, 5, dan 10.

Tabel 7. Hasil Skenario Uji Coba

| Skenario uji coba | Kfold | Tahap ke | Akurasi | Presisi | Recall | Jumlah data training |
|-------------------|-------|----------|---------|---------|--------|----------------------|
| 1 | 2 | 1 | 56,80% | 52,75% | 84,81% | 500 |
| | | 2 | 60,20% | 57,10% | 88,23% | 500 |
| 2 | 4 | 1 | 58,80% | 54,05% | 84,74% | 250 |
| | | 2 | 58,80% | 54,04% | 89,91% | 250 |
| | | 3 | 52,40% | 49,74% | 82,20% | 250 |
| | | 4 | 67,60% | 65,73% | 85,40% | 250 |
| 3 | 5 | 1 | 57,50% | 52,41% | 82,60% | 200 |
| | | 2 | 62% | 57,92% | 93,13% | 200 |
| | | 3 | 54,50% | 51,78% | 61,05% | 200 |
| | | 4 | 49% | 47,95% | 86,31% | 200 |
| | | 5 | 67% | 65% | 84,25% | 200 |
| 4 | 10 | 1 | 58% | 53,48% | 95,83% | 100 |
| | | 2 | 58% | 51,61% | 72,72% | 100 |
| | | 3 | 59% | 57,31% | 88,67% | 100 |
| | | 4 | 64% | 58,02% | 95,91% | 100 |
| | | 5 | 54% | 48% | 83,72% | 100 |
| | | 6 | 52% | 53,84% | 53,84% | 100 |
| | | 7 | 50% | 47,61% | 86,95% | 100 |
| | | 8 | 53% | 51,25% | 83,67% | 100 |
| | | 9 | 69% | 62,12% | 87,23% | 100 |
| | | 10 | 61% | 63,75% | 83,60% | 100 |

Sumber : hasil perhitungan jupyter notebook

Hasil nilai akurasi, presisi, dan recall pada dataset pasien penderita gagal jantung menggunakan klasifikasi Gaussian Naive Bayes (GNB). Dari semua skenario uji coba didapatkan informasi untuk akurasi tertinggi didapatkan pada skenario 4 dengan K-fold 10 dengan tahap ke – 9 bernilai 69%, sedangkan untuk presisi hasil tertinggi didapatkan pada skenario 2 dengan K-fold 4 pada tahap ke – 4 bernilai 65,73%, dan untuk recall tertinggi didapatkan pada skenario 4 dengan K-fold 10 pada tahap ke – 4 bernilai 95,91%. Sehingga skenario paling optimal pada dataset pasien penderita gagal jantung terjadi di skenario 4 dengan K-fold 10 pada tahap ke – 9 dengan nilai akurasi 69%, presisi 65,73% pada skenario 2 dengan K-fold 4 pada tahap ke 4, dan recall 95,91% pada skenario 4 dengan K-fold 10 pada tahap ke-4.

5. KESIMPULAN DAN SARAN

1. Kesimpulan

Berdasarkan penelitian hasil prediksi data pasien penderita gagal jantung menggunakan algoritma Gaussian Naive Bayes, maka dapat diambil kesimpulan bahwa nilai akurasi dan recall yang diperoleh dari penerapan algoritma Gaussian Naive Bayes dan dilakukan pengujian menggunakan K-fold, 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%.

2. Saran

Pada penelitian ini masih sangat jauh dari kata sempurna, oleh karena itu penulis berkenan jika pembaca ingin melakukan pengembangan pada penelitian ini. Beberapa aspek yang dapat dikembangkan dalam penelitian ini untuk memperoleh hasil yang lebih baik antara lain :

- Dapat membandingkan menggunakan metode klasifikasi lain untuk menemukan tingkat akurasi, presisi, dan recall yang lebih baik.
- Dapat menambahkan teknik ensemble untuk menemukan atau meningkatkan nilai akurasi, presisi, dan recall.

DAFTAR PUSTAKA

- Bustami. (2014). *PENERAPAN ALGORITMA NAIVE BAYES*. 8(1), 884–898.
- Fitria. (2013). Gambaran Tipe Kepribadian pada Pasien Gagal Jantung Kongestif di RSUD Dr. Moewardi Surakarta. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Han, J., Kamber, M., & Pei, J. (2012). Third Edition: Data Mining Concepts and Techniques. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. Retrieved from <http://library.books24x7.com/toc.aspx?bkid=44712>
- hidayah nur umi, & dkk. (2021). hidayah nur umi dkk. *NALISIS METODE K NEAREST NEIGHBOR TERHADAP KLASIFIKASI DATA PASIEN PENDERITA GAGAL JANTUNG*, 28, 145–158. Retrieved from <http://www.riss.kr/link?id=A99932365>
- <https://jupyter.org/>
- <https://www.kaggle.com/jackleenrasmybar/eh/heart-failure>
- Kamel, H., Abdulah, D., & Al-Tuwaijari, J. M. (2019). Cancer Classification Using Gaussian Naive Bayes Algorithm. *Proceedings of the 5th International Engineering Conference, IEC 2019*, 165–170. <https://doi.org/10.1109/IEC47844.2019.8950650>
- Kemendes RI. (2014). Situasi kesehatan jantung. *Pusat Data Dan Informasi Kementerian Kesehatan RI*, 3. <https://doi.org/10.1017/CBO9781107415324.004>
- Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*.
- Maggioni, A. P., Dahlström, U., Filippatos, G., Chioncel, O., Leiro, M. C., Drozd, J., ... Asserlund, B. (2013). EURObservational Research Programme: Regional differences and 1-year follow-up results of the Heart Failure Pilot Survey (ESC-HF Pilot). *European Journal of Heart Failure*, 15(7), 808–817. <https://doi.org/10.1093/eurjhf/hft050>
- Naparin, H. (2016). *KLASIFIKASI PEMINATAN SISWA SMA MENGGUNAKAN*. 02(01), 25–32.
- PERKI. (2015). Current capacity and thermal transport in carbon nanofiber interconnects. *4th IEEE International Conference on Nano/Micro Engineered and Molecular Systems, NEMS 2009*, 848–853. <https://doi.org/10.1109/NEMS.2009.5068708>
- Ponikowski, P., Anker, S. D., AlHabib, K. F., Cowie, M. R., Force, T. L., Hu, S., ... Filippatos, G. (2014). Heart failure: preventing disease and death worldwide. *ESC Heart Failure*, 1(1), 4–25. <https://doi.org/10.1002/ehf2.12005>
- Purwatiningsih, I., & Nilogiri, A. (2020). *No Title* □□□ □□□□□□□□ □□ □□□ □□ □□□□□□ 291–294.

Raharja, K. Y., Oktavianto, H., & Umilasari, R. (2021). *PERBANDINGAN KINERJA ALGORITMA GAUSSIAN NAIVE BAYES DAN K-NEAREST NEIGHBOR (KNN) UNTUK MENGLASIFIKASI PENYAKIT HEPATITIS C VIRUS (HCV) C . Cara untuk melakukan diagnosa dini terhadap penyakit Hepatitis C adalah potensial dan berguna yang tersimpan da. 1–12.*

Wibowo, A. P., & Jumiati, E. (2016). *Sentiment Analysis Masyarakat Pekalongan Terhadap Pembangunan Jalan Tol Pemalang-Batang Di Media Sosial. (0285).*