

# Comparison Performance of K-NN and NBC Algorithm for Classification of Heart Disease

## Perbandingan Kinerja Algoritma K-NN dan NBC Untuk Klasifikasi Penyakit Jantung

Ravega Widyawati Putri<sup>1</sup>, Aidina Ristyawan<sup>2</sup>, M. Najibulloh Muzaki<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, Fakultas Teknik, Universitas Nisantara PGRI Kediri

E-mail: <sup>1</sup>ravegaputri9@gmail.com, <sup>2</sup>aidinaristi@unpkediri.co.id, <sup>3</sup>m.n.muzaki@gmail.com

**Abstract** – Heart disease is one of the most common diseases suffered by Indonesian people, in 2018 there were 1.5% of people with heart disease according to a doctor's diagnosis. With so many cases that have occurred, this research was conducted with the aim of implementing an algorithm to help classify whether a person has the potential to have heart disease or not. The algorithms that will be used are K-Nearest Neighbor (K-NN) and Naive Bayes Classifier (NBC). The accuracy of the two algorithms will be compared, so that it can be seen which algorithm is better in classifying heart disease. The final results obtained show that the NBC algorithm has a higher accuracy rate of 86.17%, while K-NN has a slightly lower accuracy rate of 85.11%. By implementing a data mining algorithm to compare the K-NN and NBC algorithms using the heart disease data set, it can be used to help determine whether a person can be classified as having heart disease or not. And by comparing the accuracy of the two algorithms, it can be seen which algorithm is considered better in handling cases of heart disease classification.

**Keywords** — data mining, heart disease, K-NN, NBC

**Abstrak** – Penyakit jantung (*Heart Disease*) adalah salah satu penyakit yang banyak diderita oleh masyarakat Indonesia, pada tahun 2018 tercatat 1,5% orang yang mengidap penyakit jantung sesuai diagnosa dokter. Dengan banyaknya kasus yang telah terjadi maka dilakukanlah penelitian ini dengan tujuan mengimplementasikan algoritma untuk membantu mengklasifikasikan apakah seseorang berpotensi mengidap penyakit jantung atau tidak. Algoritma yang akan digunakan adalah *K-Nearest Neighbor* (KNN) dan *Naive Bayes Classifier* (NBC). Akurasi dari kedua algoritma tersebut akan dibandingkan, sehingga dapat diketahui algoritma mana yang lebih baik dalam mengklasifikasikan penyakit jantung. Hasil akhir yang diperoleh menunjukkan bahwa algoritma NBC memiliki tingkat akurasi yang lebih tinggi yaitu 86,17%, sedangkan K-NN memiliki tingkat akurasi sedikit lebih rendah yaitu 85,11%. Dengan mengimplementasikan algoritma K-NN dan NBC menggunakan *dataset heart disease* dapat digunakan untuk membantu mengetahui apakah seseorang dapat diklasifikasikan menghidap penyakit jantung atau tidak, serta dengan membandingkan akurasi kedua algoritma tersebut dapat diketahui algoritma mana yang dinilai lebih baik dalam menangani kasus klasifikasi penyakit jantung.

**Kata Kunci** — data mining, K-NN, NBC, penyakit jantung

### 1. PENDAHULUAN

Jantung merupakan salah satu organ vital yang berfungsi memompa darah untuk memenuhi kebutuhan oksigen dan nutrisi ke seluruh tubuh. Penyakit jantung adalah sebutan umum yang menggambarkan gangguan terhadap fungsi jantung [1]. Penyakit jantung (*Heart Disease*) adalah salah satu penyakit yang cukup banyak diderita oleh masyarakat terutama Lansia (orang lanjut usia), di Indonesia sendiri berdasarkan data Riskesdas (Riset Kesehatan Dasar) pada tahun 2018 tercatat 1,5%

penyakit jantung sesuai diagnosis dokter. Oleh karena itu diterapkan sebuah algoritma yang dapat memprediksi apakah seseorang dapat dikatakan mengidap penyakit jantung atau tidak, salah satunya yaitu dengan teknik *data mining* menggunakan metode klasifikasi.

Klasifikasi adalah sebuah proses menemukan satu set pola atau fungsi yang menggambarkan dan memisahkan satu kelas data dengan data yang lain untuk menyatakan objek yang dimasukkan dalam kategori tertentu yang telah ditentukan [2]. Untuk proses klasifikasi membutuhkan beberapa data yang dapat digunakan untuk proses klasifikasi, data-data yang dibutuhkan seperti usia, kadar gula darah, kolesterol dan lain sebagainya, data-data tersebut nanti akan diolah menggunakan salah satu metode klasifikasi untuk mengetahui hasilnya.

Ada beberapa metode klasifikasi yang sering digunakan salah satunya yaitu algoritma *K-nearest Neighbor* (K-NN) dan *Naïve Bayes Classifier* (NBC). Dalam penelitian ini akan menggunakan dua algoritma tersebut yang bertujuan untuk mengimplementasikan algoritma K-NN dan NBC untuk mengklasifikasikan *dataset* penyakit jantung, serta untuk membandingkan dua algoritma tersebut mana yang dinilai memiliki tingkat kinerja yaitu akurasi atau kemiripan yang tinggi terhadap *dataset*.

Dalam penelitian terdahulu telah dilakukan beberapa studi terkait dengan dua algoritma yaitu K-NN dan NBC antara lain, pada penelitian yang dilakukan oleh Waliyansyah dan Fitriyah pada tahun 2019 [3] tentang membandingkan akurasi antara algoritma NBC dan K-NN untuk mengklasifikasikan citra kayu jati. Kemudian penelitian yang dilakukan oleh Maricar dan Dian Pramana pada tahun 2019 [4] tentang membandingkan akurasi antara algoritma NBC dan K-NN pada klasifikasi untuk meramalkan status pekerjaan alumni ITB STIKOM Bali, dan juga penelitian tentang membandingkan kinerja NBC dan K-NN untuk mengklasifikasikan artikel Bahasa Indonesia yang dilakukan oleh Devita, Herwanto, dan Wibawa pada tahun 2018 [5], dan yang terakhir penelitian yang sama menggunakan *dataset heart disease* atau penyakit jantung yaitu penelitian yang dilakukan oleh Nawawi, Purnama, dan Hikmah pada tahun 2019 [6] tentang komparasi algoritma NN (*Neural Network*) dan NBC untuk memprediksi penyakit jantung.

Dari beberapa penelitian yang telah dilakukan oleh peneliti terdahulu belum pernah ada peneliti yang mengambil masalah tentang membandingkan hasil akurasi antara algoritma K-NN dan NBC dengan menggunakan *dataset heart disease*. Sehingga diputuskan untuk penelitian ini menggunakan judul “Perbandingan kinerja Algoritma K-NN dan NBC untuk klasifikasi penyakit jantung” sebagai judul skripsi dan jurnal, hasil yang diharapkan dari penelitian ini adalah dapat menghasilkan kinerja atau akurasi terbaik pada kedua algoritma, dimana algoritma dengan akurasi terbaik akan bermanfaat dalam membantu memprediksi seseorang yang berpotensi mengidap penyakit jantung dan yang tidak berpotensi mengidap penyakit jantung.

## 2. METODE PENELITIAN

### 2.1. Jenis dan Lokasi Penelitian

Jenis penelitian yang digunakan dalam penelitian ini yaitu metode percobaan (Eksperimental), metode eksperimen adalah metode terbaik untuk menetapkan sebab akibat untuk melakukan eksperimen yang dirancang dengan hati-hati di mana dampak variabel tersembunyi yang mungkin dikendalikan [7]. Metode ini dipilih karena akan membandingkan dua algoritma yaitu algoritma NBC dan algoritma K-NN, di mana pada salah satu algoritma yaitu algoritma K-NN akan mencoba menggunakan beberapa nilai K (jumlah tetangga terdekat) yang berbeda pada setiap prosesnya untuk mendapatkan nilai akurasi tertinggi. Ada beberapa faktor yang terkait dengan penelitian yang menggunakan metode eksperimental salah antara lain [8]: *Independent Variable* (VI), *Dependent Variable* (DV), *Experimental Condition* (group), *Confounding variable*, dan *An uncontrolled variable*. Pada penelitian ini yang termasuk faktor *Independent Variable* adalah nilai *range* (jarak minimal dan maksimal) pada proses normalisasi untuk metode KNN dan nilai K (jumlah tetangga terdekat) pada metode K-NN, kemudian untuk faktor *Dependent Variable* adalah nilai *performance* yang akan dihasilkan seperti nilai *accuracy*, *precision* dan *recall*.

## 2.2. Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan pada penelitian ini adalah dengan cara mencari bahan yang mendukung dalam pendefinisian permasalahan melalui internet (Penelitian lain, jurnal, proposal atau skripsi) serta buku yang erat kaitannya dengan objek permasalahan. Untuk penelitian ini menggunakan *dataset* publik yang diunduh dari <https://www.kaggle.com>. *Dataset* terdapat dua macam yaitu *dataset* privat dan *dataset* publik. *Dataset* privat adalah *dataset* yang dapat diambil dari organisasi atau tempat yang dijadikan objek penelitian, sedangkan *dataset* publik adalah *dataset* yang dapat diambil dari *repository* publik yang telah disepakati oleh para peneliti [9].

## 2.3. Sumber Data

Sumber data yang digunakan pada penelitian ini yaitu menggunakan *dataset public* yang diunduh dari <https://www.kaggle.com>. *Dataset* tersebut terdiri dari 75 atribut tetapi semua eksperimen yang dipublikasikan mengacu pada penggunaan subset sebanyak 14 atribut saja. 14 atribut tersebut antara lain seperti yang disajikan oleh Tabel 1.

Tabel 1. Atribut Data Set

No	Atribut	Keterangan
1	Age	Usia dalam hitungan tahun
2	Sex	1 = male, 0 = female
3	Chest pain type	1 = typical angina, 2 = atypical angina, 3 = non-angina pain, 4 = asymptomatic
4	BP	Tekanan darah (dalam mmHg saat masuk ke rumah sakit)
5	Cholesterol	Kolesterol serum dalam mg/dl
6	FBS over 120	Fasting blood sugar (1 = true, 0 = false)
7	EKG result	Hasil elektrokardiografi (0 = normal, 1 =memiliki kelainan gelombang ST-T, 2 = menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri menurut kriteria estes)
8	Max HR	Maximum heart rate achieved
9	Exercise angina	1 = yes, 0 = no
10	ST depression	Depresi ST yang diinduksi oleh olahraga relatif terhadap istirahat
11	Slope of ST	Value 1 = upsloping, Value 2 = flat, Value 3 = downsloping
12	Number of vessels fluoro	Jumlah pembuluh darah besar (0-3) yang diwarnai dengan fluoroscopy
13	Thallium	3 = normal, 6 = fixed defect, 7 = reversible defect
14	Heart Disease	Presence (berpotensi mengidap penyakit jantung), Absence (tidak berpotensi mengidap penyakit jantung)

## 2.4. Metode Analisa Data

Metode analisa data yang digunakan pada penelitian ini adalah *Data Mining* dengan Algoritma K-NN dan NBC. *Data mining* adalah proses untuk mengidentifikasi pola pada data yang memiliki potensi dan memiliki nilai guna [10]. Untuk mengelola *dataset* dalam jumlah besar dibantu dengan aplikasi *RapidMiner Studio ver.9.10* untuk mengelola data dan memprediksi apakah seseorang berpotensi mengidap penyakit jantung atau tidak. *Dataset* yang didapat dari *kaggle* sebanyak 270 data kemudian dibagi menggunakan metode *Hold Out* (metode pembagian data secara acak) untuk dibagi menjadi data *testing* sebanyak 35% (94 data) dan data *training* sebanyak 65% (176 data).

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini akan menggunakan langkah-langkah dalam *Knowledge Discovery in Database (KDD)*, yaitu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan pola atau hubungan dalam sebuah set data yang berukuran besar, atau bisa juga diartikan sebagai proses yang terorganisir berguna untuk menemukan pola baru dan dapat dimengerti dari kumpulan data yang besar dan kompleks [11]. Adapun langkah-langkah yang akan digunakan dalam mengklasifikasikan *dataset heart disease* menggunakan algoritma NBC dan K-NN adalah sebagai berikut:

#### 3.1. Metode Analisa Data

Pada tahap ini yaitu tahap untuk memahami *dataset* yang akan digunakan dalam penelitian. *Dataset* yang akan digunakan adalah *dataset* tentang *heart disease* yang terdiri dari 75 atribut tetapi semua eksperimen yang dipublikasikan mengacu pada penggunaan subset sebanyak 14 atribut saja. Adapun karakteristik dari *dataset* tersebut antara lain:

Tabel 2. Karakteristik *Dataset Heart Disease*

<i>Dataset Characteristics</i>	<i>Multivariate</i>
<i>Attribute Characteristics:</i>	<i>Categorical, Integer, Real</i>
<i>Associated Taks :</i>	<i>Classification</i>
<i>Nuber of Attributes:</i>	75

#### 3.2. Metode Analisa Data

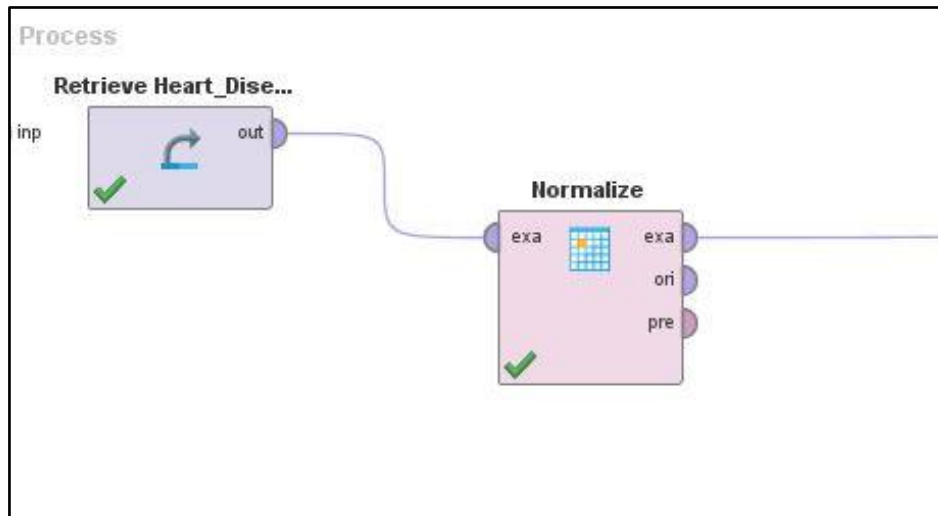
Data yang digunakan dalam penelitian ini adalah *dataset heart disease* yang terdiri dari 14 atribut data, dan target data yang akan digunakan adalah kolom *heart disease* yang berisi *Presence* dan *Absence*. *Presence* yang berarti seseorang yang diklasifikasikan berpotensi mengidap penyakit jantung, sedangkan *Absence* adalah seseorang yang diklasifikasikan tidak berpotensi mengidap penyakit jantung, berikut adalah tabel *dataset* yang akan digunakan.

Tabel 3. *Dataset Heart Disease*

<i>Age</i>	<i>Sex</i>	<i>Chest pain type</i>	<i>BP</i>	<i>Chol</i>	<i>FBS over 120</i>	<i>EKG result</i>	<i>Max HR</i>	<i>Exang</i>	<i>ST depression</i>	<i>Slope of ST</i>	<i>No. of vessels fluro</i>	<i>Thal</i>	<i>Heart disease</i>
70	1	4	130	322	0	2	109	0	2.4	2	3	3	<i>Presence</i>
57	1	2	124	261	0	0	141	0	0.3	1	0	7	<i>Presence</i>
64	1	4	128	263	0	0	105	1	0.2	2	1	7	<i>Absence</i>
74	0	2	120	269	0	2	121	1	0.2	1	1	3	<i>Absence</i>
65	1	4	120	177	0	0	140	0	0.4	1	0	7	<i>Absence</i>

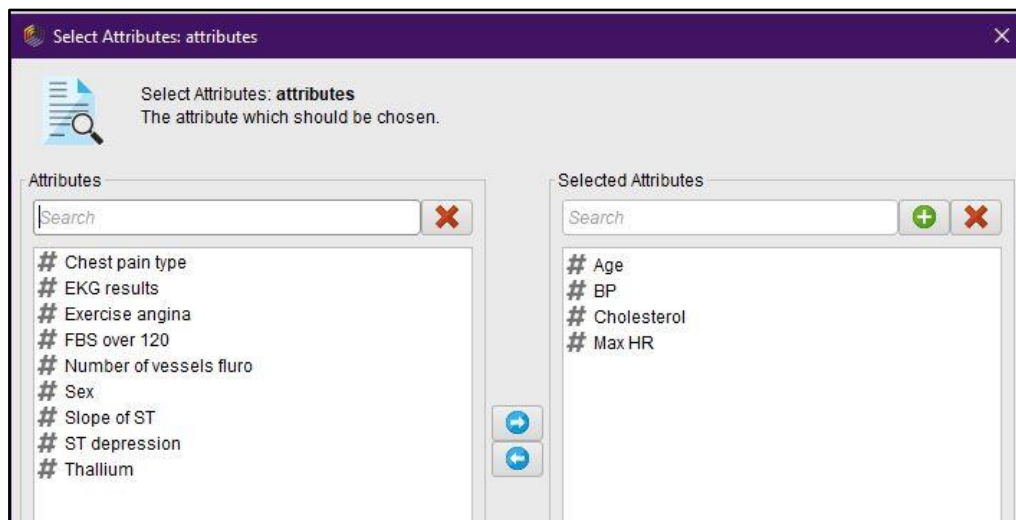
#### 3.3. Data Transformation

Pada tahap *transformation* ini dilakukan proses normalisasi pada data khususnya pada proses K-NN, proses normalisasi ini digunakan untuk meminimalkan range/jarak nilai dari masing-masing data, karena setiap fitur yang ada pada *data set heart disease* tersebut tidak sama yaitu ada data yang bernilai satuan, puluhan dan ratusan, sehingga diperlukan proses normalisasi untuk meminimalkan jarak. Gambar 1 adalah operator normalisasi pada aplikasi *RapidMiner*.



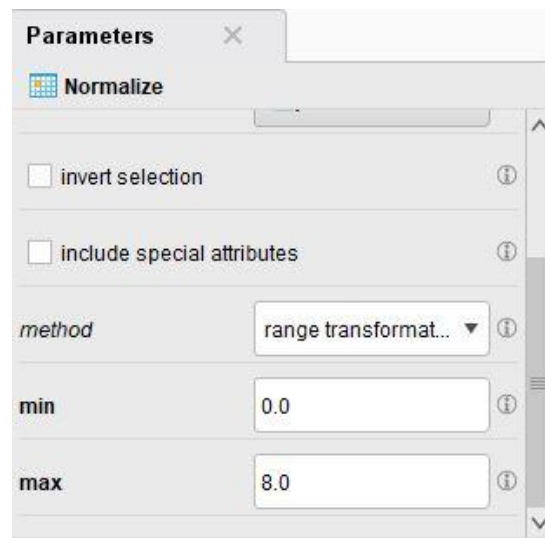
Gambar 1. Pengaturan Normalisasi Menggunakan *RapidMiner*

Sebelum menjalankan proses normalisasi maka menentukan terlebih dahulu atribut mana saja yang perlu dilakukan proses normalisasi seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Menentukan Atribut Yang Dinormalisasi

Setelah menentukan atribut mana saja yang akan dinormalisasi selanjutnya menentukan jarak *Min* dan *Max* seperti yang ditunjukkan oleh Gambar 3. Pada penelitian ini memakai jarak *Min* = 0 dan *Max* = 8.

Gambar 3. Menentukan nilai *Min-Max*

Berikut merupakan data yang sudah dinormalisasi, bisa dilihat data pada kolom *age*, *BP*, *cholesterol* dan *Max HR* yang tadinya bernilai puluhan dan ratusan (pada Tabel 3) setelah dinormalisasi menjadi data bernilai satuan atau desimal:

Tabel 3. Data setelah dinormalisasi

<i>Age</i>	<i>Sex</i>	<i>Chest pain type</i>	<i>BP</i>	<i>Chol</i>	<i>FBS over 120</i>	<i>EKG result</i>	<i>Max HR</i>	<i>Exang</i>	<i>ST depression</i>	<i>Slope of ST</i>	<i>No. of vessels fluro</i>	<i>Thal</i>	<i>Heart disease</i>
5.9	1	4	2.3	3.1	0	2	2.03	0	2.7	2	3	3	<i>Presence</i>
5.5	1	2	1.3	7	0	0	4.7	0	1.8	1	0	7	<i>Presence</i>
4.8	1	4	1.9	2.1	0	0	3.7	1	0.3	2	1	7	<i>Absence</i>
5.1	0	2	2.2	2.1	0	2	1.8	1	0.2	1	1	3	<i>Absence</i>
6.6	1	4	1.7	2.2	0	0	2.6	0	0.2	1	0	7	<i>Absence</i>

### 3.4. Choosing the Suitable Data Mining Tasks

Proses *data mining* yang akan digunakan pada penelitian ini adalah klasifikasi, karena sesuai dengan judul penelitian yaitu untuk mengklasifikasikan apakah seseorang dapat dikatakan mengidap penyakit jantung atau tidak.

### 3.5. Choosing the Suitable Data Mining Algorithm

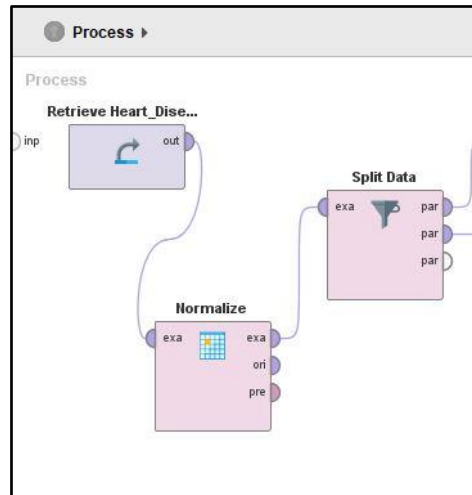
Algoritma yang digunakan untuk diterapkan pada *dataset* adalah algoritma K-NN dan NBC. Dengan dibantu oleh *tools RapidMiner Studio ver.9.10* untuk mengolah data.

### 3.6. Employing Data Mining Algorithm

#### 3.6.1. Proses Pembagian Data

Sebelum data diaplikasikan menggunakan algoritma K-NN dan NBC, data yang secara keseluruhan berjumlah 270 data akan dinormalisasi kemudian dibagi terlebih dahulu menggunakan

teknik *Hold Out* (pembagian data secara acak), pada *RapidMiner* akan menggunakan operator bernama *split data* untuk membagi *data training* dan *data testing* seperti Gambar 4.

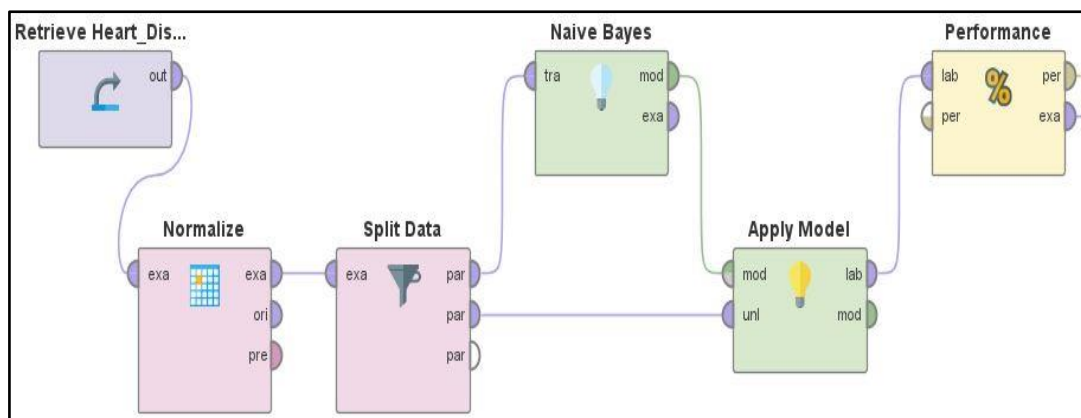


Gambar 4. *Split Data*

Kemudian data akan dibagi menjadi *data training* sebanyak 0.65 atau 65% menjadi 176 data dan *data testing* sebanyak 0.35 atau 35% menjadi 94 data.

### 3.6.2. Proses Implementasi Algoritma NBC

Alur yang dapat dilakukan jika menggunakan *tools RapidMiner* adalah setelah operator *split data* kemudian dapat ditambahkan operator *Naive Bayes* dan *Apply Model*. *Data training* sebanyak 176 data akan disambungkan pada operator *Naive Bayes* untuk dipelajari terlebih dahulu, kemudian untuk *data testing* dari *split data* langsung disambungkan ke *Apply Model* dan data yang sudah dipelajari oleh operator *Naive Bayes* akan disambungkan *Apply Model* (Gambar 5) dan yang terakhir adalah operator *performance* yang berfungsi untuk memberikan daftar nilai kinerja secara otomatis sesuai dengan tugas yang diberikan (Gambar 6).



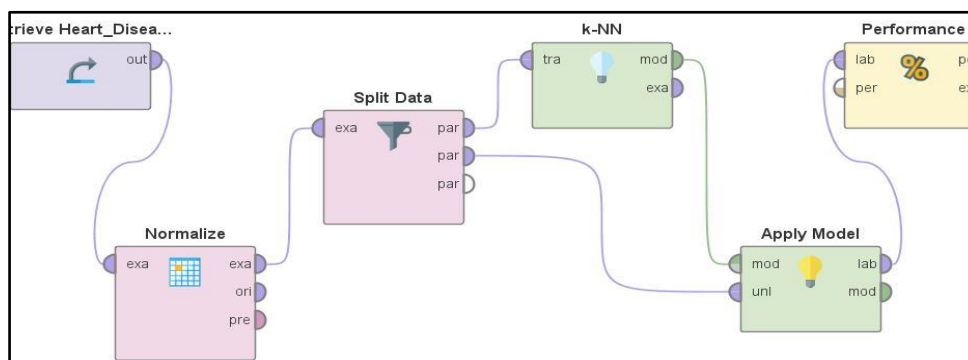
Gambar 5. Proses Implementasi Algoritma NBC pada *RapidMiner*

accuracy: 86.17%			
	true Presence	true Absence	class precision
pred. Presence	37	8	82.22%
pred. Absence	5	44	89.80%
class recall	88.10%	84.62%	

Gambar 6. Hasil Kinerja Menggunakan Algoritma NBC

### 3.6.3. Proses Implementasi Algoritma K-NN

Proses awal sama seperti dengan proses pada algoritma NBC yaitu proses pembagian *data training* dan *data testing* terlebih dahulu. Kemudian setelah operator *split data* ditambah dengan operator *normalize*, tujuannya untuk meminimalkan *range*/jarak nilai dari masing-masing data, karena fitur pada *dataset* yang tidak sama yaitu bernilai satuan, puluhan dan ratusan. Proses normalisasi pada penelitian ini menggunakan metode *range transformation* dengan nilai *min* 0 dan nilai *max* 8 serta akan mencoba menghitung kinerja dengan nilai *K* yang berbeda yaitu 3, 5, 7 dan 9, dari perbedaan jarak tersebut akan menghasilkan nilai kinerja yang berbeda-beda, yang di kemudian akan dicari nilai akurasi tertinggi dan akan dibandingkan dengan algoritma NBC (Gambar 7).



Gambar 7. Proses Implementasi Algoritma K-NN Pada RapidMiner

Pada implementasi K-NN ini menerapkan algoritma eksperimental untuk mendapatkan nilai dengan akurasi tertinggi, dan hasil yang didapat untuk nilai akurasi tertinggi yaitu dengan menetapkan nilai *K* sebesar 9 dan *range*/ jarak minimal 0, maksimal 8, *measure type* menggunakan *Numerical Measure* dan jarak yang dipakai adalah jarak *default* yaitu *Euclidean Distance*. Sebelum mendapatkan hasil tertinggi dari algoritma K-NN yang telah ditetapkan tersebut, maka terlebih dahulu dilakukan beberapa kali percobaan dengan nilai *K* 3, 5, 7 dan 9 dan didapatkan hasil akhir yang disajikan oleh Tabel 5 dan Gambar 8.

Tabel 5. Data Hasil Eksperimen Pada Algoritma K-NN

<i>Range</i>	Nilai <i>K</i>	Accuracy	Precision	Recall
Min = 0 dan Max = 8	3	80,85%	76,19%	80,00%
	5	81,91%	76,19%	82,05%
	7	82,98%	76,19%	84,21%
	9	85,11%	80,95%	85,00%



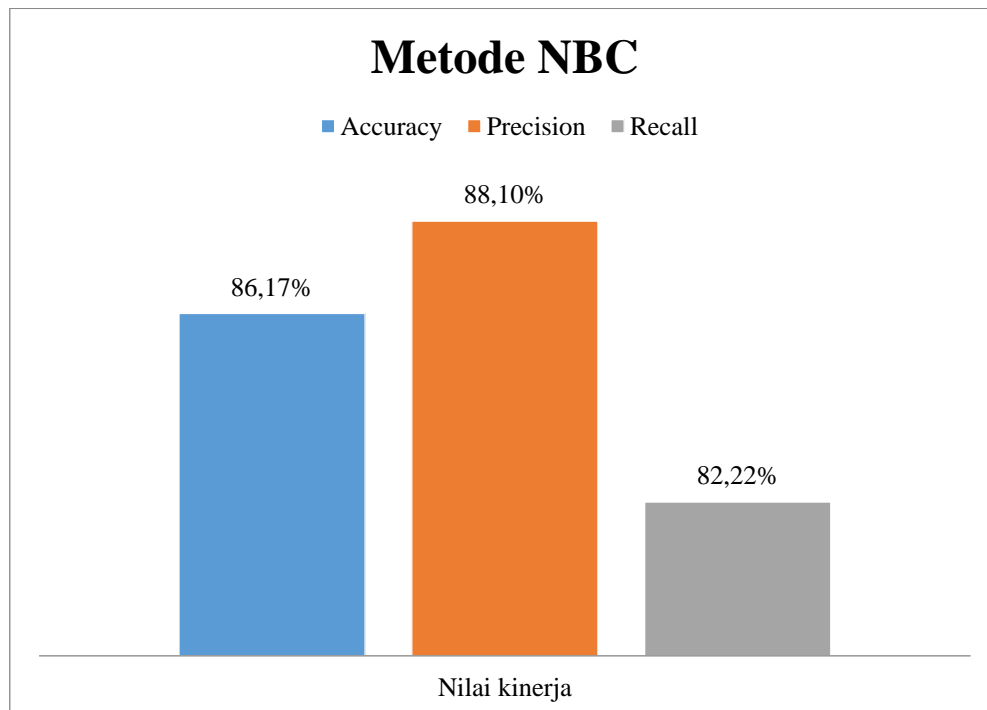
accuracy: 85.11%

	true Presence	true Absence	class precision
pred. Presence	34	6	85.00%
pred. Absence	8	46	85.19%
class recall	80.95%	88.46%	

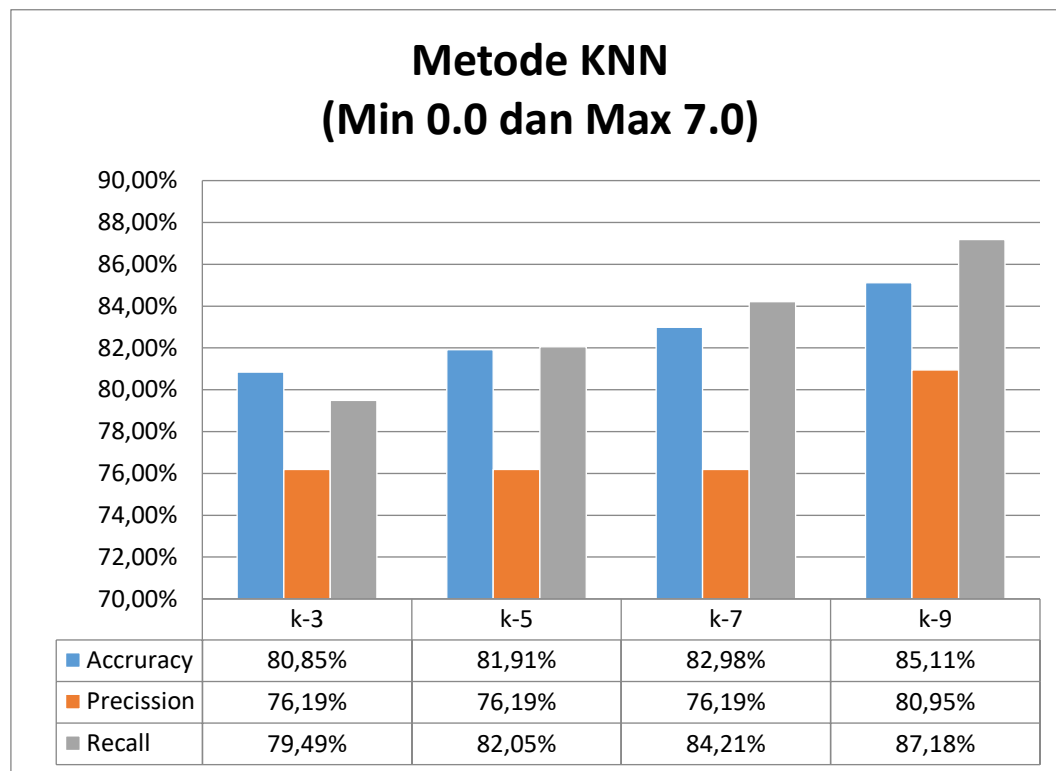
Gambar 8. Hasil Kinerja Menggunakan Algoritma K-NN

#### 3.6.4. Using Discovery Knowledge

Pada tahap terakhir proses *data mining* adalah menggunakan pengetahuan yang telah ditemukan berdasarkan setiap tahapan proses yang dilalui. Pengetahuan tersebut akan digunakan sebagai ilmu bagi orang lain, oleh karena itu sebisanya harus dikemas dengan menggunakan cara yang mudah dipahami agar apa yang didapat dari penelitian ini bisa dipelajari oleh orang lain. Salah satu cara agar data yang didapat bisa dimengerti adalah dengan menyajikannya dalam bentuk grafik. Grafik yang pertama berisi tentang data *accuracy*, *precision* dan *recall* dari algoritma NBC dan algoritma K-NN yang menggunakan nilai  $k=9$ ,  $min = 0$  dan nilai  $max = 8$ . Gambar 9 adalah grafik yang menyajikan nilai *accuracy*, *precision* dan *recall* dari algoritma NBC dan Gambar 10 adalah grafik yang menyajikan nilai *accuracy*, *precision* dan *recall* dari algoritma K-NN.

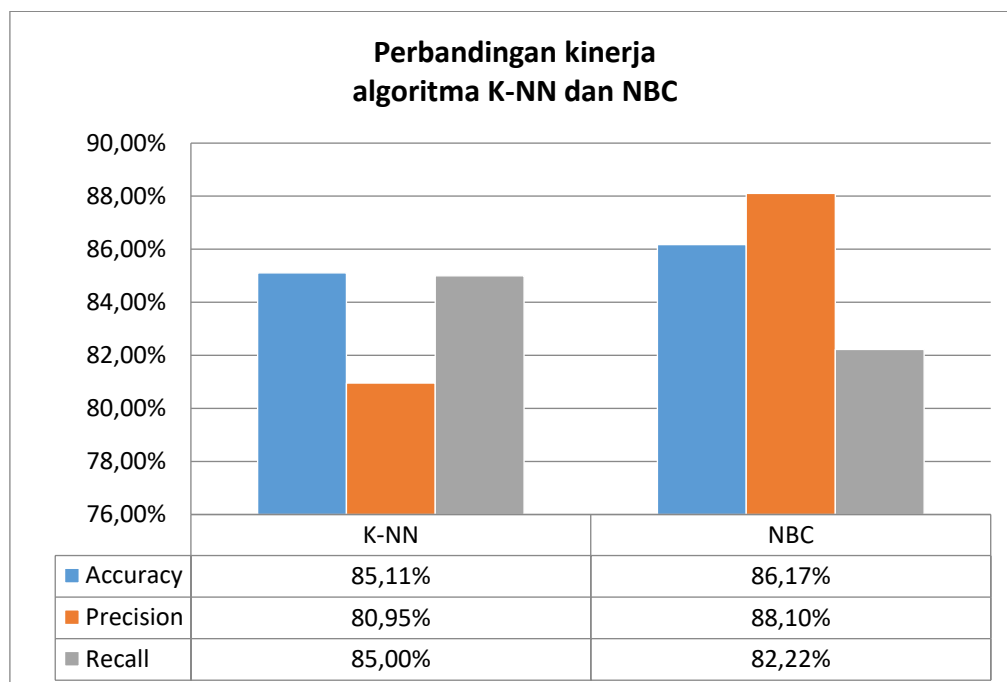


Gambar 9. Hasil Kinerja Metode NBC



Gambar 10. Hasil Kinerja Metode K-NN

Dari Hasil Kinerja metode NBC dan K-NN dapat dibandingkan Kinerja metode NBC dan K-NN seperti yang ditunjukkan oleh Gambar 11.



Gambar 11. Perbandingan Nilai Accuracy, Precision, dan Recall

Tabel 6. Perbandingan kinerja Pada Algoritma K-NN dan NBC

Algoritma	Accuracy	Precision	Recall
K-NN	85,11%	80,95%	85,00%
NBC	86,17%	88,10%	82,22%

Berdasarkan data pada grafik dan Tabel 6 diatas dapat dilihat algoritma NBC memiliki tingkat akurasi yang lebih tinggi yaitu 86,17%, sedangkan untuk algoritma K-NN memiliki tingkat akurasi sedikit lebih rendah yaitu 85,11%. Sehingga dapat disimpulkan bahwa algoritma NBC lebih baik dalam menangani kasus untuk mengklasifikasikan penyakit jantung.

#### 4. KESIMPULAN

##### 4.1. Kesimpulan

- Dengan mengimplementasikan algoritma NBC dan K-NN menggunakan nilai  $K = 3, 5, 7$  dan  $9$ ,  $range = 0 - 8$ , dapat menjadi salah satu solusi dari permasalahan yang ada yaitu untuk mengetahui apakah seseorang dapat diklasifikasikan berpotensi mengidap penyakit jantung atau tidak.
- Dengan menggunakan  $176$  data training dan  $94$  data testing, algoritma NBC memiliki tingkat akurasi lebih tinggi yaitu 86,17% sedangkan algoritma K-NN dengan  $K$  terbaik adalah  $9$ ,  $range$   $min = 0$  dan  $max = 8$  memiliki akurasi yang sedikit lebih rendah yaitu 85,11%.
- Jadi disimpulkan bahwa algoritma NBC adalah algoritma yang lebih baik untuk diimplementasikan dalam kasus klasifikasi menggunakan dataset heart disease.

##### 4.2. Saran

Saran untuk rencana penelitian yang akan datang dapat dicoba untuk menerapkan proses cleaning/pembersihan data dan pre-processing pada saat melakukan klasifikasi data, agar dapat kemungkinan untuk meningkatkan tingkat akurasi dari kedua metode menjadi hasil yang lebih baik lagi. Karena pada penelitian ini tidak menerapkan proses tersebut pada saat melakukan klasifikasi dataset.

#### DAFTAR PUSTAKA

- [1] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," J. Media Inform. Budidarma, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [2] A. Novandya, "Penerapan Algoritma Klasifikasi Data Mining C4.5 pada Dataset Cuaca Wilayah Bekasi," KNiST, pp. 368–372, 2017.
- [3] R. R. Waliyansyah and C. Fitriyah, "Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN)," J. Edukasi dan Penelit. Inform., vol. 5, no. 2, p. 157, 2019, doi: 10.26418/jp.v5i2.32473.
- [4] M. A. Maricar and Dian Pramana, "Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," J. Sist. dan Inform., vol. 14, no. 1, pp. 16–22, 2019, doi: 10.30864/jsi.v14i1.233.
- [5] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [6] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, "Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung," J. Pilar Nusa Mandiri, vol. 15, no. 2, pp. 189–194, 2019, doi: 10.33480/pilar.v15i2.669.

- [7] M. Rifal and N. Sinaga, “Kaji Eksperimental Rasio Metanol-Bensin Terhadap Konsumsi Bahan Bakar, Emisi Gas Buang, Torsi Dan Daya,” *Gorontalo J. Infrastruct. Sci. Eng.*, vol. 1, no. 1, p. 47, 2018, doi: 10.32662/gojise.v1i1.140.
- [8] Z. A. Hasibuan, “Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi,” *Konsep, Tek. Dan Apl.*, no. Universitas Indonesia, p. 194, 2007.
- [9] L. Wijaya and N. A. Pratiwi, “Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur Pendahuluan berbagai manusia bidang dan terutama bidang dengan jenis dan jumlah yang mencukupi sehingga obat dapat diperoleh dengan ce,” vol. 3, no. 2, pp. 64–73, 2020.
- [10] J. Miharja, P. S. Informatika, F. Teknik, and U. Majalengka, “Proceeding SENDIU 2021 PENERAPAN DATA MINING PENERIMAAN KARYAWAN MENGGUNAKAN METODE,” pp. 978–979, 2021.
- [11] T. Barat, J. J. Selatan, and C. Matrix, “Analisa Particle Swarm Optimization Terhadap Kepuasan,” pp. 143–148, 2020..