

Rancang Bangun Sistem Informasi Prediksi Penyakit Jantung Berbasis Algoritma Naive Bayes

Agung Suko Wijoyo^{*1}, Arief Jananto²

^{1,2}Unisbank Semarang; Jalan Tri Lomba Juang Semarang

e-mail: ^{*1}agungwijoyo8@gmail.com, ²ajananto09@edu.unisbank.ac.id

Abstrak

Penyakit jantung menyebabkan kematian di dunia/tahun. Kematian yang tinggi akibat penyakit jantung dapat dicegah dan faktor risiko berkurang jika masyarakat memiliki informasi tentang gejala penyakit jantung. Banyaknya faktor yang dikumpulkan untuk menentukan seseorang menderita penyakit kardiovaskuler atau tidak memerlukan sistem pengolahan data yang cukup besar sehingga jantung aplikasi data mining dengan metode *Naive Bayes Classifier* (NBC). Berdasarkan pengujian dari pembagian data latih 80 %, 85 % dan 90 % didapatkan akurasi terbaik menggunakan metode yaitu pada latih 90 % dengan akurasi sebesar 84 %. Semakin banyak data latih yang digunakan maka akan semakin baik pula akurasi dari metode NBC yang dihasilkan.

Kata kunci— Jantung, Prediksi, Naive Bayes Classifier

Abstract

Heart disease causes death in the world / year. High mortality from heart disease can be prevented and risk factors reduced if people have information about symptoms of heart disease. The number of factors collected to determine whether a person has cardiovascular disease or does not require a large enough data processing system is the heart of the data mining application with the Naive Bayes Classifier (NBC) method. Based on testing of the distribution of training data 80%, 85% and 90% the best accuracy was obtained using the method, namely at 90% training with an accuracy of 84%. The more training data used, the better the accuracy of the resulting NBC method.

Keywords— Heart, Naive Bayes Classifier, Prediction

1. PENDAHULUAN

Penyakit jantung menyebabkan kematian di dunia tiap tahun [1]. Pada tahun 2008, di dunia diperkirakan sebanyak 17,3 juta kematian disebabkan oleh penyakit kardiovaskuler. Lebih dari 3 juta kematian tersebut terjadi sebelum usia 60 tahun [2]. Terjadinya kematian dini yang disebabkan oleh penyakit jantung berkisar sebesar 4% di negara berpenghasilan tinggi, dan 42% terjadi di negara berpenghasilan rendah [3], [4]. Pada tahun 2030, kematian akibat penyakit jantung diproyeksikan akan terus meningkat mencapai 23,3 juta kematian [5]. Penyakit jantung terus meningkat di Indonesia, menyebabkan morbiditas, kecacatan, dan beban sosial ekonomi bagi keluarga, masyarakat, dan negara yang terkena dampak [6].

Penyakit jantung adalah penyakit karena disfungsi jantung dan pembuluh darah. Penyebab utama jantung adalah merokok, kurang olahraga, tidak sehatnya pola makan, dan minum alkohol [7]. Faktor yang tidak dapat dimodifikasi dari jantung antara lain riwayat keluarga, usia, jenis kelamin dan obesitas, sedangkan faktor yang dapat dimodifikasi antara lain tekanan darah tinggi, diabetes, disiplin, kurang aktivitas fisik, pola makan yang tidak sehat dan stres [8].

Setiap orang berisiko terkena penyakit jantung. Kematian yang tinggi akibat penyakit jantung dapat dicegah dan faktor risiko berkurang jika masyarakat memiliki informasi tentang gejala penyakit jantung. Berapa banyak faktor yang dikumpulkan untuk menentukan apakah seseorang terkena penyakit kardiovaskuler atau tidak membutuhkan sistem pengolahan data yang besar. Oleh karena itu, kebutuhan pengumpulan data merupakan langkah strategis [9]. Penambahan data adalah disiplin yang mempelajari metode untuk mengekstraksi informasi atau menemukan pola dalam data dalam jumlah besar. Data mining terbukti menjadi langkah terbaik untuk mengolah data dalam jumlah besar dan cepat menghasilkan data dalam jumlah besar dengan akurasi tinggi serta mengolah data dari berbagai sumber [10].

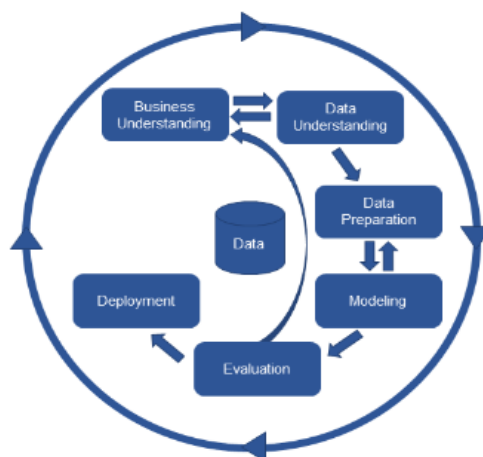
Metode data mining dalam penelitian ini adalah *Naive Bayes Classifier* (NBC). Metode NBC merupakan metode klasifikasi, meskipun juga dapat digunakan untuk estimasi dan prediksi. Metode

NBC memiliki kelebihan yang dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan dalam data besar [11], [12].

Penelitian sebelumnya terkait dengan prediksi menggunakan algoritma NBC yang diterapkan pada data kesehatan memiliki akurasi yang sangat baik. Penelitian oleh Devita dkk (2018) membandingkan metode NBC dengan KNN dengan hasil NBC lebih baik dari metode KNN [13]. Penelitian selanjutnya akan menggunakan metode NBC untuk melakukan prediksi penyakit jantung + atau - berdasarkan data dengan pembandingan data latih 80, 85 dan 90 %.

2. METODE PENELITIAN

Metode analisis data dalam penelitian ini menggunakan metode CRISP-sepeti gambar 3.1



Gambar 1. CRISP-DM

- a. **Pemahaman Bisnis (*Business Understanding*)**
Tahap ini memahami proses prediksi penyakit jantung yang diharapkan mampu melakukan penanganan terbaik terhadap penyakit jantung. Hal tersebut untuk mengurangi tingkat penyakit jantung yang dapat mengakibatkan kematian
- b. **Pemahaman Data (*Data Understanding*)**
Dataset penelitian diperoleh dari Kaggle Data Set yang terdiri dari 11 atribut dan 1 label dengan tipe data integer dan *binary*. Total 1.000 data penyakit jantung, dinyatakan 523 orang tidak terkena penyakit jantung dan 477 orang lainnya terkena penyakit jantung. Penjelasan mengenai atribut-atribut yang terdapat dalam dataset penyakit jantung dijabarkan sebagai tabel 1.

Tabel 1. Atribut Penyakit Jantung

No	Atribut	Tipe Data	Keterangan
1.	<i>Age</i>	Integer	Usia yang dinyatakan dalam tahun
2.	<i>Gender</i>	Binary	Jenis kelamin (1 : wanita / 2 : pria.
3.	<i>Body Mass Index</i>	Integer	Status gizi dihitung dari rumus $\frac{\text{berat badan (kg)}}{\text{tinggi badan}^2(m)}$
4.	<i>Systolic</i>	Integer	Tekanan darah atas (sistolik) dalam satuan mmHg.
5.	<i>Diastolic</i>	Integer	Tekanan darah bawah (diastolik) dalam satuan mmHg.
6.	<i>Cholesterol</i>	Integer	Kadar kolesterol dimana jika normal diberi nilai 1, diatas normal diberi nilai 2 sedangkan jika jauh diatas normal diberi nilai 3
7.	<i>Glucose</i>	Integer	Kadar glukosa dimana jika normal diberi nilai 1, diatas normal diberi nilai 2 sedangkan jika jauh diatas normal diberi nilai 3
8	<i>Smoking</i>	Binary	Kondisi merokok atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
9	<i>Alcohol Intake</i>	Binary	Konsumsi alkohol atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
10	<i>Physical Activity</i>	Binary	Melakukan olahraga atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
11	<i>Result</i>	Binary	Terkena penyakit jantung atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1

c. Pengolahan Data (*Data Preparation*)

Pada tahap ini dilakukan konversi dataset penyakit jantung yang berbentuk bilangan integer dan binominal akan di konversi menjadi bentuk permisalan yaitu

a) *Age*

Atribut *age* dibagi dalam 2 kategori yaitu:

Tabel 2. Atribut *Age*

Kondisi	Konversi
< 45	Muda
>45	Tua

b) *Gender*

Atribut *gender* dibagi dalam 2 kategori yaitu:

Tabel 3. Atribut *Gender*

Kondisi	Konversi
1	Wanita
2	Pria

c) *Body Mass Index (BMI)*

Atribut BMI dibagi dalam 3 kategori yaitu:

Tabel 4. Atribut BMI

Kondisi	Konversi
<18,5	Kurus
18,5-25	Normal
> 25	Gemuk

d) *Systolic*

Atribut *systolic* dibagi dalam 2 kategori yaitu:

Tabel 5. Atribut *Systolic*

Kondisi	Konversi
<=120	Normal
> 120	Tinggi

e) *Diastolic*

Atribut *diastolic* dibagi dalam 2 kategori yaitu (Wulandhari, 2019):

Tabel 6. Atribut *Diastolic*

Kondisi	Konversi
<=80	Normal
> 80	Tinggi

f) *Cholesterol*

Atribut *cholesterol* dibagi dalam 3 kategori yaitu:

Tabel 7. Atribut *Cholesterol*

Kondisi	Konversi
1	Normal
2	Diatas Normal
3	Jauh Diatas Normal

g) *Glucose*

Atribut *glucose* dibagi dalam 3 kategori yaitu:

Tabel 8. Atribut *Glucose*

Kondisi	Konversi
1	Normal
2	Diatas Normal
3	Jauh Diatas Normal

h) *Smoking*

Atribut *smoking* dibagi dalam 2 kategori yaitu:

Tabel 9. Atribut *Smoking*

Kondisi	Konversi
0	Tidak
1	Ya

i) *Alcohol Intake*

Atribut *alcohol intake* dibagi dalam 2 kategori yaitu:

Tabel 10. Atribut *Alcohol Intake*

Kondisi	Konversi
0	Tidak
1	Ya

j) *Physical Activity*

Atribut *physical activity* dibagi dalam 2 kategori yaitu:

Tabel 11. Atribut *Physical Activity*

Kondisi	Konversi
0	Tidak
1	Ya

k) *Result*

Atribut *result* dibagi dalam 2 kategori yaitu:

Tabel 12. Atribut *Result*

Kondisi	Konversi
0	Tidak
1	Ya

d. Pemodelan (*Modelling*)

Tahap ini dilakukan prediksi dari data uji dari penyakit jantung dengan metode NBC.

e. Evaluasi (*Evaluation*)

Tahap ini model yang terbentuk diuji menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi dengan menggunakan rumus akurasi = $\frac{TP+TN}{TP+TN+FP+FN} \times 100\%$...

3. HASIL DAN PEMBAHASAN

3.1. Data Latih 80%

Penggunaan data latih 80 % dari 1.000 data akan menghasilkan data latih dari nomor 1 sampai dengan 800 sedangkan data uji akan dimulai dari nomor 801 sampai dengan nomor 1.000. Dari 800 data latih penyakit jantung diketahui sebagai berikut:

- Jumlah Data Latih : 800
- Kasus Positif : 387
- Kasus Negatif : 413

Data uji yang dihitung adalah data nomor 801 dengan kriteria sebagai berikut:

Age : Tua
 Gender : Wanita
 BMI : Gemuk
 Systolic : Tinggi
 Diastolic : Normal
 Cholesterol : Jauh Diatas Normal
 Glucose : Normal
 Smoking : Tidak
 Alcohol Intake : Tidak
 Physical Activity : Ya

Proses perhitungan data uji dengan algoritma NBC untuk probabilitas kasus positif yaitu:

$$\begin{aligned}
 P(\text{Status} = \text{Positif}) &= \frac{387}{1.000} \\
 P(\text{Age} = \text{Tua} \mid \text{Hasil} = \text{Positif}) &= \frac{353}{387} \\
 P(\text{Gender} = \text{Wanita} \mid \text{Hasil} = \text{Positif}) &= \frac{242}{387} \\
 P(\text{BMI} = \text{Gemuk} \mid \text{Hasil} = \text{Positif}) &= \frac{271}{387} \\
 P(\text{Sys} = \text{Tinggi} \mid \text{Hasil} = \text{Positif}) &= \frac{242}{387} \\
 P(\text{Dias} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{240}{387} \\
 P(\text{Chol} = \text{JDA} \mid \text{Hasil} = \text{Positif}) &= \frac{83}{387} \\
 P(\text{Glu} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{297}{387} \\
 P(\text{Smoke} = \text{Tidak} \mid \text{Hasil} = \text{Positif}) &= \frac{351}{387} \\
 P(\text{Alco} = \text{Tidak} \mid \text{Hasil} = \text{Positif}) &= \frac{367}{387} \\
 P(\text{Active} = \text{Ya} \mid \text{Hasil} = \text{Positif}) &= \frac{311}{387} \\
 P(X \mid \text{Hasil} = \text{Positif}) &= \frac{387}{1.000} \times \frac{353}{387} \times \frac{242}{387} \times \frac{271}{387} \times \frac{242}{387} \times \frac{240}{387} \times \frac{83}{387} \times \frac{297}{387} \times \frac{351}{387} \times \frac{367}{387} \times \frac{311}{387} \\
 &= 0,0072457665075193
 \end{aligned}$$

Proses perhitungan data uji dengan algoritma NBC untuk probabilitas kasus negatif yaitu:

$$\begin{aligned}
 P(\text{Status} = \text{Negatif}) &= \frac{413}{1.000} \\
 P(\text{Age} = \text{Tua} \mid \text{Hasil} = \text{Negatif}) &= \frac{315}{413} \\
 P(\text{Gender} = \text{Wanita} \mid \text{Hasil} = \text{Negatif}) &= \frac{256}{413} \\
 P(\text{BMI} = \text{Gemuk} \mid \text{Hasil} = \text{Negatif}) &= \frac{222}{413} \\
 P(\text{Sys} = \text{Tinggi} \mid \text{Hasil} = \text{Negatif}) &= \frac{72}{413} \\
 P(\text{Dias} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{362}{413} \\
 P(\text{Chol} = \text{JDA} \mid \text{Hasil} = \text{Negatif}) &= \frac{27}{413} \\
 P(\text{Glu} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{365}{413} \\
 P(\text{Smoke} = \text{Tidak} \mid \text{Hasil} = \text{Negatif}) &= \frac{379}{413} \\
 P(\text{Alco} = \text{Tidak} \mid \text{Hasil} = \text{Negatif}) &= \frac{384}{413} \\
 P(\text{Active} = \text{Ya} \mid \text{Hasil} = \text{Negatif}) &= \frac{319}{413} \\
 P(X \mid \text{Hasil} = \text{Negatif}) &= \frac{413}{1.000} \times \frac{315}{413} \times \frac{256}{413} \times \frac{222}{413} \times \frac{72}{413} \times \frac{362}{413} \times \frac{27}{413} \times \frac{365}{413} \times \frac{379}{413} \times \frac{384}{413} \times \frac{319}{413} \\
 &= 0,00076334673897975
 \end{aligned}$$

Hasil prediksi penyakit jantung adalah positif, karena nilai *posterior probability* dari penyakit jantung dengan hasil positif lebih tinggi dibanding dengan nilai *posterior probability* dari penyakit jantung dengan dengan hasil negatif. Hasil perhitungan dengan data latih 80 % yang lain menyesuaikan dan hasil prediksi penyakit jantung dengan algoritma *Naïve Bayes Classifier* yang diperlihatkan seperti tabel 14

Tabel 14 Hasil Prediksi Data Latih 80%

No.	Hasil	Prediksi	No.	Hasil	Prediksi
801	-	+	976	+	-
802	-	-	977	-	-
803	-	-	978	+	+
804	+	-	979	+	+
805	+	+	980	-	-
806	+	-	981	-	-
807	+	+	982	-	+
808	-	-	983	-	-
809	-	-	984	+	+
810	+	+	985	-	-
811	-	-	986	+	+
812	+	-	987	+	-
813	+	+	988	-	-
814	+	-	989	+	+
815	+	+	990	-	+
816	-	-	991	-	-
817	+	+	992	+	-
818	+	-	993	+	-
819	-	+	994	-	+
820	+	+	995	+	+
821	-	-	996	+	-
822	+	+	997	-	-
823	+	+	998	-	-
824	+	+	999	+	-
825	-	-	1000	+	-

Dari perhitungan dengan menggunakan seluruh data uji, selanjutnya dilakukan pengujian menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh dari data latih 80 % algoritma NBC yang ditunjukkan seperti tabel 15.

Tabel15. Hasil *Confusion Matrix* Data Latih 80 %

	True Positif	True Negatif
Pred. Positif	58	22
Pred. Negatif	28	92

Hasil dari tabel *confusion matrix* 15 didapatkan nilai akurasi sebesar 75 % dengan perhitungan sebagai berikut $\text{akurasi} = \frac{58+92}{58+22+28+92} \times 100 \% = 75 \%$. Hasil prediksi dan akurasi pada sistem prediksi penyakit jantung dengan data latih 80 %

3.2. Data Latih 85%

Penggunaan data latih 85 % dari 1.000 data akan menghasilkan data latih dari nomor 1 sampai dengan 850 sedangkan data uji akan dimulai dari nomor 851 sampai dengan nomor 1.000. Dari 850 data latih penyakit jantung diketahui sebagai berikut:

- Jumlah Data Latih : 850
- Kasus Positif : 409
- Kasus Negatif : 441

Hasil perhitungan dengan data latih 85 % dengan algoritma NBC diperlihatkan seperti tabel 16.

Tabel 16. Hasil Prediksi Data Latih 85 %

No.	Hasil	Prediksi	No.	Hasil	Prediksi
851	-	-	976	+	-
852	-	+	977	-	-
853	-	+	978	+	+
854	+	-	979	+	+
855	-	-	980	-	-
856	+	-	981	-	-
857	-	-	982	-	+
858	-	-	983	-	-
859	-	-	984	+	+
860	+	-	985	-	-
861	+	+	986	+	+
862	+	-	987	+	-
863	-	+	988	-	-
864	-	-	989	+	+
865	+	+	990	-	+
866	-	-	991	-	-
867	+	+	992	+	-
868	-	+	993	+	-
869	-	-	994	-	+
870	-	-	995	+	+
871	-	+	996	+	-
872	+	+	997	-	-
873	-	-	998	-	-
874	-	-	999	+	-
875	-	-	1000	+	-

Dari perhitungan dengan menggunakan seluruh data uji, selanjutnya dilakukan pengujian menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh dari data latih 85 % algoritma *Naïve Bayes Classifier* yang ditunjukkan seperti tabel 17.

Tabel 17. Hasil *Confusion Matrix* Data Latih 85 %

	True Positif	True Negatif
Pred. Positif	44	15
Pred. Negatif	20	71

Hasil dari tabel *confusion matrix* 17 didapatkan nilai akurasi sebesar 77 % dengan perhitungan yaitu $\text{akurasi} = \frac{44+71}{44+15+20+71} \times 100\% = 77\%$. Hasil prediksi dan akurasi pada sistem prediksi penyakit jantung dengan data latih 85 %

3.3. Data Latih 90 %

Penggunaan data latih 90 % dari 1.000 data akan menghasilkan data latih dari nomor 1 sampai dengan 900 sedangkan data uji akan dimulai dari nomor 901 sampai dengan nomor 1.000. Dari 900 data latih penyakit jantung diketahui sebagai berikut:

- Jumlah Data Latih : 900
- Kasus Positif : 431
- Kasus Negatif : 469

Hasil perhitungan dengan data latih 90 % pada prediksi penyakit jantung dengan algoritma NBC diperlihatkan seperti tabel 18.

Tabel 18. Hasil Prediksi Data Latih 90 %

No.	Hasil	Prediksi	No.	Hasil	Prediksi
901	-	-	976	+	-
902	-	-	977	-	-
903	-	-	978	+	+
904	-	-	979	+	+
905	+	+	980	-	-
906	-	-	981	-	-
907	-	-	982	-	+
908	-	-	983	-	-
909	-	-	984	+	+
910	-	-	985	-	-
911	-	-	986	+	+
912	+	+	987	+	-
913	+	+	988	-	-
914	-	-	989	+	+
915	+	+	990	-	+
916	+	+	991	-	-
917	+	+	992	+	-
918	+	+	993	+	-
919	-	-	994	-	+
920	+	+	995	+	+
921	-	-	996	+	-
922	-	-	997	-	-
923	-	-	998	-	-
924	-	-	999	+	-
925	+	+	1000	+	-

Dari perhitungan dengan menggunakan seluruh data uji, selanjutnya dilakukan pengujian menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh dari data latih 90 % algoritma *Naïve Bayes Classifier* yang ditunjukkan seperti tabel 19.

Tabel 19. Hasil *Confusion Matrix* Data Latih 90 %

	True Positif	True Negatif
Pred. Positif	32	6
Pred. Negatif	10	52

Hasil dari tabel *confusion matrix* 19 didapatkan nilai akurasi sebesar 84 % dengan perhitungan yaitu akurasi = $\frac{32+52}{32+6+10+52} \times 100\% = 84\%$. Hasil prediksi dan akurasi pada sistem prediksi penyakit jantung dengan data latih 90 %.

Berdasarkan pengujian dari pembagian data latih 80 %, 85 % dan 90 % didapatkan akurasi terbaik menggunakan metode NBC yaitu pada latih 90 % dengan akurasi sebesar 84 %.

3.4. Pengujian

Proses pengujian menggunakan kriteria sebagai berikut:

Age : <45 Tahun (Muda)
 Gender : Pria
 BMI : > 25 (Gemuk)

<i>Systolic</i>	:	<=120 (Normal)
<i>Diastolic</i>	:	<=80 (Normal)
<i>Cholesterol</i>	:	Normal
<i>Glucose</i>	:	Normal
<i>Smoking</i>	:	Tidak
<i>Alcohol Intake</i>	:	Tidak
<i>Physical Activity</i>	:	Ya

Proses perhitungan data uji dengan algoritma NBC untuk probabilitas kasus positif sebagai berikut:

$$\begin{aligned}
 P(\text{Status} = \text{Positif}) &= \frac{431}{1.000} \\
 P(\text{Age} = \text{Muda} \mid \text{Hasil} = \text{Positif}) &= \frac{37}{431} \\
 P(\text{Gender} = \text{Pria} \mid \text{Hasil} = \text{Positif}) &= \frac{155}{431} \\
 P(\text{BMI} = \text{Gemuk} \mid \text{Hasil} = \text{Positif}) &= \frac{299}{431} \\
 P(\text{Sys} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{165}{431} \\
 P(\text{Dias} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{227}{431} \\
 P(\text{Chol} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{271}{431} \\
 P(\text{Glu} = \text{Normal} \mid \text{Hasil} = \text{Positif}) &= \frac{336}{431} \\
 P(\text{Smoke} = \text{Tidak} \mid \text{Hasil} = \text{Positif}) &= \frac{394}{431} \\
 P(\text{Alco} = \text{Tidak} \mid \text{Hasil} = \text{Positif}) &= \frac{409}{431} \\
 P(\text{Active} = \text{Ya} \mid \text{Hasil} = \text{Positif}) &= \frac{345}{431} \\
 P(X \mid \text{Hasil} = \text{Positif}) &= \frac{431}{1.000} \times \frac{37}{431} \times \frac{155}{431} \times \frac{299}{431} \times \frac{165}{431} \times \frac{227}{431} \times \frac{271}{431} \times \frac{336}{431} \times \frac{394}{431} \times \frac{409}{431} \times \frac{345}{431} \\
 &= 0,00030840667674065
 \end{aligned}$$

Proses perhitungan data uji dengan algoritma NBC untuk probabilitas kasus negatif sebagai berikut:

$$\begin{aligned}
 P(\text{Status} = \text{Negatif}) &= \frac{469}{1.000} \\
 P(\text{Age} = \text{Muda} \mid \text{Hasil} = \text{Negatif}) &= \frac{106}{469} \\
 P(\text{Gender} = \text{Pria} \mid \text{Hasil} = \text{Negatif}) &= \frac{177}{469} \\
 P(\text{BMI} = \text{Gemuk} \mid \text{Hasil} = \text{Negatif}) &= \frac{252}{469} \\
 P(\text{Sys} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{384}{469} \\
 P(\text{Dias} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{411}{469} \\
 P(\text{Chol} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{389}{469} \\
 P(\text{Glu} = \text{Normal} \mid \text{Hasil} = \text{Negatif}) &= \frac{415}{469} \\
 P(\text{Smoke} = \text{Tidak} \mid \text{Hasil} = \text{Negatif}) &= \frac{429}{469} \\
 P(\text{Alco} = \text{Tidak} \mid \text{Hasil} = \text{Negatif}) &= \frac{436}{469} \\
 P(\text{Active} = \text{Ya} \mid \text{Hasil} = \text{Negatif}) &= \frac{366}{469} \\
 P(X \mid \text{Hasil} = \text{Negatif}) &= \frac{469}{1.000} \times \frac{106}{469} \times \frac{177}{469} \times \frac{252}{469} \times \frac{384}{469} \times \frac{411}{469} \times \frac{389}{469} \times \frac{415}{469} \times \frac{429}{469} \times \frac{436}{469} \times \frac{366}{469} \\
 &= 0,0069880998881851
 \end{aligned}$$

Hasil prediksi penyakit jantung adalah negatif atau tidak terkena penyakit jantung, karena nilai *posterior probability* dari penyakit jantung dengan hasil negatif lebih tinggi dibanding dengan nilai *posterior probability* dari penyakit jantung dengan dengan hasil positif.

4. KESIMPULAN

Berdasarkan pengujian dari pembagian data latih 80 %, 85 % dan 90 % didapatkan akurasi terbaik

menggunakan metode *Naive Bayes Classifier* yaitu pada latih 90 % dengan akurasi sebesar 84 %. Semakin banyak data latih yang digunakan maka akan semakin baik pula akurasi dari metode *Naive Bayes Classifier* yang dihasilkan:

DAFTAR PUSTAKA

- [1] R. R. Juslim and F. Herawati, Penyakit Kardiovaskular, Yogyakarta: Graha Ilmu, 2018.
- [2] R. M. Benjamin, "The Million Heart Initiative: Progress In Preventing Heart Attacks And Stroke," *Surgeon General's Perspectives*, pp. 558-560, 2012.
- [3] S. Fadlilah, A. Sucipto and T. Amestiasih, "Usia, Jenis Kelamin, Perilaku Merokok dan IMT Berhubungan Dengan Resiko Penyakit Kardiovaskular," *Jurnal Keperawatan*, vol. XII, no. 4, pp. 261-268, 2019.
- [4] A. Nugroho, A. B. Gumelar, A. G. Sooi, D. Sarvasti and P. L. Tahalele, "Perbandingan Performansi Algoritma Pengklasifikasian Terpandu Untuk Kasus Penyakit Kardiovaskular," *RESTI*, vol. IV, no. 5, pp. 998-1006, 2019.
- [5] J. W. Rossano, "Congenital heart disease: a global public health concern," *Health Metrics and Evaluation*, vol. IV, pp. 168-169, 2020.
- [6] A. Sakti, I. Alwi, Muhadi and H. Shatri, "Karakteristik Mortalitas Jemaah Haji Indonesia Akibat Penyakit Kardiovaskular," *Jurnal Penyakit Dalam Indonesia*, vol. VI, no. 4, pp. 178-181, 2019.
- [7] D. Zhao, "Epidemilogy Features of Cardiovascular Disease in Asia," *JACC*, vol. I, no. 1, pp. 1-13, 2021.
- [8] N. A. Arsani, N. S. Wahyuni and N. M. Agustini, "Deteksi Dini Dan Pencegahan Penyakit Kardiovaskular," *Proceeding Senadimas Undiksha*, pp. 663-668, 2022.
- [9] Q. Yang and X. Wu, "10 Challenging Problems in Data Mining Research," *International Journal of Information Technology & Decision Making*, vol. V, pp. 1-9, 2006.
- [10] F. A. Hermawati, Data Mining, Yogyakarta: Andi, 2016.
- [11] Kusriani and E. T. Luthfi, Algoritma Data Mining, Yogyakarta: Andi, 2015.
- [12] J. Han and M. Kamber, Data Mining: Concept and Techniques, Third Edition, Waltham: Morgan Kaufmann, 2016.
- [13] R. N. Devita, H. W. Herwanto and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. V, pp. 427-434, 2018.