

# SISTEM PREDIKSI PENYAKIT JANTUNG KORONER MENGUNAKAN METODE *NAÏVE BAYES*

Devina Larassati<sup>1)</sup>, Ati Zaidiah<sup>2)</sup>, Sarika Afrizal<sup>3)</sup>

<sup>1, 2, 3)</sup> Program Studi S-1 Sistem Informasi, Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jakarta  
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok,  
Daerah Khusus Ibukota Jakarta 12450

e-mail: [devinal@upnvj.ac.id](mailto:devinal@upnvj.ac.id)<sup>1)</sup>, [atizaidiah@upnvj.ac.id](mailto:atizaidiah@upnvj.ac.id)<sup>2)</sup>, [sarika.afrizal@upnvj.ac.id](mailto:sarika.afrizal@upnvj.ac.id)<sup>3)</sup>

## ABSTRAK

*Penyebab dari penyakit jantung koroner yaitu penyumbatan pembuluh darah koroner, penyakit ini sangat diperhatikan oleh seluruh kalangan masyarakat dikarenakan pengaruh yang disebabkan oleh penyakit tersebut. Penelitian ini memiliki tujuan untuk membuat prediksi yang akan membantu para dokter untuk melakukan diagnose dengan tepat dan akurat sehingga penyakit jantung koroner dapat ditangani lebih awal. Salah satu algoritma klasifikasi data mining yang digunakan pada penelitian ini adalah algoritma Naïve Bayes Classifier. Algoritma ini diterapkan dengan tujuan untuk menghitung probabilitas kemungkinan seseorang pasien berdasarkan data rekam medis pasien. Rekam medis pasien diperoleh dari kaggle untuk dilakukan percobaan pada sistem yang akan dibuat. Dataset awal memuat 303 record setelah dilakukan preprocessing memuat 296 record. Percobaan pada penelitian ini dilakukan sebanyak 3 percobaan dengan membagi data latih dan data uji. Hasil yang diperoleh dalam percobaan pertama memiliki akurasi paling tinggi yaitu sebesar 83.1%. Diharapkan dengan adanya sistem ini dapat membantu dokter untuk mendiagnosis penyakit jantung koroner.*

**Kata Kunci:** Penyakit Jantung Koroner, Prediksi, Klasifikasi, Naïve Bayes Classifier.

## ABSTRACT

*Coronary heart disease, caused by blockage of coronary arteries, is a disease that gets attention from all circles of society considering its impact. This research was conducted to make predictions that could later assist a doctor in determining the right diagnosis and early treatment for coronary heart disease. One of the data mining classification algorithms used in this study is the Naive Bayes Classifier algorithm. This algorithm is applied to calculate the probability of a patient based on the patient's medical record data. Patient medical records were obtained from Kaggle for experimentation on the system to be made. Aval dataset contains 303 records after preprocessing contains 296 records. In this study, three experiments were conducted by dividing training data and test data. The results obtained in the first experiment had the highest accuracy, namely of 83.1% It is hoped that this system can help doctors to diagnose: coronary heart disease.*

**Keywords:** Coronary Heart Disease, Prediction, Clarifikasi, Naive Bayes Classifier

## I. PENDAHULUAN

Jantung adalah adalah suatu organ tubuh dari manusia yang berperan dalam sistem peredaran darah. Ketika otot jantung tidak memiliki aliran darah yang cukup, maka sindrom koroner akut atau serangan jantung dapat mengganggu kerja jantung dengan serius, sehingga fungsi jantung untuk mengedarkan darah menuju seluruh tubuh menjadi terganggu [1]. Menurut penelitian Perhimpunan Dokter Spesialis Kardiovaskular Indonesia (PERKATI) tahun 2020, penyakit jantung koroner adalah pembunuh pertama di seluruh dunia, sekitar 36% dari semua kematian, dimana dua kali lebih tinggi dari tingkat kematian yang disebabkan oleh kanker. Menurut laporan, penyakit jantung koroner di Indonesia merupakan penyebab nomor satu berdasarkan semua kematian, yaitu sebanyak 26,4% [2].

Data mining merupakan suatu proses menemukan informasi baru dengan mencari pola dan kecenderungan dengan menggali informasi tersembunyi dalam sejumlah data yang sangat besar yang tersimpan menggunakan metode pengenalan pola seperti metode statistik dan matematika. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery Databases (KDD) [3].

Teknik data mining digunakan untuk menganalisis data berjumlah besar di dalam *database*. Dalam bidang kesehatan, potensi data mining telah diakui secara luas. Penelitian dengan menggunakan teknik data mining modern telah banyak dilakukan, termasuk *classification* dan *predictive* yang diaplikasikan dalam rekam medis elektronik. Dalam hal ini sudah dilakukan perbandingan terhadap ketiga metode klasifikasi. Untuk percobaan pada Algoritma

*Naïve Bayes* di dapatkan nilai akurasi paling tinggi sebesar 86.88%, pada Algoritma Decision Tree memiliki nilai akurasi sebesar 85.24% dan pada *K-Nearest Neighbors* didapatkan nilai akurasi paling rendah yaitu sebesar 72.13%. Dengan demikian, Algoritma *Naïve Bayes* dipilih pebeliti untuk mendiagnosa penyakit jantung koroner dikarenakan nilai akurasi yang diperoleh paling tinggi. Metode Naive Bayes memiliki kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana dan berakurasi tinggi [4].

Sistem adalah jaringan proses yang saling berhubungan yang terorganisir untuk melakukan aktivitas dan mencapai tujuan tertentu [5]. Perancangan sistem merupakan salah satu unsur atau tahapan dari keseluruhan pembangunan sistem komputerisasi. Perancangan sistem untuk pengembangan sistem informasi biasanya memerlukan jangka waktu yang lebih lama daripada pemecahan masalah pada umumnya memperlihatkan aliran data utama pada sistem [6].

Prediksi adalah sebuah proses memperkirakan kejadian pada masa depan dengan sistematis menurut informasi yang berasal dari masa lalu serta meminimalkan kesalahan (perbedaan antara hal yang sudah terjadi dan hal yang diharapkan). Prediksi bukan serta merta memberi hasil atau jawaban jelas, tetapi merupakan upaya untuk mencari dan memperoleh jawaban yang sedekat mungkin dengan apa yang mungkin terjadi [7].

Peramalan atau disebut juga *forecasting* ialah sebuah cara menganalisis perhitungan yang dilakukan menggunakan pendekatan kuantitatif atau kualitatif untuk memprediksi peristiwa di masa depan dengan merujuk pada referensi data di masa lampau [8].

Prediksi bisa berdasarkan metode ilmiah maupun subjektif belaka. Misalkan, prediksi cuaca, gempa, gunung meletus ataupun bencana secara umum, selalu berdasarkan data dan informasi terbaru yang didasarkan pengamatan termasuk oleh satelit. Namun, prediksi seperti pertandingan sepakbola, olahraga, dan lain-lainnya, umumnya berdasarkan pandangan subjektif dengan sudut pandang sendiri dari yang memprediksinya. [9]

*Python* adalah bahasa pemrograman yang ditafsirkan serbaguna. Berbeda dengan bahasa lain yang sangat sulit untuk dipahami dan dibaca, *python* menekankan pada keterbacaan kode untuk membuat perintah lebih mudah dipahami. Maka mempelajari *Python* merupakan hal yang sangat mudah bagi pemula dan siapa saja yang telah mempelajari bahasa pemrograman lain [10].

*Flask* merupakan kerangka kerja web yang ditulis dengan Python yang diklasifikasikan sebagai kerangka mikro. *Flask* bertindak sebagai *framework* aplikasi dan tampilan *Web*. *Flask* dan *Python* memungkinkan pengembang untuk membuat jaringan terstruktur dan mengelola perilaku jaringan dengan lebih mudah [11].

Ada persamaan dan perbedaan antara penelitian penyakit jantung koroner saat ini dan penelitian sebelumnya. Persamaannya adalah peneliti akan menggunakan metode pengklasifikasian menggunakan algoritma yang sama yakni Algoritma *Naïve Bayes* untuk memprediksi penyakit jantung koroner.

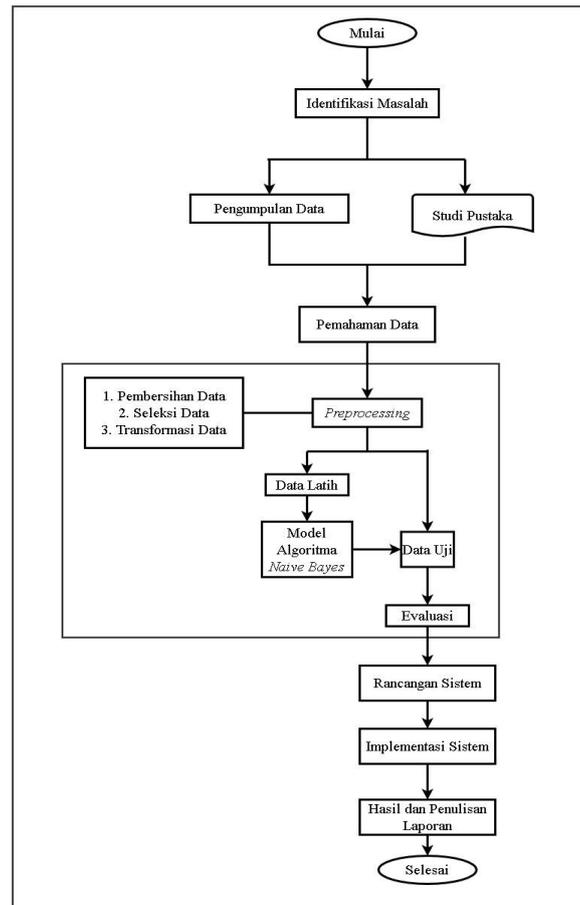
Perbedaannya, peneliti membuat sistem prediksi sederhana untuk membantu dokter dalam memprediksi penyakit jantung koroner berdasarkan gejala atau data pasien yang tersedia. Sistem yang dibangun ini menggunakan Python Flask dengan menarik dataset yang tersedia sebagai sara perhitungan untuk mendapatkan hasil prediksi. Langkah pertama peneliti menggunakan beberapa variabel yang telah ditentukan dan melakukan beberapa percobaan dengan proporsi *data training* dan *data testing* yang berbeda di setiap percobaan untuk melihat hasil akurasi tertinggi, pada langkah selanjutnya peneliti membandingkan hasil akurasi pertama dengan hasil akurasi lainnya sesuai dengan percobaan yang dilakukan.

Berdasarkan uraian tersebut, peneliti tertarik melakukan penelitian terhadap kumpulan data yang merupakan data sekunder bersumber dari website kaggle ([www.kaggle.com](http://www.kaggle.com)) yang diupload oleh Rashik Rahman pada tahun 2021 dengan judul *Heart Attack Analysis & Prediction Dataset* [12]. *Dataset* bisa digunakan untuk memberikan informasi mengenai keadaan seseorang apakah terkena penyakit jantung koroner atau tidak. Peneliti juga tertarik untuk merancang sistem sederhana yang dapat memprediksi penyakit jantung koroner dengan menerapkan Algoritma *Naïve Bayes*. Sistem tersebut diharapkan dapat digunakan untuk memprediksi penyakit jantung koroner berdasarkan gejala atau rekam medis pasien yang ada.

## II. METODE PENELITIAN

### A. Alur Penelitian

Berikut adalah alur penelitian yang dilakukan oleh peneliti seperti terlihat pada Gambar 1.



Gambar 1. Alur Penelitian

### B. Pengumpulan Data

Data di peroleh dari website kaggle ([www.kaggle.com](http://www.kaggle.com)) yang diupload oleh Rashik Rahman pada tahun 2021 dengan judul *Heart Attack Analysis & Prediction Dataset*. Setelah mendapatkan dataset peneliti melakukan validasi dataset dengan melakukan wawancara kepada pakar atau dokter spesialis jantung. Wawancara dilakukan dengan dokter dr. Bambang Dwiputra, SpJP.

### C. Pemahaman Data

Data utama pada penelitian berasal dari *dataset* [www.kaggle.com](http://www.kaggle.com). *Dataset* yang diperoleh berjumlah 303 record yang terdiri dari 13 variabel dan 1 kelas. Variabel yang diberikan berdasarkan [www.kaggle.com](http://www.kaggle.com) berupa Age (usia), Sex (jenis kelamin), Chest Pain Type/CP (jenis nyeri dada), Resting Blood Pressure/Trestbps (tekanan darah), Cholesterol (kolesterol), Fasting Blood Sugar/Fbs (gula darah), Resting Electrocardiographic/Restecg (hasil elektrokardiografi), Thalach (detak jantung maksimum), Exang (nyeri dada apabila berolahraga), Oldpeak (segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat), Slope (kemiringan segmen ST dalam latihan maksimum), Ca (jumlah vessel utama yang diwarnai oleh fluoroskopi), Thall (status jantung), dan Output (kelas dari fitur) [10].

### D. Data Processing

Pengolahan data yang dilakukan adalah membersihkan data dengan cara mengecek apakah data yang digunakan memiliki informasi (fitur) yang hilang atau tidak atau biasa yang disebut dengan missing value dan mereduksi data pada fitur yang tidak digunakan untuk memastikan data pada rekam medik pemeriksaan yang dipilih telah layak untuk proses pengolahan, kemudian data diproses menggunakan Algoritma Naïve Bayes.

#### 1. Data Cleaning (Pembersihan Data)

Tujuan dari tahapan ini yaitu guna melengkapi data yang hilang dengan mengidentifikasi outlier dan mengoreksi ketidak konsistenan data.

#### 2. Data Selection (Seleksi Data)

Pada tahapan ini data diolah, di mana data akan diseleksi serta dipisahkan antara data yang relevan atau

dibutuhkan dengan data yang tidak diperlukan.

### 3. *Data Transformation* (Transformasi Data)

Transformasi data yang dilakukan untuk mendeteksi keluaran dari prediksi seseorang dengan melihat apakah data tersebut sesuai atau tidak dengan algoritma yang telah dilakukan. Data yang berjenis numerikal seperti variabel usia. Pada *dataset* penyakit jantung koroner yang diperoleh, seluruh variabelnya menggunakan numerikal sehingga akan dilakukan perubahan data menjadi bentuk kategorikal.

### E. *Pemodelan Data*

Data pada tahapan ini telah melalui data preprocessing akan dilakukan implementasi Algoritma Naïve Bayes pada *dataset* untuk mengetahui hasil memprediksi penyakit jantung koroner.

### F. *Evaluasi*

Tahap ini merupakan tahap evaluasi dari hasil pemodelan yang telah diterapkan. Untuk mengetahui tingkat keakurasian model tersebut, maka perhitungan akurasi akan dilakukan dengan perumusan sebagai berikut:

$$Akurasi = \frac{\sum Data Benar}{n} \times 100\%$$

Keterangan :

Akurasi : Akurasi sistem

$\sum Data Benar$  : Jumlah benar dalam sampel uji

$n$  : Banyaknya jumlah sampel uji

### G. *Rancangan Sistem*

Tahap ini merupakan tahap merancang sistem yang sederhana, dengan adanya tahap ini dimaksudkan untuk merancang sebuah sistem yang sederhana menggunakan model yang telah dibangun sebelumnya. Pada tahap ini peneliti merancang sistem prediksi PJK dan menggambarannya dengan menggunakan DFD (Data Flow Diagram) dikarenakan pembuatan sistem ini secara struktural.

### H. *Implementasi Sistem*

Tahap ini merupakan tahap mengimplementasikan sistem, dengan adanya tahap ini dimaksudkan untuk menggunakan sistem prediksi PJK yang telah dirancang sebelumnya. Pada tahap ini diharapkan sistem prediksi PJK dapat digunakan dengan baik.

## III. HASIL DAN PEMBAHASAN

### A. *Dataset*

Data ini digunakan sebagai percobaan model Algoritma Naïve Bayes yang digunakan. *Dataset* yang diperoleh merupakan variabel rekam medik pasien asli dan belum dilakukan Data Preprocessing. Pada *dataset* yang dimiliki terdapat tiga belas variabel dan satu kelas, ringkasan variabel *dataset* dapat dilihat pada tabel 1.

TABEL I.  
RINGKASAN VARIABEL *DATASET*

No	Variabel	Nilai
1	<i>age</i>	29-77
2	<i>sex</i>	0, 1
3	<i>cp</i>	0, 1, 2, 3
4	<i>trtbps</i>	94-200
5	<i>chol</i>	126-564
6	<i>fbs</i>	0, 1
7	<i>restecg</i>	0, 1, 2
8	<i>thalachh</i>	71-202
9	<i>exng</i>	0, 1

10	<i>oldpeak</i>	0,0-6,2
11	<i>slp</i>	0, 1, 2
12	<i>caa</i>	0, 1, 2, 3, 4
13	<i>thall</i>	0, 1, 2, 3
14	<i>output</i>	0, 1

*Dataset* rekam medik pada tahap ini akan dibersihkan serta diseleksi dari data yang mengandung *missing value*. Data awal rekam medik terdapat 303 *records* ditemukan data dengan *missing value*, sehingga *missing value* tersebut akan dihapus dan akan berpengaruh pada data awal yang nanti akan berubah.

### B. Data Processing

Tahap ini mengolah data yang diinput ke sistem dan sesuai dengan metode yang digunakan. Berdasarkan data yang diperoleh terdapat 303 data pasien dengan variabel berupa *Age* (usia), *Sex* (jenis kelamin), *CP/Chest Pain Type* (jenis nyeri dada), *trtbps/Trestbps* (tekanan darah), *Chol* (kolesterol), *Fbs/Fasting Blood Sugar* (gula darah), *Restecg* (hasil elektrokardiografi), *Thalachh* (detak jantung maksimum), *Exng/Exang* (nyeri dada apabila berolahraga), *Oldpeak* (segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat), *Slp/Slope* (kemiringan segmen ST dalam latihan maksimal), *Caa* (total vessel utama yang dilakukan pewarnaan oleh fluoroskopi), *Thall* (status jantung), dan *Output* (kelas dari variabel). Contoh *dataset* awal dapat dilihat pada tabel 2.

TABEL II.  
DATASET AWAL

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1

### 4. Data Cleaning

Data yang dibersihkan sebagai data aktif., lalu dibutuhkan seleksi variabel untuk memilih data yang relevan. Terdapat 5 record *missing value* pada variabel *caa* dan 2 record *missing value* pada variabel *thall*. Hasil dari pembersihan data tersebut menghasilkan jumlah *record* baru yaitu 296 *records*.

TABEL III.  
DATA MISSING VALUE PADA VARIABEL CAA

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
52	1	2	138	223	0	1	169	0	0	2	4	2	1
58	1	1	125	220	0	1	144	0	0.4	1	4	3	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
43	1	0	132	247	1	0	143	1	0.1	1	4	3	0

TABEL IV.  
DATA MISSING VALUE PADA VARIABEL THALL

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	out-put
53	0	2	128	216	0	0	115	0	0	2	0	0	1
52	1	0	128	204	1	1	156	1	1	1	0	0	0

### 5. Data Selection

Data yang dipilih hanya data yang memiliki variabel penting yang berguna dalam kategori prediksi penyakit jantung koroner tersebut. Variabel yang ada harus saling berhubungan satu sama lain untuk menentukan informasi yang didapat agar mendapatkan keluaran yang tergantung dengan kepentingannya.

Seleksi variabel dilakukan dalam tahapan ini, *dataset* seluruh variabel yang digunakan berjumlah 13, yaitu Age (*usia*), Sex (*jenis kelamin*), CP/Chest Pain Type (*jenis nyeri dada*), trtbps/Trestbps (*tekanan darah*), Chol (*kolesterol*), Fbs/Fasting Blood Sugar (*gula darah*), Restecg (*hasil elektrokardiografi*), Thalachh (*detak jantung maksimum*), Exng/Exang (*nyeri dada apabila berolahraga*), Oldpeak (*segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat*), Slp/Slope (*kemiringan segmen ST dalam latihan maksimal*), Caa (*total vessel utama yang dilakukan pewarnaan oleh fluoroskopi*), Thall (*status jantung*), dan Output (*kelas dari variabel*). Semua variabel ini saling berhubungan guna mencari informasi yang tepat untuk keluaran yang dibutuhkan.

### 6. Data Transformation

Nilai dari variabel kategori diubah menjadi simbol, sebagai berikut:

- Jenis Kelamin (sex), perempuan diberi nilai 0 dan laki-laki diberi nilai 1.
- Jenis Nyeri Dada (Chest Pain Type/cp), Asymptomatic Pain diberi nilai 0, Typical Angina diberi nilai 1, Atypical Angina diberi nilai 2, dan Non-Anginal Pain diberi nilai 3.
- Gula Darah (Fasting Blood Sugar/fbs), false diberi nilai 0 dan True diberi nilai 1.
- Hasil ECG (Resting Electrocardiographic/restecg), hype diberi nilai 0, Normal diberi nilai 1, dan Abnormal diberi nilai 2.
- Nyeri Dada Apabila Berolahraga (Exercise Induced Angina/exang), false diberi nilai 0 dan True diberi nilai 1.
- Slope, down disimbolkan dengan 0, flat diberi nilai 1, dan up diberi nilai 2.
- Status Jantung (Thalassemia/thall), fixed diberi nilai 1, Normal diberi nilai 1, dan Reversal diberi nilai 2.

### C. Algoritma Naïve Bayes

Tahap pertama pengambilan sampel. Data rekam medik yang diambil adalah data training yang sudah terbagi kedalam komposisi 60% dengan perolehan data sebanyak 178 record. Data sudah melalui tahap pra-proses maka dapat pada tabel 5 berikut:

TABEL V.  
KLASIFIKASI DATA TRAINING DAN TESTING

Keterangan	Jumlah Data (296 records)	Data Training (178 records)	Data Testing (118 records)
Tidak	136	87	49
Ya	160	91	69

Terdapat dua klasifikasi yang dibentuk pada kolom Output, yaitu YA dan TIDAK. Untuk menentukan seseorang tersebut terkena penyakit jantung koroner atau tidak ditentukan oleh beberapa parameter yaitu Age (*usia*), Sex (*jenis kelamin*), CP/Chest Pain Type (*jenis nyeri dada*), trtbps/Trestbps (*tekanan darah*), Chol (*kolesterol*), Fbs/Fasting Blood Sugar (*gula darah*), Restecg (*hasil elektrokardiografi*), Thalachh (*detak jantung maksimum*), Exng/Exang (*nyeri dada apabila berolahraga*), Oldpeak (*segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat*), Slp/Slope (*kemiringan segmen ST dalam latihan maksimal*), Caa (*total vessel utama yang dilakukan pewarnaan oleh fluoroskopi*), Thall (*status jantung*), dan Output (*kelas dari variabel*).

Data rekam medik yang diambil adalah data training yang sudah terbagi kedalam komposisi 60% dengan perolehan data sebanyak 178 record. Peluang penyakit jantung coroner (Ya) = 0.51 sedangkan (Tidak) = 0.49. nilai paling tinggi untuk variabel Age Tua Tidak PJK yaitu 0.82, nilai paling tinggi untuk variabel Sex Laki-laki Tidak PJK yaitu 0.82. Nilai paling tinggi untuk variabel CP 0: Asymptomatic Pain Tidak PJK yaitu 0.71. Nilai paling tinggi untuk variabel Trestbps Normal Ya PJK yaitu 0.81. Nilai paling tinggi untuk variabel Chol Tinggi Tidak PJK yaitu 0.86. Nilai paling tinggi untuk variabel Fbs Normal Tidak PJK yaitu 0.85. Nilai paling tinggi untuk variabel Restecg Normal Ya PJK yaitu 0.62. Nilai paling tinggi untuk variabel Thalach Tinggi Ya PJK yaitu 0.85. Nilai paling tinggi untuk variabel Exang Tidak Nyeri Ya PJK adalah 0.86. Nilai paling tinggi untuk variabel Oldpeak Normal Ya PJK adalah 0.99. Nilai paling tinggi untuk variabel Slope Flat Tidak PJK adalah 0.69. Nilai paling tinggi untuk variabel Caa 0 Ya PJK adalah 0.79. Nilai paling tinggi untuk variabel Thall 2 Ya PJK adalah 0.75. Tahapan berikutnya yaitu menghitung Joint Probability Distribution (JPB) suatu gejala untuk mengetahui pasien tersebut terkena PJK atau tidak. Berdasarkan hasil perhitungan JPB diatas didapatkan hasil untuk Ya PJK sebesar 0.000482 sedangkan untuk Tidak PJK sebesar 0.000002, maka Ya PJK > Tidak PJK atau berdasarkan data baru tersebut kemungkinan terkena penyakit jantung koroner.

Berikut dijelaskan proses perhitungan manual dengan metode *Naïve Bayes* dengan contoh data berikut:

TABEL VI.  
CONTOH DATA UJI

No	Variabel	Inputan Dokter
1	Age	35 (0: Dewasa)
2	Sex	0: Perempuan
3	Chest Pain Type	0: Asymptomatic Pain
4	Resting Blood Pressure	138 (0: Normal)
5	Cholesterol	183 (0: Normal)
6	Fasting Blood Sugar	0: False
7	Resting ECG Results	1: Normal
8	Maximum Heart Rate	182: Tinggi
9	Exercise Induced Angina	No (0: Tidak Nyeri)
10	ST Depression Induced	1.4 (0: Normal)
11	Slope	2: Up
12	Number of Vessels Colored by Flourosopy	0
13	Thalassemia	2: Normal
14	Output	NA

Berdasarkan contoh uji data pada Tabel 6, maka tahapan berikutnya dalam proses perhitungan *Naïve Bayes* adalah menghitung *Joint Probability Distribution* suatu gejala untuk dapat mengetahui apakah pasien tersebut terkena PJK atau tidak. Pertama-tama masukkan perhitungan peluang dan *prior* sesuai dengan parameter yang diinput data baru tersebut mulai dari peluang PJK iya dan tidak lalu perhitungan *prior* semua variabel yang ada, setelah itu dilanjutkan melakukan perhitungan *Joint Probability Distribution*.

Penyelesaian:

Peluang PJK  $P(\text{PJK} | \text{Ya})$

$$\frac{P(\text{Jumlah klasifikasi YA pada kolom Output})}{\text{Jumlah data}} = \frac{91}{178} = 0.51$$

*Prior Age*  $P(\text{Dewasa} | \text{Ya})$

$$\frac{P(\text{Jumlah data usia Dewasa YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{29}{91} = 0.32$$

*Prior Sex*  $P(\text{Perempuan} | \text{Ya})$

$$\frac{P(\text{Jumlah data jenis kelamin perempuan YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{44}{91} = 0.48$$

*Prior CP*  $P(\text{Asymptomatic Pain} | \text{Ya})$

$$\frac{P(\text{Jumlah data CP Asymptomatic YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{22}{91} = 0.24$$

*Prior Trestbps*  $P(\text{Normal} | \text{Ya})$

$$\frac{P(\text{Jumlah data Trestbps Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{74}{91} = 0.81$$

*Prior Chol*  $P(\text{Normal} | \text{Ya})$

$$\frac{P(\text{Jumlah data Chol Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{16}{91} = 0.18$$

*Prior Fbs* P(Normal | Ya)

$$\frac{P(\text{Jumlah data Fbs Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{76}{91} = 0.84$$

*Prior Restecg* P(Normal | Ya)

$$\frac{P(\text{Jumlah data Restecg Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{56}{91} = 0.62$$

*Prior Thalachh* P(Tinggi | Ya)

$$\frac{P(\text{Jumlah data Thalachh Tinggi YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{77}{91} = 0.85$$

*Prior Exang* P(Tidak Nyeri | Ya)

$$\frac{P(\text{Jumlah data Exang Tidak Nyeri YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{78}{91} = 0.86$$

*Prior Oldpeak* P(Normal | Ya)

$$\frac{P(\text{Jumlah data Oldpeak Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{90}{91} = 0.99$$

*Prior Slope* P(Up | Ya)

$$\frac{P(\text{Jumlah data Slope Up YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{58}{91} = 0.64$$

*Prior Caa* P(0 | Ya)

$$\frac{P(\text{Jumlah data Caa 0 YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{72}{91} = 0.79$$

*Prior Thall* P(Normal | Ya)

$$\frac{P(\text{Jumlah data Thall Normal YA PJK})}{\text{Jumlah data klasifikasi YA PJK}} = \frac{68}{91} = 0.75$$

*Joint Probability Distribution* (Ya PJK)

Hitung P(PJK | databaru)

$$\begin{aligned} &= P(\text{PJK}) * P(\text{Dewasa | Ya}) * P(\text{Perempuan | Ya}) * P(\text{Asymptomatic Pain | Ya}) * P(\text{Normal | Ya}) * P(\text{Normal | Ya}) \\ &* P(\text{Normal | Ya}) * P(\text{Normal | Ya}) * P(\text{Tinggi | Ya}) * P(\text{Tidak Nyeri | Ya}) * P(\text{Normal | Ya}) * P(\text{Up | Ya}) * P(0 | \\ &\text{Ya}) * P(\text{Normal | Ya}) \\ &= 0.51 * 0.32 * 0.48 * 0.24 * 0.81 * 0.18 * 0.84 * 0.62 * 0.85 * 0.86 * 0.99 * 0.64 * 0.79 * 0.75 \\ &= 0.000390 \text{ atau } 0.0390\% \end{aligned}$$

Peluang tidak PJK P(PJK | Tidak)

$$\frac{P(\text{Jumlah klasifikasi TIDAK pada kolom Output})}{\text{Jumlah data}} = \frac{87}{178} = 0.49$$

*Prior Age* P(Dewasa | Tidak)

$$\frac{P(\text{Jumlah data usia Dewasa TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{16}{87} = 0.18$$

*Prior Sex* P(Perempuan | Tidak)

$$\frac{P(\text{Jumlah data jenis kelamin perempuan TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{16}{87} = 0.18$$

*Prior CP* P(Asymptomatic Pain | Tidak)

$$\frac{P(\text{Jumlah data CP Asymptomatic Pain TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{62}{87} = 0.71$$

*Prior Trestbps* P(Normal | Tidak)

$$\frac{P(\text{Jumlah data Trestbps Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{60}{87} = 0.69$$

*Prior Chol* P(Normal | Tidak)

$$\frac{P(\text{Jumlah data Chol Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{12}{87} = 0.14$$

*Prior Fbs* P(Normal | Tidak)

$$\frac{P(\text{Jumlah data Fbs Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{74}{87} = 0.85$$

*Prior Restecg* P(Normal | Tidak)

$$\begin{aligned} \text{Prior Thalachh } P(\text{Tinggi} \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Restecg Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{37}{87} = 0.43 \\ \text{Prior Exang } P(\text{Tidak Nyeri} \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Thalachh Tinggi TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{49}{87} = 0.56 \\ \text{Prior Oldpeak } P(\text{Normal} \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Exang Tidak Nyeri TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{40}{87} = 0.46 \\ \text{Prior Slope } P(\text{Up} \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Oldpeak Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{79}{87} = 0.91 \\ \text{Prior Caa } P(0 \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Slope Up TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{19}{87} = 0.22 \\ \text{Prior Thall } P(\text{Normal} \mid \text{Tidak}) &= \frac{P(\text{Jumlah data Caa 0 TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{31}{87} = 0.36 \\ &= \frac{P(\text{Jumlah data Thall Normal TIDAK PJK})}{\text{Jumlah data klasifikasi TIDAK PJK}} = \frac{25}{87} = 0.29 \end{aligned}$$

**Joint Probability Distribution (Tidak PJK)**Hitung  $P(\text{Tidak PJK} \mid \text{databaru})$ 

$$\begin{aligned} &= P(\text{PJK}) * P(\text{Dewasa} \mid \text{Tidak}) * P(\text{Perempuan} \mid \text{Tidak}) * P(\text{Asymptomatic Pain} \mid \text{Tidak}) * P(\text{Normal} \mid \text{Tidak}) \\ &* P(\text{Normal} \mid \text{Tidak}) * P(\text{Normal} \mid \text{Tidak}) * P(\text{Normal} \mid \text{Tidak}) * P(\text{Tinggi} \mid \text{Tidak}) * P(\text{Tidak Nyeri} \mid \text{Tidak}) \\ &* P(\text{Normal} \mid \text{Tidak}) * P(\text{Up} \mid \text{Tidak}) * P(0 \mid \text{Ya}) * P(\text{Normal} \mid \text{Ya}) \\ &= 0.49 * 0.18 * 0.18 * 0.71 * 0.69 * 0.14 * 0.85 * 0.43 * 0.56 * 0.46 * 0.91 * 0.22 * 0.36 * 0.29 \\ &= 0.000002 \text{ atau } 0.0002\% \end{aligned}$$

**D. Evaluasi**

Evaluasi kinerja klasifikasi dapat mengukur seberapa akurat algoritma dalam memprediksi data. Dalam evaluasi kinerja ini, kelas yang memiliki data yang tersebar hampir merata dan kelas tidak seimbang akan dipertimbangkan. Untuk mengetahui kinerja metode pada penelitian ini, maka hal yang dilakukan adalah mengevaluasi kemampuan prediksi menggunakan data pelatihan untuk menghasilkan nilai akurasi dari metode yang telah diterapkan. Penelitian ini dilakukan dengan percobaan, dimana dalam percobaan satu data dibagi sebanyak 60% yang menghasilkan 177 data latih dan sebanyak 40% menghasilkan 119 data uji. Percobaan pertama ini menghasilkan 99 data diprediksi benar dan 20 diprediksi salah. *Confusion Matrix* percobaan pertama dapat dilihat pada tabel 6.

TABEL VII.  
CONFUSION MATRIX PERCOBAAN PERTAMA

Kelas Aktual	Kelas Prediksi	
	Benar	Salah
Benar	45	6
Salah	14	54

$$\text{Akurasi} = \frac{99}{119} = 0.831 = 83.1\%$$

Pada percobaan kedua dilakukan pembagian data sebanyak 70% yang menghasilkan 207 data latih dan sebanyak 30% menghasilkan 89 data uji. Percobaan kedua ini menghasilkan 73 data diprediksi benar dan 16 diprediksi salah. *Confusion Matrix* percobaan kedua dapat dilihat pada tabel 7.

*CONFUSION MATRIX PERCOBAAN KEDUA*

		Kelas Prediksi	
		Benar	Salah
Kelas Aktual	Benar	34	4
	Salah	12	39

$$Akurasi = \frac{73}{89} = 0.8202 = 82.02\%$$

Pada percobaan ketiga dilakukan pembagian data sebanyak 80% yang menghasilkan 236 data latih dan sebanyak 20% menghasilkan 60 data uji. Percobaan ketiga ini menghasilkan 49 data diprediksi benar dan 11 diprediksi salah. *Confusion Matrix* percobaan ketiga dapat dilihat pada tabel 8.

TABEL IX.  
*CONFUSION MATRIX PERCOBAAN KETIGA*

		Kelas Prediksi	
		Benar	Salah
Kelas Aktual	Benar	19	4
	Salah	7	30

$$Akurasi = \frac{49}{60} = 0.816 = 81.6\%$$

TABEL X.  
HASIL PERCOBAAN

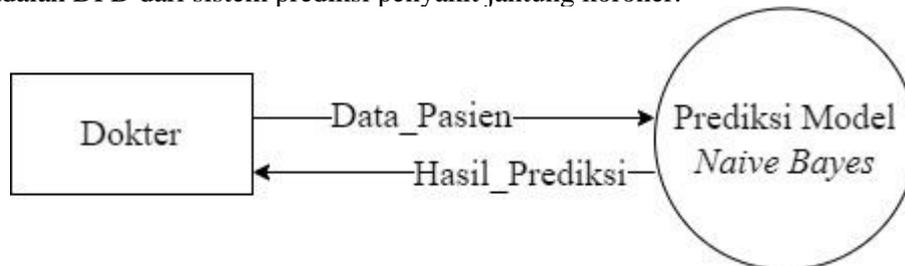
No	Percobaan	Data Latih	Data Uji	Akurasi
1	1	177 records (60%)	119 records (40%)	0.831 (83.1%)
2	2	207 records (70%)	89 records (30%)	0.8202 (82.02%)
3	3	236 records (80%)	60 records (20%)	0.816 (81.6%)

Maka dapat ditarik kesimpulan berdasarkan tabel 9 jumlah data sangat berpengaruh terhadap besar kecilnya tingkat akurasi yang akan diperoleh dan algoritma *Naïve Bayes* bisa diterapkan dalam memprediksi penyakit jantung koroner berdasarkan pemeriksaan dini pada data pasien.

**E. Rancangan Sistem**

1. Rancangan DFD (Data Flow Diagram)

Berikut ini adalah DFD dari sistem prediksi penyakit jantung koroner:



Gambar 2. Rancangan *Data Flow Diagram*

Terlihat pada Gambar 2, terdapat entitas eksternal yaitu dokter dengan menginput *data\_pasien* ke dalam sistem yang kemudian akan di proses untuk diprediksi dengan menggunakan Algoritma *Naïve Bayes*, setelah diproses sistem akan memberikan *output* berupa *hasil\_prediksi*.

**F. Implementasi Sistem**

Sistem prediksi PJK diakses melalui Spyder. Cara menjalankannya dengan running file *NaiveB.py* melalui Spyder sehingga muncul output pada Gambar 3. Running on <https://127.0.0.1:5000/> menandakan bahwa website dapat diakses melalui alamat url. Selanjutnya alamat tersebut dapat dijalankan pada halaman browser.

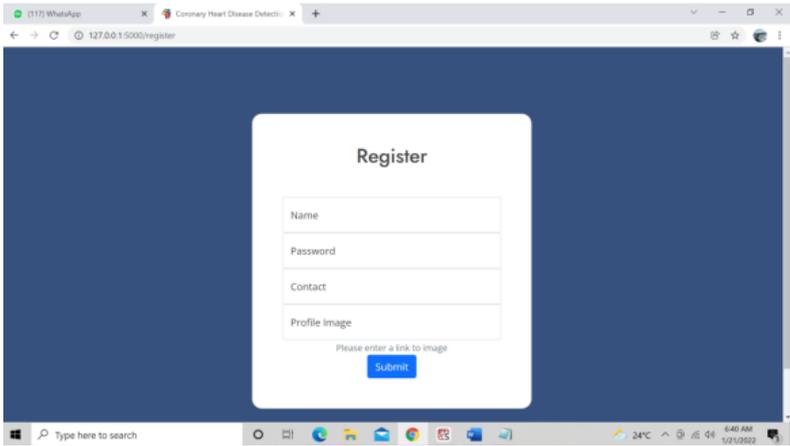
```

In [25]: runfile('C:/Users/my/Downloads/Syarat Skripsi/NaiveB/NaiveB.py',
wdir='C:/Users/my/Downloads/Syarat Skripsi/NaiveB')
* Serving Flask app "NaiveB" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production
setting.
Use a production WSGI server instead.
* Debug mode: off
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
  
```

Gambar 3. Menjalankan Sistem dengan Spyder

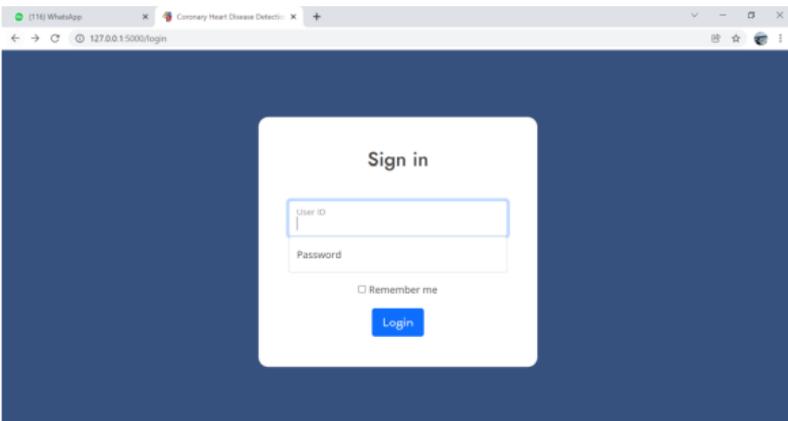
### 1. Halaman *Registration*

Pada Gambar 4 berikut, menunjukkan halaman add new user atau registration akun pada sistem prediksi PJK. Halaman ini digunakan untuk mendaftarkan akun pengguna, yaitu dokter. Pada halaman ini dokter diminta untuk mengisi beberapa data. Pada halaman ini terdapat tombol submit agar dilakukan pendaftaran akun. Dengan tombol ini, jika pengguna berhasil mendaftarkan akunnya maka akan langsung diarahkan menuju halaman home.


Gambar 4. Halaman *Registration*

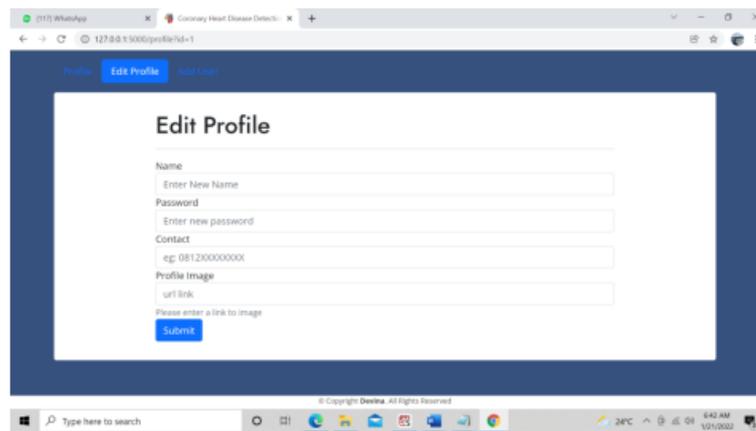
### 2. Halaman *Sign In*

Pada Gambar 5 berikut, menunjukkan halaman registrasi pada sistem prediksi PJK. Halaman ini berfungsi untuk masuk ke akun pengguna. Pada halaman ini pengguna diminta untuk mengisi *User ID* dan *password*. Pada halaman ini terdapat tombol *login*. Dengan tombol ini, jika pengguna berhasil masuk ke dalam akunnya maka akan langsung diarahkan menuju halaman *home*.


Gambar 5. Halaman *Sign In*

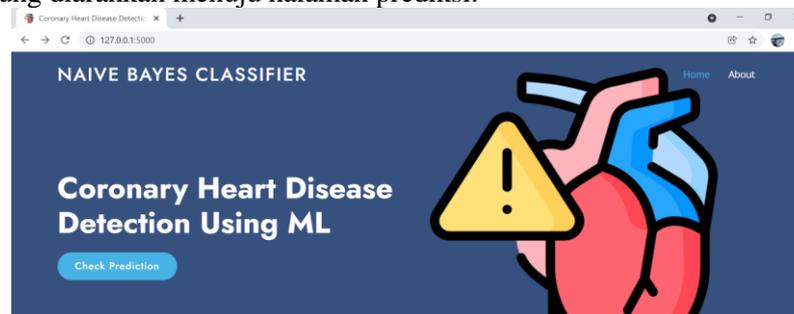
### 3. Halaman *Edit Profile*

Pada Gambar 6 berikut, menunjukkan halaman edit profile pada sistem prediksi PJK. Halaman ini digunakan untuk mengedit akun pengguna. Pada halaman ini pengguna dapat mengubah data kemudian terdapat tombol submit untuk menyimpan hasil perubahan data.

Gambar 6. Halaman *Edit Profile*

#### 4. Halaman *Home*

Pada gambar 7 bagian atas halaman utama, terdapat navigasi yang bisa memberikan arah pada pengguna untuk ke halaman *About*. Halaman tersebut terdapat juga tombol yang bertuliskan *Naïve Bayes Classifier*. Tombol ini juga akan mengarahkan pengguna ke halaman prediksi. Dengan tombol ini, pengguna juga dapat langsung diarahkan menuju halaman prediksi.

Gambar 7. Halaman *Home*

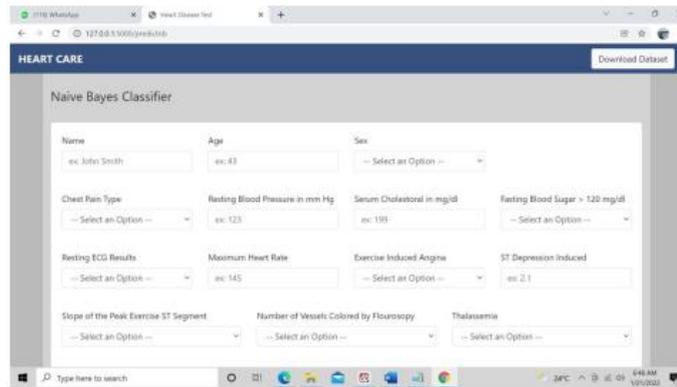
#### 5. Halaman *About*

Pada gambar 8, halaman ini berisi informasi mengenai penyakit jantung koroner, data yang akan di *input* oleh dokter serta informasi mengenai sistem yang digunakan.

Gambar 8. Halaman *About*

#### 6. Halaman *Input Data*

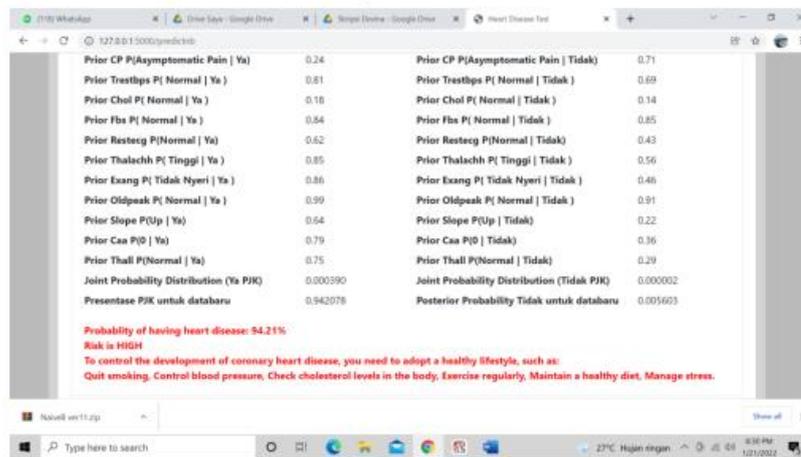
Pada gambar 9 yaitu halaman *input* data, dokter dapat menginput data pasien yang tersedia guna mendapatkan hasil prediksi. Data yang di *input* merupakan data uji untuk membandingkan hasil prediksi antara perhitungan manual dan perhitungan sistem yang dibuat. Kemudian, dokter dapat segera melakukan prediksi dengan klik tombol *result*.



Gambar 9. Halaman *Input* Data

7. Lihat Hasil Prediksi

Pada gambar 10 merupakan halaman dimana dokter sudah menginput data pasien dan klik tombol *result* maka akan tampil hasil prediksi.



Gambar 10. Halaman Lihat Hasil Prediksi

Dalam hal ini disimpulkan bahwa prediksi yang diperoleh adalah sama dan sesuai. Dengan hasil Ya pasien tersebut berisiko tinggi terkena PJK, tingkat akurasi yang didapat yaitu sebesar 0.942078 pada perhitungan manual atau hasil dari Posterior Probability dan sebesar 94.21% pada perhitungan sistem.

8. *Blackbox Testing*

Peneliti menggunakan blackbox testing untuk menguji dan mengetahui sistem apakah dapat bekerja dengan baik. Pada pengujian ini, sistem harus memberikan respon yang benar dan tepat agar sistem prediksi dapat digunakan.

TABEL XI.  
 BLACKBOX TESTING

Modul	Input	Proses	Output	Status
	Mengakses halaman url	Mengarahkan ke system	Menampilkan halaman awal	Berhasil
Halaman awal	Klik tombol <i>Naive Bayes Classifier</i>	Mengarahkan ke halaman prediksi	Menampilkan halaman prediksi	Berhasil
	Klik <i>About</i>	Mengarahkan ke halaman <i>About</i>	Menampilkan halaman <i>About</i>	Berhasil
Halaman Prediksi	Klik tombol <i>Result</i>	Menghitung hasil prediksi	Menampilkan hasil prediksi	Berhasil

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan tersebut, maka didapatkan kesimpulan bahwa *Algoritma Naive Bayes* dapat melakukan prediksi penyakit jantung koroner berdasarkan pemeriksaan dini pada pasien. Pada pemilihan pembagian data training dan data testing untuk model yang akan dibangun, hasil akan lebih optimal jika akurasi tertinggi yang dipilih. Berdasarkan evaluasi, hasil tertinggi yang didapatkan sebesar 83,1% pada percobaan pertama, maka perbandingan data training dan data testing yang dipilih yaitu 60:40.

#### DAFTAR PUSTAKA

- [1] Bella, Airindya. (2022). Macam-Macam Penyakit Jantung, Gejala, dan Penyebabnya. Diakses pada: <https://www.alodokter.com/macam-macam-penyakit-jantung-gejala-dan-penyebabnya>
- [2] P2PTM Kemenkes RI. (2019). Hari Jantung Sedunia (World Heart Day): Your Heart is Our Heart Too. Diakses pada: <http://p2ptm.kemkes.go.id/kegiatan-p2ptm/pusat-/hari-jantung-sedunia-world-heart-day-your-heart-is-our-heart-too>
- [3] Syahril, Muhammad, dkk. (2017). Penerapan Data Mining untuk Menentukan Pola Penjualan Peralatan Sekolah pada Brand Wigglo dengan Menggunakan Algoritma Apriori. J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD, Vol.3, No.1.
- [4] Vulandari, Retno Tri. (2020). Penerapan Metode Naïve Bayes untuk Klasifikasi Pelanggan. Jurnal TIKomSIN, Vol. 8, No. 2.
- [5] Daniel, D. J. T. S., dkk. (2020). Analisa dan Perancangan Sistem Informasi Administrasi Kursus Bahasa Inggris pada Intensive English Course di Ciledug Tangerang. JURNAL IPSIKOM Vol. 8 No.1.
- [6] Ayu, Fitri. (2018). Perancangan Sistem Informasi Pengolahan Data Praktek Kerja Lapangan (PKL) Pada Devisi Humas PT. PEGADAIAN. Jurnal Intra-Tech, Vol. 2, No.2.
- [7] Initu.id. (2018). Pengertian dan Jenis Prediksi, Ramalan (Forecasting) dan Implementasi. Diakses pada: <https://initu.id/amp/pengertian-dan-jenis-prediksi-ramalan-forecasting-dan-implementasi/>
- [8] Rabbani, Aletheia. (2020). Pengertian Prediksi dan Bentuknya. Diakses pada: <https://www.sosial79.com/2020/08/pengertian-prediksi-dan-bentuknya.html>
- [9] Kanal Pengetahuan. (2020). Pengertian Prediksi. Diakses pada: <https://www.kanal.web.id/pengertian-prediksi>
- [10] Belajarpython. (2021). Pendahuluan Python. Diakses pada: <https://belajarpython.com/tutorial/apa-itu-python>
- [11] Maulid, Reyvan. (2021). Mengenal Flask, Library Machine Learning Python Idaman Developer. Diakses pada: <https://dqlab.id/mengenal-flask-library-machine-learning-python-idaman-developer>
- [12] Rahman, Rashik. (2021). Heart Attack Analysis & Prediction Dataset. Diakses pada: <https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>