

Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes

Data Mining Implementation to Predict Heart Disease using Naive Bayes Method

Ade Riani^{*1}, Yessy Susianto², Nur Rahman³

^{1,2,3}Universitas Darwan Ali

e-mail: aderiani1013@gmail.com¹ Susiantoyessy@gmail.com², nurrahman.ikhtiar@gmail.com³

Abstrak

Penyakit jantung adalah penyakit dengan penyumbang angka kematian yang cukup tinggi dalam dunia kesehatan. Adapun penyakit ini biasanya jarang sekali disadari penyebabnya. Namun, ada beberapa parameter yang bisa digunakan untuk memprediksi bahwa seseorang itu memiliki resiko penyakit jantung atau tidak. Adapun untuk penelitian kali ini peneliti akan menggunakan beberapa indikator yang diantaranya adalah Age, Sex, Chest pain type, Trestbps, Cholesterol, Fasting blood sugar, Resting ecg, Max heart rate, Exercise induced angina, Oldpeak, Slope Number of vessels colored, dan Thal. Penelitian ini akan melakukan perhitungan menggunakan metode Data Mining dengan Algoritma Naive Bayes. Hasil dari penelitian ini mendapatkan akurasi sebesar 86% untuk 303 dataset yang yang diuji.

Kata Kunci: Data Mining, Naive Bayes, dan Penyakit Jantung

Abstract

Heart disease is a disease with a high mortality rate in the world of health. The disease is usually rarely realized the cause. However, there are several parameters that can be used to predict whether a person has a risk of heart disease or not. As for this study, researchers will use several indicators including Age, Sex, Chest pain type, Trestbps, Cholesterol, Fasting blood sugar, Resting ecg, Max heart rate, Exercise induced angina, Oldpeak, Slope, Number of vessels colored, and Thal This research will perform calculations using the Data Mining method with the Naive Bayes Algorithm. The results of this study get an accuracy of 86% for the 303 datasets tested.

Keywords: Data Mining, Naive Bayes, and Heart Disease

Pendahuluan

Penyakit jantung menjadi masalah utama dalam dunia kesehatan. Pada tahun 2015 data World Health Organization (WHO) menyatakan bahwa sekitar 17 juta lebih orang di dunia meninggal diakibatkan oleh penyakit jantung dan kerusakan pembuluh darah. Dimana kematian itu setara dengan 31% dari jumlah kematian yang ada diseluruh dunia. Adapun sekitar 8,7 juta orang meninggal diakibatkan oleh penyakit jantung. Fenomena penyakit ini biasanya terjadi di negara berkembang dimana rata-rata penduduknya memiliki penghasilan menengah kebawah.

Di Indonesia sendiri penyakit jantung telah memasuki fase yang cukup mengawatirkan. Hal ini dikarenakan trend penyakit jantung sudah tidak memandang usia. Dari data RISKESDAS (Riset Kesehatan Dasar) tahun 2018 meng-indikasikan bahwa sekitar 15 dari 1.000 penduduk Indonesia telah menderita penyakit jantung. Apabila dilihat dari penyebab kematian, menurut SRS (*Survei Sample Registration System*) tahun 2014 menunjukkan bahwa sekitar 12,9% kematian di Indonesia terjadi akibat penyakit jantung. Melihat terjadinya hal ini tentu saja memerlukan perhatian yang lebih untuk mengidentifikasi sejak dini kemungkinan – kemungkinan adanya resiko penyakit jantung.

Pada dasarnya penyakit jantung dapat dianalisis kemunculannya sejak dini. Namun, dikarenakan kurangnya pengetahuan menyebabkan banyak orang terlambat untuk menyadarinya. Keadaan ini apabila dibiarkan akan menjadi masalah yang serius untuk kesehatan kedepannya. Kemungkinan untuk sembuh dari penyakit jantung juga sangat kecil serta bagi mereka yang sudah terdokter penyakit jantung biasa sangat sulit untuk kembali sehat lagi. Satu – satunya jalan terbaik yang dapat diterapkan adalah menangani

*) Penulis Korespondensi : aderiani1013@gmail.com

penyakit jantung sejak dini. Berdasarkan penjelasan diatas, perlu sekali untuk kita mengetahui faktor apa saja yang mempengaruhi adanya resiko penyakit jantung agar kedepannya penyakit ini bisa ditangani sejak dini dengan langkah yang cepat dan tepat. Adapun sebelumnya telah ada beberapa penelitian yang dijadikan acuan untuk membantu dalam proses penyelesaian penelitian ini.

Penelitian yang dilakukan oleh M.Sabransyah, Yuki Novia Nasution, dan Fidia Deny Tisna Amijaya (2017). Menggunakan metode naive bayes dalam melakukan prediksi resiko penyakit jantung untuk mengurangi angka kematian mulai sejak dini. Penelitian tersebut menggunakan data pasien yang ada pada RSUD AWS dengan rentang waktu mulai dari November hingga Desember tahun 2016. Adapun objek yang diteliti adalah pria berusia 47 tahun, memiliki kadar kolesterol 198 mg / dL, tekanan darah 131 mmHg, orang tua yang memiliki catatan medis penyakit jantung, menderita diabetes mellitus, memiliki obesitas, memiliki dislipidemia tinggi. Dari hasil penelitian disimpulkan bahwa objek tersebut jatuh ke dalam "kategori potensial" memiliki penyakit jantung. Hasil klasifikasi yang telah dilakukan pada 25 data mendapatkan jumlah akurasi 80%, sedangkan hasil uji menggunakan 50 data mendapat jumlah akurasi 78% [1].

Penyakit Jantung

Salah satu fungsi jantung yang tidak normal adalah penyakit jantung *atherosclerosis* yang dihasilkan dari penumpukan plak yang mengeras dan membuat arteri menyempit. Tumpukan Plak ini menyebabkan aliran darah tidak lancar dan akhirnya terhambat. Serangan jantung dapat terjadi ketika aliran darah tersumbat karena bekuan darah [2].

Penyakit jantung terjadi ketika kumpulan plak yang menumpuk menghalangi aliran darah ke otot jantung pada arteri koronaria [3]. Ketika penimbunan lemak tumbuh mencapai sekitar 70% dari diameter arteri koroner, hal ini mungkin akan mulai menimbulkan gejala [4]. Gejala utama dari penyakit jantung adalah nyeri dada atau angina. Nyeri ini biasanya dapat terasa di bagian tengah dada, tetapi bisa menyebar atau hanya terasa di leher, bahu, lengan, atau rahang bawah, terutama pada sisi kiri badan. Untuk sebagian besar orang, gejala ini hampir selalu terjadi seiring atau setelah aktivitas fisik, perasaan emosional dan mungkin terjadi setelah aktifitas makan atau dalam cuaca dingin [5].

Data Mining

Data mining adalah sebuah proses pengekstraksian informasi yang di dapat dari sekumpulan data melalui penggunaan algoritma, teknik penarikan dalam bidang statistik, pembelajaran mesin dan juga sistem manajemen basis data [6]. Data mining merupakan proses penganalisaan data dari perspektif yang berbeda serta menyimpulkannya menjadi sebuah informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya [7].

Dalam definisi lain juga mengatakan bahwa data mining adalah sebuah kegiatan yang meliputi pengumpulan data, pemakaian data historis, pola atau hubungan dalam data yang berukuran besar. Dari beberapa definisi diatas dapat disimpulkan bahwa data mining adalah suatu proses atau kegiatan yang dilakukan untuk mengumpulkan data yang berukuran besar, lalu kemudian mengekstraksi data tersebut sehingga menjadi sebuah informasi yang dapat digunakan untuk sesuatu yang bermanfaat [8].

Tahap-tahap dalam Data Mining

Data Mining dibagi menjadi beberapa tahap proses. Adapun tahapan tersebut bersifat interaktif, pengguna terlibat langsung atau melalui perantaraan knowledge base. Tahapan – tahapan tersebut antara lain adalah sebagai berikut[9]:

a. Data Preprocessing

Langkah pertama yang harus dilakukan dalam pengecekan terhadap dataset pasien penyakit jantung yaitu melakukan tahapan Preprocessing. Data preprocessing adalah proses membersihkan data dari noise. Salah satu teknik data preprocessing untuk mengatasi ukuran database yang besar adalah dengan membagi database menjadi beberapa bagian sehingga akan mempercepat proses scanning data saat algoritma data mining diterapkan[10].

b. Seleksi Fitur

Proses ini dilakukan untuk mereduksi dataset pasien penyakit jantung dan memilih fitur-fitur yang relevan terhadap diagnosis penyakit jantung. Penelitian ini melakukan seleksi fitur menggunakan aplikasi Weka 3.8.

c. Pengacakan Data

Proses pengacakan data pada dataset pasien penyakit jantung dapat mempengaruhi performa dari diagnosis penyakit jantung. Hal tersebut disebabkan karena adanya model klasifikasi yang dilakukan selama pelatihan terhadap data. Penerapan pengacakan terhadap data akan dilakukan setelah hasil dari preprocessing dan hasil reduksi.

d. Klasifikasi Data

Klasifikasi terhadap dataset menggunakan Algoritma Naive Bayes untuk penyelesaiannya. Performa klasifikasi yang disajikan dengan melakukan 10-fold cross validation pada dataset hasil pengacakan. Penerapan 10-fold cross validation dilakukan berdasarkan default yang diperoleh dari software weka yang biasanya digunakan secara umum dalam berbagai penelitian. Sembilan fold pertama data akan dijadikan sebagai training sedang fold terakhir akan dijadikan sebagai testing. Kemudian sembilan fold kedua akan dijadikan testing sedangkan sisanya akan dijadikan testing.

e. Evaluasi Performa

Tahapan Evaluasi performa dilakukan dengan melihat nilai akurasi dari metode yang digunakan. Rumus akurasi yang digunakan adalah sebagai berikut:

$$\text{Akurasi} = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

Dari uraian diatas dapat dilihat bahwa Algoritma Naive Bayes dapat digunakan dalam membantu menghitung kemungkinan adanya resiko penyakit jantung, dengan begitu penyakit jantung bisa ditangani sejak dini. Serta dengan banyaknya penelitian mengenai prediksi penyakit jantung diharapkan dapat mengurangi angka kematian yang ada pada saat ini. Adapun pada penelitian ini hanya menjelaskan mengenai penerapan metode *Naive Bayes* ke dalam data mining untuk keperluan memprediksi adanya resiko penyakit jantung pada seseorang.

Metode Penelitian

Metode penelitian juga dilakukan untuk menemukan jawaban dari permasalahan yang terjadi dalam sistem, penyebab permasalahan tersebut dan cara mengatasi permasalahan yang ada. Dalam penyusunan laporan, digunakan beberapa metode dan berpegang pada beberapa cara, dimana cara-cara tersebut saling melengkapi untuk memperoleh informasi yang diperlukan dalam pembuatan Aplikasi Diagnosis Penyakit Jantung Koroner Menggunakan Metode Naive Bayes.

Adapun cara-cara tersebut adalah sebagai berikut:

a) Studi Pustaka (Library Research).

Studi pustakatrftgi adalah usaha-usaha yang dilakukan untuk memperoleh informasi tentang pembuatan dan perancangan sistem ini dengan literatur-literatur yang ada. Pada metode ini akan dilakukan pengumpulan informasi atribut atau hasil pengecekan laboratorium dari Machine Learning University of California Irvine yang berupa atribut-atribut dari jantung koroner.

b) Studi Dokumen

Mengumpulkan seluruh data-data dokumen ataupun file yang berisi angka, huruf, gambar, dan sample. Dengan tujuan untuk bahan analisis yang akan digunakan dalam pembuatan sistem.

Hasil dan Pembahasan

Pada bab ini akan membahas tentang tahapan – tahapan analisis terhadap dataset yang digunakan. Kemudian melakukan seleksi terhadap fitur-fitur dari dataset. Hasil seleksi fitur akan dianalisis menggunakan metode Naive Bayes. Hasil dari nilai performa akurasi akan dilakukan uji Paired t-Test untuk menentukan signifikansi dari setiap nilai akurasi metode yang digunakan.

Dataset

Dataset pada penelitian ini memiliki jumlah instance sebanyak 304, dengan data berupa angka dan huruf. Dataset memiliki atribut atau fitur yang mendeskripsikan setiap instance. Berikut ini adalah daftar atribut pada penelitian ini

Tabel 1. Ringkasan Atribut Dataset

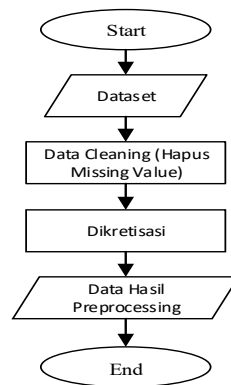
Atribut	Nilai
Age	29,0 – 77,0
Sex	Male, Female
Chest Pain type	Abnang, Angina, Asympt, Notang
Trestbps	94,0 – 200,0
Cholesterol	126,0 – 564,0
Fasting blood sugar	True, False
Resting ecg	Norm, Hyp, Abn
Max heart rate	99,0 – 103,0
Exercise induced angina	True, False
Oldpeak	0,0 – 6,2
Slope	Down, Flat, Up
Number of vessels colored	0,0 – 3,0
Thal	Normal, Rever, Fixed
Healthy	Healthy, Sick

Berikut penjelasan dari masing-masing fitur pada *dataset* pasien penyakit jantung:

1. *Age*: Usia dalam tahun (numerik);
2. *Sex*: 1: Laki-laki, 0: Perempuan;
3. *Chest pain type* (CP): (a) *typical angina* (1: *angina*), (b) *atypical angina* (2: *abnang*), (c) *non-anginal pain* (3: *notang*), (d) *asymptomatic* (4: *asympt*). Pengertian secara medis :
 - a. *Typical angina* adalah kondisi rekam medis pasien menunjukkan gejala umum nyeri dada sehingga kemungkinan memiliki penyumbatan arteri koroner yang tinggi.
 - b. *Atypical angina* adalah kondisi dimana gejala pasien tidak rinci sehingga kemungkinan penyumbatan lebih rendah.
 - c. *Non-anginal pain* adalah rasa sakit yang menusuk seperti pisau atau kondisi menyakitkan yang dapat berlangsung dalam jangka waktu pendek atau panjang
 - d. *Asymptomatic pain* tidak menunjukkan gejala penyakit dan kemungkinan tidak akan menyebabkan atau menunjukkan gejala penyakit.
4. *Trestbps*: *resting blood pressure* pasien dalam *mm Hg* pada saat masuk rumah sakit (numerik);
5. *Chol*: Serum kolesterol dalam *mg/dl* (numerik);
6. *Fbs*: Ukuran *boolean* yang menunjukkan apakah *fasting blood sugar* lebih besar dari 120 *mg/dl* atau tidak: (1: *True*; 0: *false*);
7. *Restecg*: Hasil *ECG* selama istirahat : 0: normal, 1: abnormal (memiliki kelainan gelombang *ST-T*), 2: *hipertrofi ventrikel*;
8. *Thalac*: detak jantung maksimum yang dicapai (numerik);
9. *Exang*: Ukuran *boolean* yang menunjukkan apakah latihan *angina* induksi telah terjadi (1: ya, 0: tidak);
10. *Oldpeak*: segmen *ST* yang diperoleh dari latihan relatif terhadap istirahat (numerik);
11. *Slope*: kemiringan segmen *ST* untuk latihan maksimum (puncak). Terdapat tiga jenis nilai yaitu 1: *upsloping*, 2: *flat*, 3: *downsloping*;
12. *Ca*: jumlah *vessel* utama yang diwarnai oleh fluoroskopi (numerik);
13. *Thal*: status jantung (3: normal, 6: cacat tetap, 7: cacat reversibel);
14. *The class attributes*: 0: sehat dan 1: sakit.

Data Preprocessing

Data preprocessing adalah langkah awal dalam mendiagnosis penyakit jantung. Pada penelitian ini data *preprocessing* terbagi menjadi dua langkah yaitu data *cleaning* dan *diskretisasi*. Gambar 1 memperlihatkan langkah-langkah pada tahap data *preprocessing*.



Gambar 1. Data Preprocessing

Data Cleaning

Langkah pertama yang harus dilakukan adalah data cleaning atau menghapus missing value, dataset yang terdiri dari 303 instance, tujuh instance diantaranya memiliki missing value pada fitur-fitur tertentu. Tabel 2 menunjukkan rincian missing value yang terdapat pada dataset.

Tabel 2. Rincian Atribut Dataset

Age	Sex	Cpt	Trest	Ch	Fbs	Restecg	Mhr	Eia	Oid	Slope	N.O.V.C	Thal	Classes
53	0	3	128	216	0	2	115	0	0	1	0	?	0
52	1	3	138	223	0	0	169	0	0	1	?	3	0
38	1	3	138	175	0	0	173	0	0	1	?	3	0
43	1	4	132	247	1	2	143	1	0	2	?	7	1
52	1	4	128	204	1	0	156	1	1	2	0	?	1
58	1	2	125	220	0	0	144	0	0	2	?	7	0
38	1	3	138	175	0	0	173	0	0	1	?	3	0

Diskterisasi

Setelah proses data cleaning selesai maka selanjutnya adalah tahap terakhir yaitu diskretisasi pada data dengan mengkategorikan nilai dari setiap atribut. Adapun penjelasan dari kategori dataset tersebut akan dijelaskan sebagai berikut:

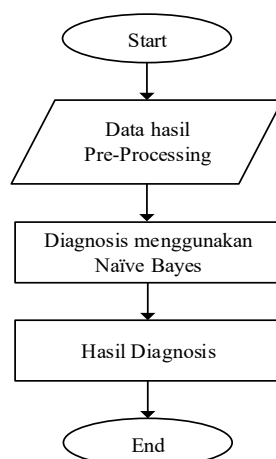
- a. *Age*
Atribut ini dikategorikan menjadi 2 nilai, yaitu umur dibawah 54 tahun dikategorikan dengan nilai 0, dan umur diatas 54 tahun dikategorikan menjadi nilai 1.
- b. *Trestbps*
Atribut ini dikategorikan menjadi 2 nilai, yaitu nilai yang kurang dari 148 dikategorikan dengan nilai 0, dan nilai lebih besar dari 148 dikategorikan menjadi nilai 1.
- c. *Cholestral*
Atribut ini dikategorikan menjadi 2 nilai, yaitu nilai yang kurang dari 346 dikategorikan dengan nilai 0, dan nilai lebih besar dari 346 dikategorikan menjadi nilai 1.
- d. *Mhr*
Atribut ini dikategorikan menjadi 2 nilai, yaitu nilai yang kurang dari 136 dikategorikan dengan nilai 0, dan nilai lebih besar dari 136 dikategorikan menjadi nilai 1.
- e. *Oldpeak*
Atribut ini dikategorikan menjadi 2 nilai, yaitu nilai yang kurang dari 3,2 dikategorikan dengan nilai 0, dan nilai lebih besar dari 3,2 dikategorikan menjadi nilai 1.
- f. *Novc*
Atribut ini adalah atribut numerik tetapi sudah mempunyai simbol seperti atribut kategori jadi sudah memiliki nilai yaitu 0,1,2 dan 3.

Setelah pengkategorisasian atribut numerik menjadi kategori selesai, maka selanjutnya adalah mengubah nilai dari atribut kategori menjadi simbol, berikut nilai dari atribut kategori yang akan dirubah menjadi simbol:

- a. *Sex*
Atribut ini mempunyai 2 kategori, jadi hanya akan dirubah menjadi 2 simbol. Male disimbolkan dengan nilai 0 dan female disimbolkan dengan nilai 1.
- b. *Cp*
Atribut ini mempunyai 4 kategori, dan dirubah menjadi 4 simbol. Notang disimbolkan dengan nilai 0, asympt disimbolkan dengan nilai 1, angina disimbolkan dengan nilai 2, dan abnang disimbolkan dengan nilai 3.
- c. *Fbs*
Atribut ini mempunyai 2 kategori, jadi hanya akan dirubah menjadi 2 simbol. False disimbolkan dengan nilai 0 dan True disimbolkan dengan nilai 1.
- e. *Restecg*
Atribut ini mempunyai 3 kategori, dan dirubah menjadi 3 simbol. Norm disimbolkan dengan nilai 0, hype disimbolkan dengan nilai 1, dan abn disimbolkan dengan nilai 2.
- f. *Eia*
Atribut ini mempunyai 2 kategori, jadi hanya akan dirubah menjadi 2 simbol. False disimbolkan dengan nilai 0 dan True disimbolkan dengan nilai 1.
- g. *Slope*
Atribut ini mempunyai 3 kategori, dan dirubah menjadi 3 simbol. Down disimbolkan dengan nilai 0, flat disimbolkan dengan nilai 1, dan up disimbolkan dengan nilai 2.
- h. *Thal*
Atribut ini mempunyai 3 kategori, dan dirubah menjadi 3 simbol. Norm disimbolkan dengan nilai 0, rev disimbolkan dengan nilai 1, dan fix disimbolkan dengan nilai 2.

Diagnosis Penyakit Jantung

Setelah data preprocessing selesai dilakukan, langkah selanjutnya adalah melakukan data processing. Adapun tahapannya antara lain sebagai berikut:



Gambar 2. Data Processing (Alir diagnosis penyakit jantung)

Berdasarkan Gambar 2, alur data dimulai dari input hasil data *preprocessing*, kemudian data tersebut diproses menggunakan metode *Naive Bayes*, dan setelah itu akan didapatkan hasil dari diagnosis tersebut apakah pasien mengidap penyakit atau tidak.

Diagnosis Menggunakan Naive Bayes

Berikut merupakan persamaan dari teorema Bayes[7] :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

Keterangan :

X :Data dengan class yang belum diketahui

H :Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

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

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Rumus teorema bayes ini digunakan untuk mendapatkan nilai persentase atribut yang didapat dari perhitungan probabilitas tiap atribut dengan klasifikasi 0 dan 1 berdasarkan tabel dataset yang ada.

Adapun langkah perhitungan probabilitas diagnosis penyakit jantung berdasarkan hasil laboratorium yaitu :

1. Tahap 1
Memisahkan 2 Class menjadi J1 untuk “Healthy = Healthy” dan J2 untuk “Healthy = Sick”.
2. Tahap 2
Mencari Prior Probability:
 - a. Probabilitas Healthy dengan hasil Healthy.
 - b. Probabilitas Healthy dengan hasil Sick.
3. Tahap 3
Mencari Likelihood dari setiap atribut.
4. Tahap 4
Mencari Likelihood Probability:
 - a. Probabilitas Diagnosis dengan hasil Healthy = Healthy berdasarkan hasil pemeriksaan pasien.
 - b. Probabilitas Diagnosis dengan hasil Healthy = Sick berdasarkan hasil pemeriksaan pasien.
5. Tahap 5
Menghitung Posterior Probability:
 - a. Mengalikan Probabilitas Diagnosis dengan hasil healthy = healthy berdasarkan hasil laboratorium pasien dari Likelihood probability dengan Probabilitas healthy = sick dari Prior Probability

- b. Mengalikan Probabilitas Diagnosis dengan hasil healthy = healthy berdasarkan hasil laboratorium seorang pasien dari Likelihood probability dengan Probabilitas healthy = sick dari Prior Probability
6. Tahap 6
 Membandingkan hasil dari Posterior Probability antara hasil healthy dan sick. Nilai yang lebih besar merupakan hasil dari diagnosis jantung koroner berdasarkan hasil laboratorium menggunakan metode Naive Bayes.

Berdasarkan hasil dari laboratorium menggunakan metode *Naive Bayes*. Berikut adalah langkah perhitungan probabilitas diagnosis penyakit jantung dengan kasus sebagai berikut:

Tabel 3. Contoh Kasus

<i>Age</i>	> =54
<i>Sex</i>	Perempuan
<i>Chest_Pain_Type</i>	Asympt
<i>Trestbps</i>	> =148
<i>Cholesterol</i>	1-345
<i>Fasting_Blood_Sugar</i>	Tidak
<i>Resting_Ecg</i>	Hyper
<i>Max_Heart_Rate</i>	> =136
<i>Exercise_Induced_Angina</i>	Tidak
<i>Oldpeak</i>	> =3,2
<i>Slope</i>	Down
<i>Number_Of_Vessels_Colored</i>	2
<i>Thal</i>	Rever

1. Memisahkan 2 *Class* menjadi J1 untuk “*Healthy* = Tidak” / Sehat dan J2 untuk “*Healthy* = Ya” / Sakit.
2. Mencari *Prior Probability*:

- a. Probabilitas *Healthy* dengan hasil Tidak atau (Sehat)

$$\frac{\text{Healthy dengan hasil (J1)}}{\text{Total Record}} = \frac{160}{296} = 0,540540541$$

- b. Probabilitas *Healthy* dengan hasil Ya atau (Sakit)

$$\frac{\text{Healthy dengan hasil (J2)}}{\text{Total Record}} = \frac{136}{296} = 0,459459459$$

3. Mencari *Likelihood* dari setiap atribut

- a. *Age*

Mencari *Likelihood* pada atribut *Age* dilakukan dalam 2 tahap, yakni:

1. *Likelihood Age Ya* (karena di kasus, hasil yang diperoleh adalah 1) dengan hasil class J1(Tidak/Sehat).

$$\frac{\text{Likelihood Age (Ya)}}{J1} = \frac{74}{160} = 0,462500000$$

2. *Likelihood Age Ya* (karena di kasus, hasil yang diperoleh adalah 1) dengan hasil class J2 (Ya/Sakit).

$$\frac{\text{Likelihood Age (Tidak)}}{J2} = \frac{100}{136} = 0,735294118$$

- b. *Sex*

1. J1

$$\frac{\text{Likelihood Sex (Ya)}}{J1} = \frac{71}{160} = 0,443750000$$

2. J2

$$\frac{\text{Likelihood Sex (Tidak)}}{J2} = \frac{24}{136} = 0,176470588$$

- c. *Chest_Pain_Type*

1. J1

$$\frac{\text{Likelihood Chest Pain Type (Ya)}}{J_1} = \frac{39}{160} = 0,243750000$$

2. J2

$$\frac{\text{Likelihood Chest Pain Type (Tidak)}}{J_2} = \frac{102}{136} = 0,750000000$$

d. *Trestbps*

1. J1

$$\frac{\text{Likelihood Trestbps (Ya)}}{J_1} = \frac{23}{160} = 0,143750000$$

2. J2

$$\frac{\text{Likelihood Trestbps (Tidak)}}{J_2} = \frac{30}{136} = 0,220588231$$

4. Mencari *Likelihood Probability*:

a. Probabilitas Diagnosis dengan hasil *Healthy* (Tidak / Sehat) berdasarkan 13 *Likelihood Probability* dari setiap atribut.

$$0,462500000 * 0,443750000 * 0,243750000 * 0,143750000 * 0,968750000 * 0,856250000 * 0,418750000 * 0,881250000 * 0,856250000 * 0,012500000 * 0,056250000 * 0,043750000 * 0,168750000 = 9,78417E-09$$

Probabilitas Diagnosis dengan hasil *Healthy* (Ya / Sakit) berdasarkan 13 *Likelihood Probability* dari setiap atribut.

$$0,735294118 * 0,176470588 * 0,750000000 * 0,220588235 * 0,977941176 * 0,852941176 * 0,573529412 * 0,580882353 * 0,455882353 * 0,125000000 * 0,088235294 * 0,227941176 * 0,647058824 = 4,42409E-06$$

5. Menghitung *Posterior Probability*:

a. *Likelihood J1 * J1*

$$9,78417E-09 * 0,540540541 = 5,28874E-09$$

b. *Likelihood J2 * J2*

$$4,42409E-06 * 0,459459459 = 2,03269E-06$$

Akhirnya diketahui hasil diagnosis dengan membandingkan hasil *Posterior probability* J1= 5,28874E-09 dan J2= 2,03269E-06, sehingga hasil diagnosis adalah J1, Ya atau Sehat, karena J1 lebih besar dari J2

Kesimpulan

Diagnosis terhadap penyakit jantung ini memiliki 4 (empat) buah kemungkinan hasil yang dapat dihasilkan yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). *True positive* dan *true negative* adalah hasil diagnosis yang benar. *False positive* adalah kesalahan prediksi yang terjadi ketika pasien dalam keadaan sehat tetapi diprediksi positif sakit jantung. *False negative* adalah kesalahan prediksi yang terjadi ketika pasien menderita penyakit jantung tetapi diprediksi negative (sehat). Akurasi hasil diagnosis bisa dihitung dengan rumus $\frac{\text{True positive} + \text{True Negative}}{\text{True positive} + \text{False positive} + \text{True negative} + \text{False negative}}$. Hasil akhir dari dataset sebanyak 303 memiliki akurasi 86%.

Daftar Pustaka

- [1] M. Sabransyah, Y. N. Nasution, dan F. D. T. Amijaya, "Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung," *JURNAL EKSPONENSIAL*, vol. 8, no. 2, hlm. 111–118, 2017.
- [2] M. CHABIB, "Persepsi Perempuan Tentang Penyakit Jantung Koroner di Puskesmas Jenangan, Kecamatan Jenangan Kabupaten Ponorogo," skripsi, Universitas Muhammadiyah Ponorogo, 2017.
- [3] R. Rulandi, "Hubungan Karakteristik Antara Usia, Jenis Kelamin, Tekanan Darah dan Dislipidemia dengan Kejadian Penyakit Jantung Koroner di Rumah Sakit Al-Ihsan Tahun 2014," Thesis, Fakultas Kedokteran (UNISBA), 2016.
- [4] "Studi Penggunaan Golongan Statin Pada Pada Pasien Jantung Koroner (Penelitian dilakukan di Rumah Sakit Umum Daerah Sidoarjo) - UMM Institutional Repository." [Daring]. Tersedia pada: <http://eprints.umm.ac.id/42665/>. [Diakses: 04-Des-2019].
- [5] T. Alawiyah, "Gambaran Pengetahuan Penderita Pjk Tentang Bahaya dan Akibat Makanan yang Mengandung Kolesterol," undergraduate, Universitas Muhammadiyah Semarang, 2018.

-
- [6] “Konsep Data Mining Vs Sistem Pendukung Keputusan - Dicky Nofriansyah, S.Kom., M.Kom. - Google Buku.” [Daring]. Tersedia pada:
https://books.google.co.id/books?hl=id&lr=&id=PoJyCAAQBAJ&oi=fnd&pg=PR6&dq=Data+mining+adalah+sebuah+proses+pengekstraksian+informasi+yang+didapat+dari+sekumpulan+data+melalui+penggunaan+algoritma,+teknik+penarikan+dalam+bidang+statistik,+pembelajaran+mesin+dan+juga+sistem+manajemen+basis+data.+&ots=YVLk_ktZjO&sig=talhSMKAm_91oRoQ7C7beNNF1Cg&redir_esc=y#v=onepage&q&f=false. [Diakses: 04-Des-2019].
- [7] A. Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” vol. 2, no. 3, hlm. 11, 2015.
- [8] D. S. Siltonga, S. Saifullah, dan R. Dewi, “Analisis Metode Naive Bayes dalam Memprediksi Tingkat Pemahaman Mahasiswa Terhadap Mata Kuliah Berdasarkan Posisi Duduk,” dalam *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019, vol. 1, hlm. 427–436.
- [9] Nurahman dan Prihandoko, “Perbandingan Hasil Analisis Teknik Data Mining ‘Metode Decision Tree, Naive Bayes, Smo Dan Part’ untuk Mendiagnosa Penyakit Diabetes Mellitus,” *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, Jan 2019.
- [10] D. Gunawan, “Evaluasi Performa Pemecahan Database dengan Metode Klasifikasi Pada Data Preprocessing Data mining,” *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 2, no. 1, hlm. 10-13–13, Jun 2016.