

Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung

Naive Bayes Method for a Heart Risk Disease Prediction Application

M. Sabransyah¹, Yuki Novia Nasution², dan Fidia Deny Tisna Amijaya³

¹Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

^{2,3}Jurusan Matematika FMIPA Universitas Mulawarman

¹E-mail: sabransyah.muh@gmail.com

Abstract

Classification is an activity for assessing object data which include it the data into particular class among any number of classes available. Naive Bayes is classification with probability method. This research examines the use of naive Bayes method for a heart disease risk prediction application. In this research, it will be classified a person who have the risk of heart disease by using the data of patient in RSUD AWS during November and December 2016 the sample case is 47 years old male object, has cholesterol level of 198 mg/dL, has blood pressure of 131 mmHg, parents having heart disease medical record, suffering diabetes Mellitus, has obesity, has high dyslipidemia. It is concluded that the object falls into "potential category" of having heart disease. The classification result that has been done, the exact accuracy was obtained with 25 tested data and got accuracy level in an amount of 80% and 50 tested data sample and got accuracy level in an amount of 78%.

Keywords : classification, naive Bayes, heart disease

PENDAHULUAN

Sejalan dengan pertumbuhan teknologi informasi, *data mining* merupakan salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari *database* skala besar yang semakin banyak terakumulasi. Virgiawan dan Mukhlash (2013) mendefinisikan *data mining* sebagai proses ekstraksi (meringkas) pola-pola yang menarik (implisit, tidak diketahui sebelumnya, dan berpotensi untuk dimanfaatkan) dari data yang berukuran besar. Ada beberapa macam pengklasifikasian dalam *data mining* yaitu *decision tree*, *naive Bayes*, *support vector machine* (SVM), dan lain-lain.

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan *naive* di mana diasumsikan kondisi antar petunjuk (atribut) saling bebas. Klasifikasi *naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2013). Salah satu pengaplikasian dari *naive Bayes* yaitu pada bidang kesehatan.

Kesehatan adalah keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial dan ekonomi (Yuniastuti, 2008). Tidak ada yang memiliki keinginan untuk sakit. Keadaan tubuh dalam kondisi sakit membuat apapun yang ingin dilakukan menjadi tidak maksimal atau tidak nyaman bahkan tidur, duduk, maupun makan tidak akan pernah terasa nyaman jika dalam keadaan sakit, maka menjaga kesehatan adalah

salah satu hal yang penting untuk selalu diupayakan.

Saat ini secara global Penyakit Tidak Menular (PTM) menjadi penyebab utama kematian. Data *World Health Organization* (WHO) menunjukkan bahwa dari 57 juta kematian yang terjadi di dunia pada tahun 2008, sebanyak 36 juta atau hampir dua pertiga disebabkan oleh PTM. Proporsi penyebab kematian yang disebabkan oleh PTM pada orang-orang berusia kurang dari 70 tahun, penyakit kardiovaskuler merupakan penyebab terbesar mencapai 39%, diikuti kanker sebesar 27%, sedangkan penyakit pernafasan kronis, penyakit pencernaan dan PTM yang lain bersama-sama menyebabkan sekitar 30%, serta 4% kematian disebabkan diabetes. Menurut WHO, kematian akibat PTM diperkirakan akan terus meningkat di seluruh dunia, peningkatan terbesar akan terjadi di negara-negara menengah dan miskin. Saat ini di Indonesia, menurut hasil Survei Kesehatan Rumah Tangga (SKRT) tahun 1995 dan 2001, serta hasil Riset Kesehatan Dasar (Riskesmas) tahun 2007, Data menunjukkan kematian disebabkan penyakit menular tahun 1995 sebesar 44,2%, tahun 2001 sebesar 31,2%, dan tahun 2007 sebesar 28,1%, kematian yang disebabkan PTM tahun 1995 sebesar 41,7%, tahun 2001 sebesar 49,9%, dan tahun 2007 sebesar 59,5%, sedangkan kematian yang disebabkan faktor lain yang tidak diketahui berada di bawah 19%. Penyebab kematian telah terjadi transisi epidemiologi di mana kematian disebabkan PTM semakin meningkat, sedangkan kematian karena penyakit menular semakin menurun. Dapat diketahui penyebab kematian nomor satu disebabkan dari PTM yaitu pada penyakit kardiovaskuler, penyakit kardiovaskuler adalah

penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah. Ada banyak macam penyakit kardiovaskuler, salah satunya adalah penyakit jantung (Kemenkes RI, 2014).

Pada dasarnya ciri-ciri penyakit jantung bisa dikenali sejak dini, tapi banyak orang yang belum dibekali pengetahuan yang cukup mengenai penyakit jantung menyebabkan banyak orang yang terlambat mengetahui jika dirinya terkena penyakit jantung. Jika sudah terjangkit penyakit jantung, tentu saja proses penyembuhannya akan menjadi lebih sulit, membutuhkan waktu yang lama dan biaya yang tidak sedikit. Oleh karena itu mendeteksi sejak dini harus diupayakan agar penyembuhan menjadi mudah, lebih cepat dan biaya yang sedikit.

Data Mining

Data mining, sering juga disebut *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Saat ini istilah pengenalan pola (*pattern recognition*) jarang digunakan karena ia termasuk bagian dari *data mining* (Prasetyo, 2012).

Sebagai suatu rangkaian proses, tahap-tahap dalam *Data Mining* menurut Saleh (2015) adalah sebagai berikut:

- a. Pembersihan data (*Data Cleaning*)
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
- b. Integrasi data (*Data Integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- c. Seleksi data (*Data Selection*)
Data yang ada pada *database* terkadang tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
- d. Transformasi data (*Data Transformation*)
Data diubah atau digabung ke dalam satu *database* yang sesuai untuk diproses dalam *data mining*.
- e. Proses *Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- f. Evaluasi pola (*Pattern Evaluation*)
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (*Knowledge Presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan.

Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototipe untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya (Prasetyo, 2012).

Metode Naive Bayes

Menurut Bustami (2013), algoritma *naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan *naive* di mana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Persamaan dari teorema Bayes adalah

$$P(C | F) = \frac{P(C).P(F | C)}{P(F)} \tag{1}$$

Keterangan :

$P(C|F)$: Probabilitas akhir bersyarat (*posterior*) suatu Kelas C terjadi jika diberikan Petunjuk (atribut) F terjadi

$P(C)$: Probabilitas awal (*prior*) Kelas C terjadi tanpa memandang petunjuk (atribut) apapun

$P(F|C)$: Probabilitas sebuah petunjuk (atribut) F terjadi akan mempengaruhi Kelas C

$P(F)$: Probabilitas awal (*prior*) petunjuk (atribut) F terjadi tanpa memandang Kelas apapun

Untuk menjelaskan teorema *naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk (atribut) untuk menentukan kelas apa yang cocok bagi objek yang dianalisis tersebut. Karena itu, Teorema Bayes di atas disesuaikan sebagai berikut:

$$P(C | F_1, \dots, F_n) = \frac{P(C).P(F_1, \dots, F_n | C)}{P(F_1, \dots, F_n)} \tag{2}$$

Di mana variabel C merepresentasikan kelas, sementara variable $F_1 \dots F_n$ merepresentasikan sejumlah petunjuk (atribut) yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya objek dengan petunjuk tertentu (atribut tertentu) dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuknya objek tersebut, seringkali disebut *prior*), dikali dengan

peluang kemunculan petunjuk-petunjuk (atribut) objek pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan petunjuk-petunjuk (atribut) objek secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan kelas suatu objek akan diklasifikasikan. Mengklasifikasikan suatu objek dapat ditentukan dengan memilih kelas yang memiliki *posterior* terbesar karena nilai *evidence* selalu tetap dan merupakan pembagi pada setiap kelasnya sehingga dalam perhitungan *posterior* cukup mengalikan nilai *prior* dengan *likelihood*. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $P(C|F_1, \dots, F_n)$ menggunakan aturan perkalian berikut :

$$\begin{aligned} P(C | F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n | C) \\ &= P(C)P(F_1 | C)P(F_2, \dots, F_n | C, F_1) \\ &= P(C)P(F_1 | C)P(F_2 | C, F_1) \\ &P(F_3 | C, F_1, F_2) \dots P(F_n | C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad (4)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu per satu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (*naive*), bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (*independent*) satu sama lain. Dengan asumsi tersebut, berlaku suatu kesamaan sebagai berikut :

$$P(F_i | F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (5)$$

Untuk $i \neq j$, sehingga $P(F_i | C, F_j) = P(F_i | C) \quad (6)$

Persamaan di atas dapat disimpulkan bahwa asumsi independensi *naive* tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi,

$$\begin{aligned} P(C | F_1, \dots, F_n) &= P(C)P(F_1 | C)P(F_2 | C) \dots P(F_n | C) \\ &= P(C) \prod_{i=1}^n P(F_i | C) \end{aligned} \quad (7)$$

Persamaan di atas merupakan model dari teorema *naive* Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kuantitatif atau kontinu digunakan rumus densitas *Gauss*:

$$P(F_i = f_i | C = c_j) = \frac{1}{\sqrt{2\pi s_j}} e^{-\frac{(f_i - \bar{x}_j)^2}{2s_j^2}} \quad (8)$$

Keterangan :

P :Peluang

F_i :Atribut ke- i

f_i :Nilai Atribut ke- i

C :Kelas yang dicari

c_j : Sub kelas yang dicari

\bar{x}_j : Rata-rata hitung (*Mean*) dari suatu atribut pada kelas ke- j

s_j^2 : Variansi dari suatu atribut pada kelas ke- j

s_j : Simpangan baku (Standar Deviasi) dari suatu atribut pada kelas ke- j

Pengukuran Tingkat Akurasi

Dalam melakukan klasifikasi diharapkan dapat melakukan klasifikasi pada semua objek dengan benar, sehingga dalam klasifikasi juga harus diukur tingkat akurasi. Untuk mendapatkan tingkat akurasi dari hasil prediksi maka digunakan rumus berikut (Siahaan, 2015):

$$Akurasi = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah data}} \times 100\% \quad (9)$$

Penyakit Jantung

Aliran darah dari dan ke jantung sangat dipengaruhi oleh kondisi jantung beserta seluruh sistem pembuluh darah pada setiap orang. Kondisi jantung seseorang dapat dilihat dari tekanan darah atas (sistolik) maupun tekanan darah bawah (diastolik) dan kondisi pembuluh-pembuluh darah yang digunakan sebagai jalan aliran darah ke seluruh tubuh. Bila kondisi jantung tidak optimal dan pembuluh darah banyak menyempit maka berbagai keluhan akan dirasakan. Keluhan tersebut dirasakan sebagai penyakit jantung. Berdasarkan proses terjadinya. Penyakit jantung dapat dibedakan menjadi penyakit jantung sejak lahir (penyakit jantung bawaan) dan penyakit jantung yang timbul pada saat umur tertentu (Mursito, 2002).

Hasil dan Pembahasan

Tahap awal yang dilakukan dalam penelitian ini adalah analisis deskriptif, bertujuan untuk menggambarkan karakteristik data pasien rawat inap di RSUD AWS bulan November dan Desember 2016. Karakteristik yang digambarkan pada analisis deskriptif adalah jenis kelamin, umur, kadar kolesterol, tekanan darah, riwayat penyakit jantung orang tua, diabetes melitus, obesitas, dislipidemia, dan penyakit jantung (kelas).

Klasifikasi Naive Bayes

Dalam proses menghitung klasifikasi *naive* Bayes, terdapat 3 alur yaitu membaca data *training*, menghitung nilai probabilitas setiap atribut pada setiap kelasnya dengan memperhatikan apakah data tersebut merupakan data kualitatif atau kuantitatif dan menentukan probabilitas akhir. Probabilitas akhir yang didapat digunakan untuk menentukan termasuk dalam kelas ya atau tidak untuk data yang ingin diuji. Misalkan, sebagai contoh kasus akan memprediksi kelas pada seseorang dengan nilai-nilai atribut jenis kelamin adalah laki-laki,

berumur 47 tahun, memiliki kadar kolesterol sebesar 198 mg/dL termasuk kategori tidak, memiliki tekanan darah sebesar 131 mmHg termasuk kategori prehipertensi, orang tua pernah terkena penyakit jantung, menderita diabetes melitus, tinggi badan 159 cm berat badan 70 kg termasuk kategori obesitas, dengan kadar *High Density Lipoprotein* (HDL) sebesar 30 mg/dL termasuk dislipidemia kategori tinggi. Selanjutnya akan ditentukan apakah orang tersebut termasuk kelas ya (artinya orang tersebut memiliki potensi besar terkena penyakit jantung) atau tidak (artinya orang tersebut memiliki potensi kecil terkena penyakit jantung). Jika termasuk kelas ya maka dianjurkan segera cek kesehatan jantung di rumah sakit terdekat.

Untuk menentukan data yang nantinya akan dianalisis dengan metode *naive* Bayes maka tahap pertama yang dilakukan adalah membaca data *training*. Tahap kedua adalah menentukan nilai probabilitas setiap atribut pada setiap kelasnya dengan memperhatikan apakah data tersebut merupakan data kualitatif atau kuantitatif. Pada penelitian ini, atribut yang termasuk dalam data kualitatif yaitu atribut jenis kelamin (F_1), kadar kolesterol (F_3), tekanan darah (F_4), riwayat penyakit jantung orang tua (F_5), diabetes melitus (F_6), obesitas (F_7), dislipidemia (F_8), dan penyakit jantung (C), sedangkan atribut yang termasuk dalam data kuantitatif yaitu atribut umur (F_2).

Adapun nilai probabilitas setiap atribut pada setiap kelasnya untuk data kualitatif sebagai berikut:

Berdasarkan Jenis Kelamin (F_1). Pada perhitungan nilai probabilitas atribut jenis kelamin pada kelas ya terdiri dari 90 pasien rawat inap berjenis kelamin laki-laki dan 43 pasien rawat inap berjenis kelamin perempuan, sedangkan pada kelas tidak terdiri dari 62 pasien rawat inap berjenis kelamin laki-laki dan 71 pasien rawat inap berjenis kelamin perempuan. Adapun nilai probabilitas atribut jenis kelamin pada setiap kelasnya yaitu sebagai berikut:

Probabilitas atribut jenis kelamin laki-laki untuk kelas ya, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin} = L|Y) = \frac{90}{90+43} = 0,677$$

Jadi, peluang pasien berjenis kelamin laki-laki terkena penyakit jantung sebesar 0,677.

Probabilitas atribut jenis kelamin perempuan untuk kelas ya, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin} = P|Y) = \frac{43}{90+43} = 0,323$$

Jadi, peluang pasien berjenis kelamin laki-laki terkena penyakit jantung sebesar 0,323.

Probabilitas atribut jenis kelamin laki-laki untuk kelas tidak, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin} = L|T) = \frac{62}{62+71} = 0,466$$

Jadi, peluang pasien berjenis kelamin laki-laki tidak terkena penyakit jantung sebesar 0,466.

Probabilitas atribut jenis kelamin perempuan untuk kelas tidak, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin} = P|T) = \frac{71}{62+71} = 0,534$$

Jadi, Peluang pasien berjenis kelamin perempuan tidak terkena penyakit jantung sebesar 0,534. Adapun nilai probabilitas atribut jenis kelamin pada setiap kelasnya dapat dilihat pada Tabel 1.

Tabel 1. Probabilitas Jenis Kelamin pada Setiap Kelasnya

Jenis Kelamin	Kelas		Probabilitas Kelas	
	Ya	Tidak	Ya	Tidak
Laki-laki	90	62	0,677	0,466
Perempuan	43	71	0,323	0,534

Dengan cara yang sama diperoleh probabilitas atribut kadar kolesterol (F_3), tekanan darah (F_4), riwayat penyakit jantung orang tua (F_5), diabetes melitus (F_6), obesitas (F_7), dislipidemia (F_8), dan penyakit jantung (C). Adapun nilai probabilitas seluruh atribut pada setiap kelasnya untuk data kualitatif terdapat pada Tabel 2.

Tabel 2. Probabilitas Seluruh Atribut Data Kualitatif pada Setiap Kelasnya

Atribut	Kategori	Probabilitas	
		Ya	Tidak
Jenis Kelamin	Laki-Laki	0,677	0,466
	Perempuan	0,323	0,534
Kadar Kolesterol	Ya	0,654	0,278
	Tidak	0,346	0,722
Tekanan Darah	Rendah	0,308	0,346
	Normal	0,226	0,398
	Tinggi	0,466	0,256
Riwayat Penyakit Jantung Orang Tua	Ya	0,571	0,323
	Tidak	0,429	0,677
Diabetes Melitus	Ya	0,586	0,361
	Tidak	0,414	0,639
Obesitas	Ya	0,504	0,376
	Tidak	0,496	0,624
Dislipidemia	Sangat Rendah	0,038	0,398
	Rendah	0,120	0,226
	Sedang	0,233	0,188
	Tinggi	0,301	0,120
	Sangat Tinggi	0,308	0,068
Penyakit Jantung (Kelas)		0,500	0,500

Untuk atribut yang bertipe data kuantitatif digunakan rumus densitas *Gauss* dapat dilihat pada Persamaan (8).

Probabilitas atribut umur untuk kelas ya, dihitung sebagai berikut.

$$P(\text{Umur} = 47 | \text{Ya}) = \frac{1}{10,6\sqrt{2\pi}} e^{-\frac{(47-55,5)^2}{2 \times 112,7}}$$

$$P(\text{Umur} = 47 | \text{Ya}) = \frac{1}{10,6\sqrt{2\pi}} e^{-\frac{72,441}{225,442}}$$

$$P(\text{Umur} = 47 | \text{Ya}) = \frac{1}{10,6\sqrt{2\pi}} e^{0,321}$$

$$P(\text{Umur} = 47 | \text{Ya}) = 0,052$$

Jadi, peluang pasien yang berumur 47 tahun terkena penyakit jantung sebesar 0,052.

Probabilitas atribut umur untuk kelas tidak, dihitung sebagai berikut.

$$P(\text{Umur} = 47 | \text{Tidak}) = \frac{1}{16,094\sqrt{2\pi}} e^{-\frac{(47-31,5)^2}{2 \times 259,008}}$$

$$P(\text{Umur} = 47 | \text{Tidak}) = \frac{1}{16,094\sqrt{2\pi}} e^{-\frac{238,969}{518,015}}$$

$$P(\text{Umur} = 47 | \text{Tidak}) = \frac{1}{16,094\sqrt{2\pi}} e^{0,461}$$

$$P(\text{Umur} = 47 | \text{Tidak}) = 0,039$$

Jadi, peluang pasien yang berumur 47 tahun tidak terkena penyakit jantung sebesar 0,039. Setelah mengetahui probabilitas setiap atribut pada setiap kelasnya maka selanjutnya ke tahap tiga adalah menentukan probabilitas akhir. Perhitungan probabilitas akhir setiap kelas menggunakan Persamaan (7) sebagai berikut:

$$P(\text{Ya} | F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8) = P(\text{Ya}) \times$$

$$\prod_{i=1}^8 P(F_i | \text{Ya})$$

$$= P(\text{Ya}) \times P(F_1 | \text{Ya}) \times P(F_2 | \text{Ya}) \times P(F_3 | \text{Ya}) \times P(F_4 | \text{Ya}) \times P(F_5 | \text{Ya}) \times P(F_6 | \text{Ya}) \times P(F_7 | \text{Ya}) \times P(F_8 | \text{Ya})$$

$$= P(\text{Ya}) \times P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Ya}) \times P(\text{Umur} = 47 | \text{Ya}) \times P(\text{Kadar Kolesterol} = \text{Tidak} | \text{Ya}) \times P(\text{Tekanan Darah} =$$

$$\text{Prehipertensi} | \text{Ya}) \times P(\text{Riwayat} = \text{Ya} | \text{Ya}) \times P(\text{Diabetes Melitus} = \text{Ya} | \text{Ya}) \times P(\text{Obesitas} = \text{Ya} | \text{Ya}) \times P(\text{Dislipidemia} = \text{Tinggi} | \text{Ya})$$

$$= 0,5 \times 0,677 \times 0,052 \times 0,346 \times 0,226 \times 0,571 \times 0,586 \times 0,504 \times 0,301$$

$$= 0,0000696909598444192$$

$$P(\text{Tidak} | F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8) = P(\text{Tidak})$$

$$\times \prod_{i=1}^8 P(F_i | \text{Tidak})$$

$$= P(\text{Tidak}) \times P(F_1 | \text{Tidak}) \times P(F_2 | \text{Tidak}) \times P(F_3 | \text{Tidak}) \times P(F_4 | \text{Tidak}) \times P(F_5 | \text{Tidak}) \times P(F_6 | \text{Tidak}) \times P(F_7 | \text{Tidak}) \times P(F_8 | \text{Tidak})$$

$$= P(\text{Tidak}) \times P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Tidak}) \times P(\text{Umur} = 47 | \text{Tidak}) \times P(\text{Kadar Kolesterol} = \text{Tidak} | \text{Tidak}) \times P(\text{Tekanan Darah} =$$

$$\text{Prehipertensi} | \text{Tidak}) \times P(\text{Riwayat} = \text{Ya} | \text{Tidak}) \times P(\text{Diabetes Melitus} = \text{Ya} | \text{Tidak}) \times P(\text{Obesitas} = \text{Ya} | \text{Tidak}) \times P(\text{Dislipidemia} = \text{Tinggi} | \text{Tidak})$$

$$= 0,5 \times 0,466 \times 0,039 \times 0,722 \times 0,398 \times 0,323 \times 0,361 \times 0,376 \times 0,120$$

$$= 0,0000137978556351798$$

Berdasarkan perhitungan di atas dapat diketahui bahwa kelas yang mempunyai nilai probabilitas terbesar adalah kelas ya, sehingga dapat disimpulkan contoh kasus diprediksi masuk dalam kelas ya artinya pasien tersebut memiliki potensi terkena penyakit jantung, disarankan untuk segera cek kesehatan jantung di rumah sakit.

Uji klasifikasi dengan Menggunakan Program Delphi

Langkah pertama dalam pembuatan aplikasi prediksi menggunakan Delphi yaitu membuat tampilan aplikasi. Adapun tampilan yang dibuat yaitu halaman utama (*home*), *database*, prediksi (*Prediction*), biodata (*about*) yang dapat dilihat pada Gambar 9, Gambar 10, Gambar 12, Gambar 14. Pada halaman *database* terdapat *submenu input database* ketika menekan tombol *option* yang dapat dilihat pada Gambar 11. Pada halaman *prediction* terdapat *submenu* petunjuk pengisian ketika menekan tombol petunjuk pengisian yang dapat dilihat pada Gambar 13. Aplikasi ini diberi nama HDP v1.0. HDP v1.0 singkatan dari *Heart Disease Prediction* versi 1.0



Gambar 9. Tampilan halaman utama (*home*)



Gambar 10. Tampilan halaman *database*



Gambar 11. Tampilan submenu input database



Gambar 12. Tampilan prediksi (prediction)



Gambar 13. Tampilan submenu petunjuk pengisian



Gambar 14. Tampilan biodata (about)

Uji Akurasi Naive Bayes

Dalam melakukan prediksi (klasifikasi) diharapkan dapat melakukan klasifikasi pada semua objek dengan benar. Untuk mendapatkan tingkat akurasi (ketepatan) maka menggunakan Persamaan (9). Percobaan dilakukan sebanyak 2 kali dengan jumlah data *testing* berturut-turut sebesar 25 dan 50. Adapun hasil akurasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Akurasi Klasifikasi

	Data Testing 1	Data Testing 2
Jumlah Data	25	50
Akurasi	80%	78%

Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan. Adapun kesimpulan yang diperoleh dari penelitian mengenai prediksi risiko terkena penyakit jantung sebagai berikut:

1. Metode *naive* Bayes dapat digunakan dalam memprediksi risiko seseorang terkena penyakit jantung.
2. Pengaplikasian metode *naive* Bayes dalam memprediksi risiko seseorang terkena penyakit jantung menggunakan program *Delphi 7 Enterprise* telah dilakukan dengan nama aplikasi HDP v1.0 dan dapat memprediksi risiko seseorang terkena penyakit jantung.
3. Ketepatan hasil prediksi terhadap hasil pengklasifikasian risiko penyakit jantung berdasarkan data yang didapatkan dari RSUD AWS bulan November dan Desember 2016 menggunakan program *Delphi 7 Enterprise* yaitu untuk percobaan 1 dengan jumlah data *testing* sebanyak 25 data didapat tingkat akurasi sebesar 80% dan pada percobaan 2 dengan jumlah data *testing* sebanyak 50 data diperoleh tingkat akurasi sebesar 78%.

Daftar Pustaka

Bustami. (2013). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI : Jurnal Penelitian Teknik Informatika*, 3(2), 129-132

Kemendes RI. (2014). Buletin Jendela Data dan Informasi Kesehatan Penyakit Tidak Menular. Diperoleh dari <http://www.depkes.go.id/resources/download/pusdatin/buletin/buletin-ptm.pdf>. Diakses tanggal 10 September 2016

Kemendes RI. (2014). Info Datin Situasi Kesehatan Jantung. Diperoleh dari <http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin-jantung.pdf>. Diakses tanggal 10 September 2016

- Mursito, B. (2002). *Ramuan Tradisional Untuk Pengobatan Jantung*. Jakarta: PT. Penebar Swadaya
- Prasetyo, E. (2012). *Data Mining: Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: ANDI
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 208-209
- Siahaan, D. (2016). Aplikasi Classification and Regresion Tree (CART) dan Regresi Logistik Ordinal dalam Bidang Ilmu Pendidikan dengan Studi Kasus Predikat Kelulusan Mahasiswa S1 FMIPA Universitas Mulawarman. *Jurnal Exponensial*, 7(1), 95-104
- Yuniastuti, A. (2008). *Gizi dan Kesehatan*. Yogyakarta: Graha Ilmu.

