

KLASIFIKASI DATASET STATLOG PENYAKIT JANTUNG DENGAN METODE NAÏVE BAYES



Oleh :

MUHAMMAD ARBIANSYafa SISWANTO	200411100195
ISMAIL GALANG ADITYAWAN	200411100196
KEVIN AGUNG J MAHENDRA	200411100085

**PROGRAM STUDI
TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA
2022**

KATA PENGANTAR

Alhamdulillah robbil alamin. Segala puji syukur kehadiran Allah SWT yang selalu memberikan berkah, rahmat, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul ” **KLASIFIKASI DATASET STATLOG PENYAKIT JANTUNG DENGAN METODE NAÏVE BAYES**” sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Fakultas Teknik Jurusan Teknik Informatika Program Studi Sistem Informasi. Penulis menyadari bahwa dalam terselesaikannya skripsi ini tidak lepas dari dukungan beberapa pihak sehingga penulis sangat mengucapkan terima kasih yang sebesar- besarnya kepada:

1. Allah SWT yang selalu memberikan kelancaran dan kemudahan dalam menyelesaikan kuliah dan skripsi.
2. Kedua orang tua saya yang tiada henti memberikan kasih sayang, semangat dan doa. Semoga kedua orang tua selalu sehat dan dalam lindungan Allah SWT.
3. Ibu Dr. Fika Hastarita Rachman, S.T selaku dosen pembimbing 1 dan ibu Sri Herawati, S.Kom., M.Kom selaku dosen pembimbing 2 serta ibu Imamah, S.Kom., M.Kom selaku dosen wali yang telah memberikan dukungan, masukan, semangat, serta meluangkan banyak waktu untuk memberikan bimbingan kepada penulis.
4. Bapak dan Ibu dosen Fakultas Teknik, khususnya Dosen Program Studi Sistem Informasi yang telah memberikan banyak ilmu selama perkuliahan.
5. Seluruh teman-teman Teknik Informatika khususnya angkatan 2020 yang turut memberikan dukungan dalam menyelesaikan skripsi. Penulis menyadari dalam penulisan skripsi ini tidak luput dari kekurangan dan kesalahan karena terbatasnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis memohon maaf atas segala kekurangan yang terdapat dalam skripsi. Semoga skripsi ini dapat bermanfaat bagi pembaca khususnya dalam bidang Sistem Informasi dan Teknik Informatika.

Bangkalan, 22 November 2022

ABSTRAK

Penyakit jantung adalah salah satu penyakit yang disebabkan oleh faktor seperti usia, gaya hidup, dan genetika. Terkadang, sulit untuk mengetahui kapan atau apakah itu akan terjadi, jadi para ilmuwan sedang mencari cara untuk memprediksi penyakit jantung. Ada sistem prediksi baru yang menggunakan data dari kasus sebelumnya untuk membantu mengetahui seberapa besar kemungkinan seseorang terkena penyakit jantung.

Penelitian ini menggunakan statlog heart disease dataset dengan sebanyak 270 data. Metodologi untuk penelitian ini dibuat program diagnosa penyakit jantung maka dibuatkan sistem informasi pemrograman PHP dengan metode Naïve Bayes (Naive Bayes classifier) .

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK.....	ii
DAFTAR ISI.....	iii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan dan Manfaat	1
1.4 Batasan Masalah.....	1
Bab II	3
2.1 Kasus	3
2.2 Metode.....	3
2.3 Evaluasi Akurasi	4
2.4 Penelitian Lain.....	4
Bab III	5
3.1 Alur Sistem.....	5
3.2 Skenario Uji Coba	5
Bab IV	6
4.1 Hitung Manual.....	6
4.2 Orange	11
4.3 Web / GUI	14
4.4 Evaluasi	15
BAB V	16
5.1 Menjawab pertanyaan masalah bab. 1.....	16
5.2 Jelaskan skenario mana yg terbaik akurasi tertinggi	16
Daftar pus taka	17

BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk membantu dokter yang kurang berpengalaman, diperlukan suatu sistem yang dapat membantu mereka mengidentifikasi penyakit jantung dengan lebih akurat. Secara umum, pasien disarankan untuk mengikuti sejumlah tes untuk mengidentifikasi penyakitnya memberikan kesimpulan efektif. Data ini memiliki beberapa fitur yang tidak relevan dan redundan, dapat mengakibatkan klasifikasi tidak akurat. Salah satu cara untuk mengurangi dimensi data atribut adalah dengan menggunakan teknik pemilihan fitur.

Seleksi fitur adalah teknik untuk mengurangi jumlah atribut dalam sebuah dataset. Penguatan data digunakan untuk melakukan perhitungan eksposisi untuk atribut yang berpengaruh dalam dataset pada penyakit jantung.

1.2 Rumusan Masalah

Tujuan dilakukannya penelitian ini untuk

1. Metode Naïve Bayes membantu mengklasifikasikan penyakit jantung berdasarkan ciri-cirinya.
2. Menghitung nilai akurasi ketepatan terkena penyakit jantung

1.3 Tujuan dan Manfaat

Rancangan penelitian ini untuk membantu meningkatkan akurasi klasifikasi penyakit jantung dengan menggunakan informasi yang diperoleh dari pemilihan fitur. Ini akan membantu menyederhanakan data dan membuatnya lebih mudah dipahami. Selain itu, manfaat dari penelitian ini antara lain meningkatkan akurasi dalam proses klasifikasi.

1.4 Batasan Masalah

Penelitian ini mencoba mencari tahu faktor apa saja yang berhubungan dengan penyakit jantung, dan mereka menggunakan data dari penelitian sebelumnya untuk melakukan hal ini. Mereka telah menetapkan beberapa batasan untuk melindungi privasi peserta penelitian, dan batasan tersebut termasuk mengecualikan informasi tentang kondisi kesehatan tertentu.

1. Data yang kami gunakan adalah *dataset* statlog tentang penyakit jantung dari ([UCI Machine Learning Repository: Statlog \(Heart\) Data Set](#)) berupa Dataset statlog penyakit jantung.
2. Data yang digunakan dalam penelitian ini meliputi informasi orang yang memiliki penyakit jantung dan orang yang memiliki penyakit jantung. Ukuran sampel adalah 270 data, dan memiliki 13 fitur yang berbeda. Beberapa fiturnya adalah hal seperti usia, jenis kelamin, dan jenis nyeri dada yang dialami seseorang. Untuk setiap atribut, data diberi kode variabel(nama) berdasarkan apakah orang tersebut menderita penyakit jantung atau tidak
3. Untuk implementasinya pada penelitian ini kami menggunakan pemrograman PHP.

Bab II

TINJAUAN PUSTAKA

2.1 Kasus

Penyakit Jantung disebabkan oleh masalah pada pembuluh darah besar, dapat membuat jantung bermasalah, dan aliran darah ke jantung dapat terhenti. Ada sejumlah tes yang dapat digunakan untuk mendiagnosa penyakit jantung, termasuk 13 tes berbeda yang tercantum di dataset ini.

2.2 Metode

Naïve Bayes Classifier

Klasifikasi Bayesian merupakan klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan kelas. Klasifikasi Bayesian didasarkan pada teorema Bayes. Klasifikasi Bayesian lebih dikenal sebagai klasifikasi Naïve Bayes. Algoritme ini memiliki tingkat akurasi dan kecepatan yang tinggi bila diterapkan pada data yang besar. Naïve bayes berasumsi bahwa pengaruh dari nilai atribut pada kelas yang diberikan adalah saling lepas dengan nilai-nilai atribut lainnya. Hal ini dilakukan untuk menyederhanakan perhitungan yang terlibat dan dalam pengertian ini dianggap “naive”. (Han, 2012). Teorema Bayes menyediakan cara menghitung probabilitas posterior $P(c|e)$ dari $P(c)$, $P(e)$ dan $P(e|c)$ yang ditunjukkan pada persamaan 2.4 dan 2.5. (Rahangdale, Ahirwar & Motwani, 2016).

$$P(c|e) = \frac{P(e|c).P(c)}{P(e)} \quad (2.4)$$

$$P(c|e) = P(e_1|c) * P(e_2|c) * ... * P(e_n|c) * P(c) \quad (2.5)$$

Di mana:

$P(c|e)$ = Probabilitas posterior (c merupakan kelas dan e merupakan atribut atau event)

$P(c)$ = Probabilitas prior dari kelas

$P(e|c)$ = Probabilitas likelihood

$P(e)$ = Probabilitas prior dari *predictor (event)*

Persamaan 2.4 dan 2.5 digunakan untuk mewakili data yang bersifat memiliki keterbatasan yaitu tidak dapat membuat model untuk data yang bergantung pada variabel. Sementara itu, kelebihan dari algoritma ini antara lain mudah digunakan dan seringkali dapat memberikan hasil yang baik untuk banyak kasus.

2.3 Evaluasi Akurasi

Pemilihan fitur adalah teknik yang digunakan untuk menghilangkan data yang tidak relevan atau filter dari sekumpulan data, agar model lebih akurat dan lebih cepat untuk dikerjakan.

Fitur yang membedakan signifikan dalam expositions pembelajaran menunjukkan bahwa fitur tersebut penting dan memiliki kualitas yang tinggi. Seleksi fitur memiliki kelebihan, yaitu membuat ukuran dataset lebih kecil, mengurangi ruang pencarian dan kebutuhan komputasi lebih rendah.

2.4 Penelitian Lain

No	Judul	Obyek	Metode	Hasil
1.	<i>Heart Disease Classification using Nearest Neighbor Classifier with Feature Subset Selection.</i> (Jabbar, Deekshatulu & Chandra, 2013).	Menggunakan <i>dataset Statlog (Heart)</i> yang diperoleh dari UCI <i>Machine Learning Repository</i> dengan jumlah data sebanyak 270	<i>Nearest Neighbor</i> dengan <i>Feature Subset Selection</i> .	Studi ini menemukan bahwa Metode Nearest Neighbor dengan Feature Selection dapat membantu dokter membuat keputusan tentang penyakit jantung lebih cepat dan efisien, dengan mengurangi jumlah tes yang perlu dilakukan.

Bab III

METODOLOGI PENELITIAN

3.1 Alur Sistem

Naïve Bayes merupakan metode yang diusulkan untuk memperbaiki kelemahan teknik data mining, sedangkan metode pemilihan fitur dengan Information Gain digunakan untuk meningkatkan akurasi metode tersebut. Perhitungan Information Gain dilakukan dengan menghitung entropy terlebih dahulu kemudian menghitung nilai gain dari setiap atribut untuk dapat mengetahui atribut yang relevan dengan dataset penyakit jantung. Atribut yang terpilih akan digunakan untuk proses klasifikasi menggunakan metode Naïve Bayes.

$$akurasi = \frac{jumlah\ klasifikasi\ benar}{jumlah\ data\ uji} \times 100\% \quad (2.6)$$

Keakuratan metode klasifikasi adalah seberapa baik ia memprediksi kelas dari kumpulan data uji. Prediksi ini didasarkan pada algoritme pengklasifikasi dan akurat untuk kelas kumpulan data aktual.

3.2 Skenario Uji Coba

Berikut ini adalah data

Langkah	Parameter
1	Hitung nilai <i>entropy</i>
2	Hitung nilai dari masing-masing atribut
3	Mengurutkan atribut berdasarkan nilai tertinggi. Ini akan membantu Anda memutuskan berapa banyak atribut yang ingin Anda gunakan.
4	Data dapat digunakan untuk membantu mengklasifikasikan berbagai hal.
5	Untuk menemukan koordinat tetangga terdekat dari suatu objek, kami menggunakan atribut numeriknya.
6	Untuk membangun model karakteristik seseorang menggunakan Naïve Bayes, Anda dapat menggunakan data pelatihan dari pendekatan K pada langkah Pertama

Bab IV

IMPLEMENTASI

4.1 Hitung Manual

Ada beberapa proses preprocessing untuk proses penghitungan klasifikasi dengan metode Naive Bayes

1. Data dengan banyak kriteria dijadikan range (pendiskritan data) pada atribut RBP, SC, MHRA dan Oldpeak.
2. Pengurangan data untuk sampel. Data yang digunakan untuk sampel 100 baris. Data yang dipilih adalah dengan cara sorting dari data yang paling atas.

Kalkulasi "Class Prior Probabilities, $P(C_i)$ "

$$P(\text{Group} = 1) = 0.57$$

$$P(\text{Group} = 2) = 0.43$$

Kalkulasi "Conditional Probabilities $P(x_k|C_i)$ "

Untuk menentukan hasil dari Conditional Probabilities adalah dengan menggunakan rumus $P(x_k|C_i)$, nah sebelum melakukan menghitung dengan Conditional Probabilities data yang digunakan harus melawati tahap preprocessing, Ketika sudah data akan di hitung manual, berikut cara menghitung datanya

Ketika di excel kita bisa memfilter data yang kita cari tersebut untuk mengetahui berapa banyak data yang ada, nah setelah mengetahui kita membagikan jumlah data dengan nilai Group yang sudah kita cari pada Kalkulasi Class Prior Probabilities.

$P(\text{Sex} = 0 \mid \text{Group} = 1) =$	0.46
$P(\text{Sex} = 0 \mid \text{Group} = 2) =$	0.14
$P(\text{Sex} = 1 \mid \text{Group} = 1) =$	0.54
$P(\text{Sex} = 1 \mid \text{Group} = 2) =$	0.86

$P(\text{Chest Pain Type} = 1 \mid \text{Group} = 1) =$	0.11
$P(\text{Chest Pain Type} = 1 \mid \text{Group} = 2) =$	0.05
$P(\text{Chest Pain Type} = 2 \mid \text{Group} = 1) =$	0.25

P (Chest Pain Type = 2 Group = 2) =	0.09
P (Chest Pain Type = 3 Group = 1) =	0.33
P (Chest Pain Type = 3 Group = 2) =	0.14
P (Chest Pain Type = 4 Group = 1) =	0.32
P (Chest Pain Type = 4 Group = 2) =	0.72

P Fasting Blood Sugar >120 mg/dl = 0 Group = 1) =	0.82
P Fasting Blood Sugar >120 mg/dl = 0 Group = 2) =	0.88
P Fasting Blood Sugar >120 mg/dl = 1 Group = 1) =	0.18
P Fasting Blood Sugar >120 mg/dl = 1 Group = 2) =	0.12

Resting Electrocardiographic Results = 0 Group = 1) =	0.54
Resting Electrocardiographic Results = 0 Group = 2) =	0.35
Resting Electrocardiographic Results = 1 Group = 1) =	0.02
Resting Electrocardiographic Results = 1 Group = 2) =	0.58
Resting Electrocardiographic Results = 2 Group = 1) =	0.44
Resting Electrocardiographic Results = 2 Group = 2) =	0.65

Exercise Induced Angina = 0 Group = 1) =	0.77
Exercise Induced Angina = 0 Group = 2) =	0.49
Exercise Induced Angina = 1 Group = 1) =	0.23
Exercise Induced Angina = 1 Group = 2) =	0.51

The Slope of The Peak Exercise ST Segment = 1 Group = 1) =	0.68
The Slope of The Peak Exercise ST Segment = 1 Group = 2) =	0.30
The Slope of The Peak Exercise ST Segment = 2 Group = 1) =	0.28
The Slope of The Peak Exercise ST Segment = 2 Group = 2) =	0.65
The Slope of The Peak Exercise ST Segment = 3 Group = 1) =	0.04
The Slope of The Peak Exercise ST Segment = 3 Group = 2) =	0.05

Number of Major Vessels (0-3) Colored by Flourosopy = 0 Group = 1) =	0.79
--	------

Number of Major Vessels (0-3) Colored by Flourosopy = 0 Group = 2) =	0.33
Number of Major Vessels (0-3) Colored by Flourosopy = 1 Group = 1) =	0.16
Number of Major Vessels (0-3) Colored by Flourosopy = 1 Group = 2) =	0.37
Number of Major Vessels (0-3) Colored by Flourosopy = 2 Group = 1) =	0.05
Number of Major Vessels (0-3) Colored by Flourosopy = 2 Group = 2) =	0.16
Number of Major Vessels (0-3) Colored by Flourosopy = 3 Group = 1) =	0.00
Number of Major Vessels (0-3) Colored by Flourosopy = 3 Group = 2) =	0.14

Age : 35-40 Group =1) =	0.05
Age : 35-40 Group =2) =	0.05
Age : 41-50 Group =1) =	0.32
Age : 41-50 Group =2) =	0.28
Age : 51-60 Group =1) =	0.32
Age : 51-60 Group =2) =	0.42
Age : 61-70 Group =1) =	0.25
Age : 61-70 Group =2) =	0.26
Age : 71-80 Group =1) =	0.07
Age : 71-80 Group =2) =	0.00

Resting Blood Pressure : 90 - 110 Group = 1) =	0.14
Resting Blood Pressure : 90 - 110 Group = 2) =	0.14
Resting Blood Pressure : 111 - 140 Group = 1) =	0.72
Resting Blood Pressure : 111 - 140 Group = 2) =	0.56
Resting Blood Pressure : 141 - 160 Group = 1) =	0.12
Resting Blood Pressure : 141 - 160 Group = 2) =	0.26
Resting Blood Pressure : 161 - 180 Group = 1) =	0.02
Resting Blood Pressure : 161 - 180 Group = 2) =	0.05

Serum Cholestoral (mg/dl) : 121 - 220 Group = 1) =	0.39
Serum Cholestoral (mg/dl) : 121 - 220 Group = 2) =	0.26
Serum Cholestoral (mg/dl) : 221 - 320 Group = 1) =	0.56
Serum Cholestoral (mg/dl) : 221 - 320 Group = 2) =	0.65

Serum Cholestoral (mg/dl) : 321 - 420 Group = 1) =	0.04
Serum Cholestoral (mg/dl) : 321 - 420 Group = 2) =	0.09
Serum Cholestoral (mg/dl) : 421 - 520 Group = 1) =	0.00
Serum Cholestoral (mg/dl) : 421 - 520 Group = 2) =	0.00
Serum Cholestoral (mg/dl) : 521 - 620 Group = 1) =	0.02
Serum Cholestoral (mg/dl) : 521 - 620 Group = 2) =	0.00

Maximum Heart Rate Achieved : 95 - 115 Group 1) =	0.05
Maximum Heart Rate Achieved : 95 - 115 Group 2) =	0.19
Maximum Heart Rate Achieved : 116 - 135 Group 1) =	0.11
Maximum Heart Rate Achieved : 116 - 135 Group 2) =	0.28
Maximum Heart Rate Achieved : 136-155 Group 1) =	0.26
Maximum Heart Rate Achieved : 136-155 Group 2) =	0.37
Maximum Heart Rate Achieved : 156-175 Group 1) =	0.37
Maximum Heart Rate Achieved : 156-175 Group 2) =	0.12
Maximum Heart Rate Achieved : 176 - 195 Group 1) =	0.21
Maximum Heart Rate Achieved : 176 - 195 Group 2) =	0.05

Oldpeak : 0 - 1 Group 1)=	0.77
Oldpeak : 0 - 1 Group 2)=	0.51
Oldpeak : 1,1 - 2 Group 1)=	0.16
Oldpeak : 1,1 - 2 Group 2)=	0.21
Oldpeak : 2,2 - 3 Group 1)=	0.05
Oldpeak : 2,2 - 3 Group 2)=	0.16
Oldpeak : 3,3 - 4 Group 1)=	0.00
Oldpeak : 3,3 - 4 Group 2)=	0.12
Oldpeak : 4,1 - 5 Group 1)=	0.02
Oldpeak : 4,1 - 5 Group 2)=	0.00

Thal : 3 Group 1)=	0.77
Thal : 3 Group 2)=	0.28
Thal : 6 Group 1)=	0.05
Thal : 6 Group 2)=	0.12
Thal : 7 Group 1)=	0.18
Thal : 7 Group 2)=	0.60

Contoh Sampel untuk diklasifikasikan :

$x = (\text{Age} = 35 | \text{Sex} = 0 | \text{CPT} = 1 | \text{RBP} = 112 | \text{SC} = 149 | \text{FBS} = 0 | \text{RER} = 0 | \text{MHRA} = 125 | \text{EIA} = 0 | \text{Oldpeak} = 1 | \text{SOP} = 1 | \text{NMV} = 0 | \text{Thal} = 3)$

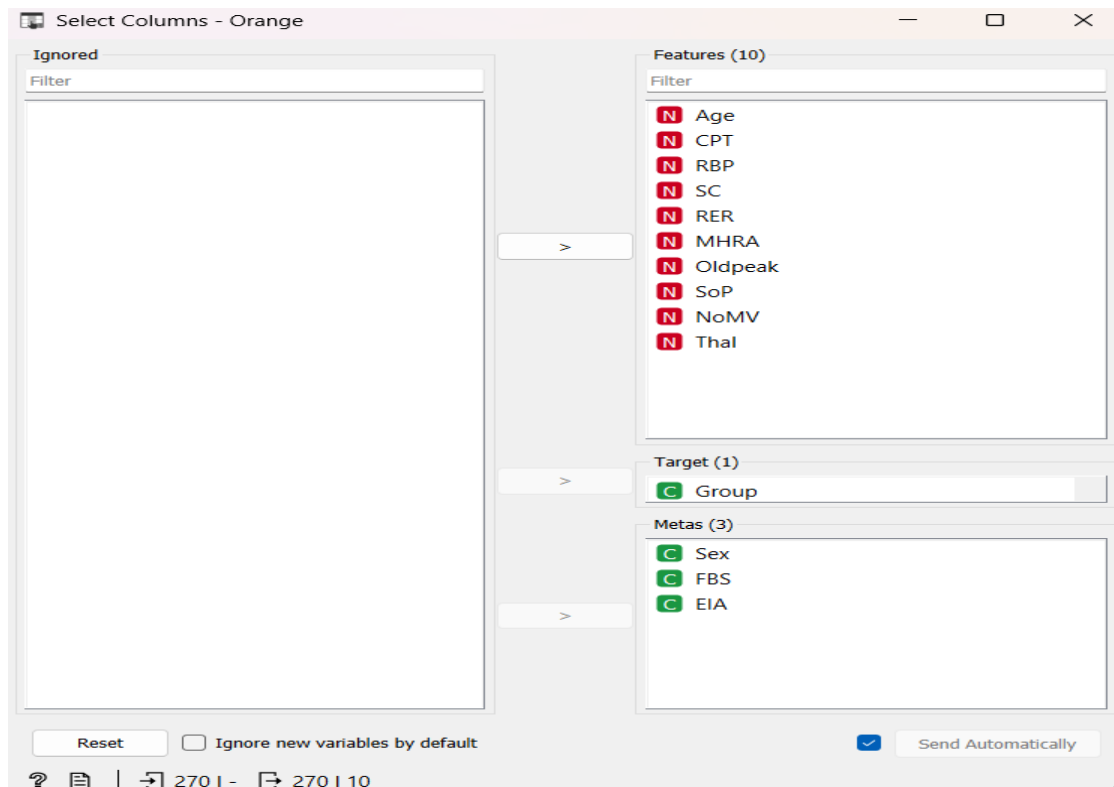
Maka untuk "Conditional probabilitas" =

$$P(X | \text{Group} = 1) = 8.48645\text{E-}06$$

$$P(X | \text{Group} = 2) = 3.04437\text{E-}08$$

4.2 Orange

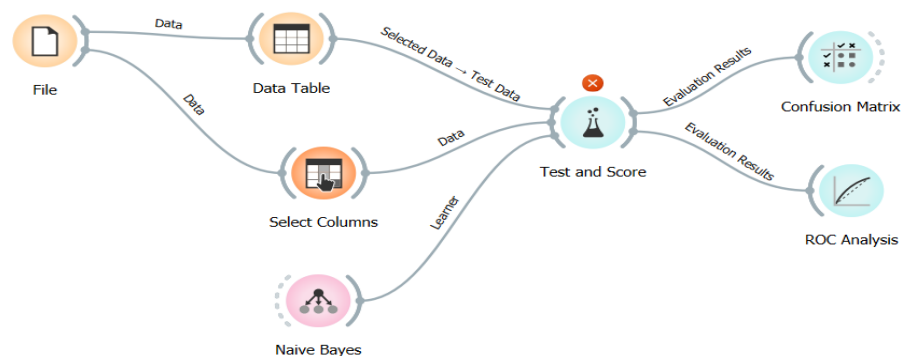
Preprocessing dataset.



Gambar 1. Penentuan data

Proses penentuan data *widget*, pilih kolom atribut yang ditentukan adalah Sex, FBS, EIA dengan *attribute* target adalah Group.

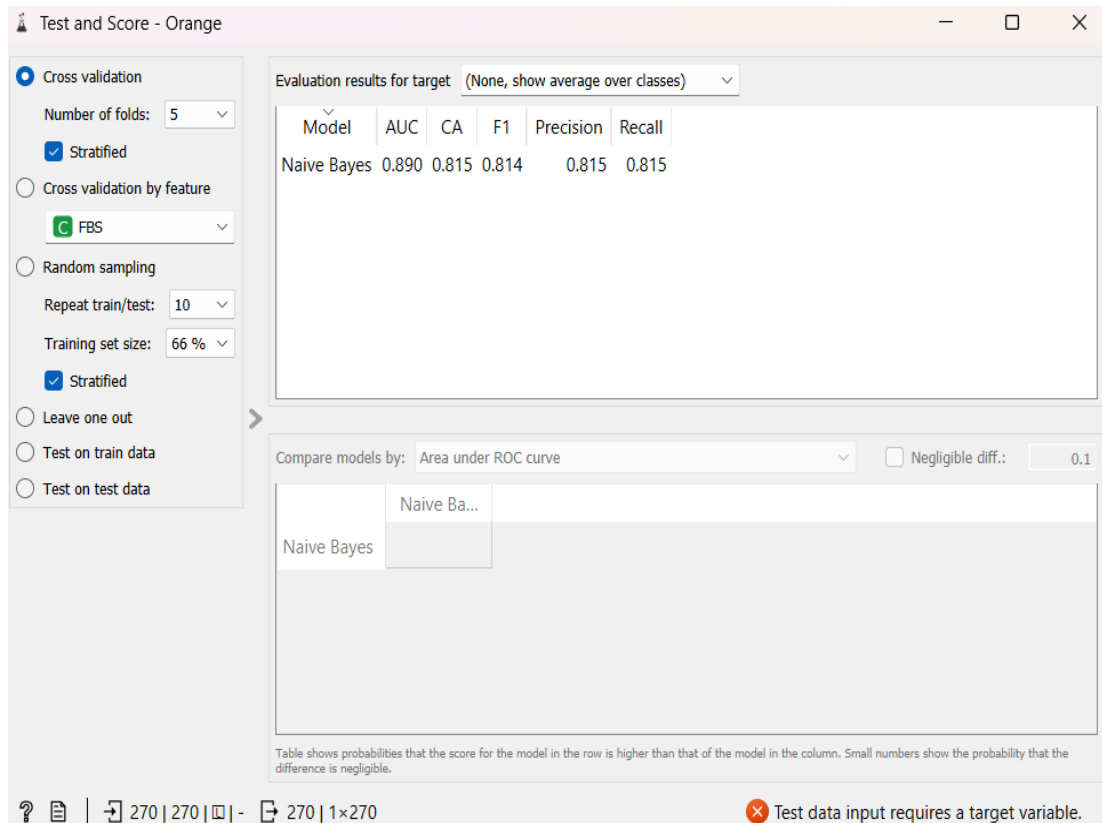
Analisa proses data, dalam performa pada aplikasi *orange*.



Gambar 2. Design widget dataset penyakit jantung.

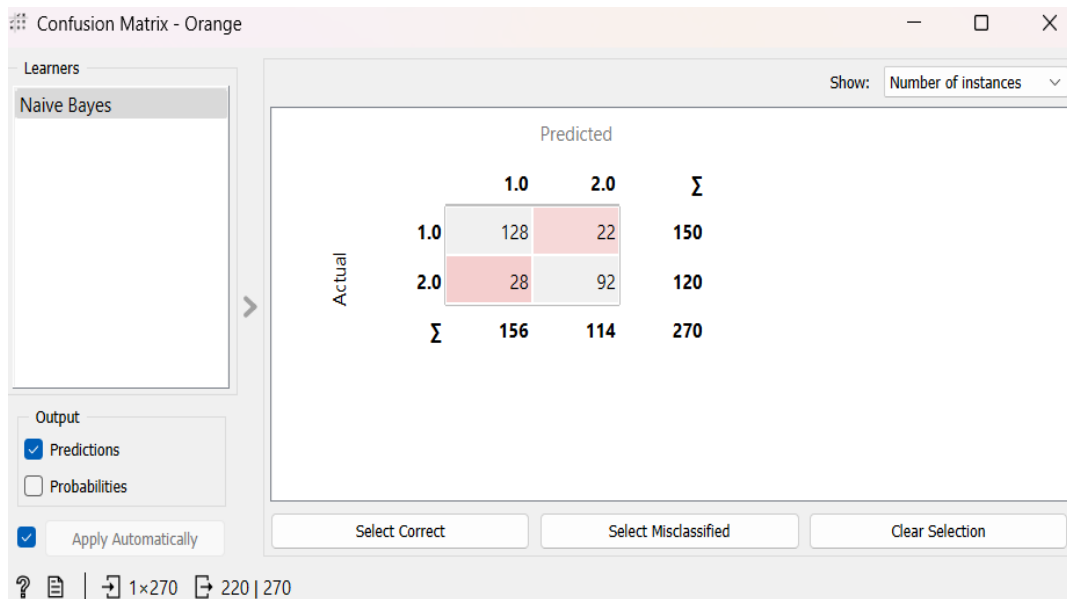
Menjadikan rancangan *widget* dalam model pada *Orange* berupa *Naïve Bayes* yang telah diinputkan *dataset*.

Hasil dari model data uji dengan *atribute1* sebagai target, 3 *atribute numeric* yaitu Sex, FBS, EIA, sehingga diperoleh hasil *test score* seperti terlihat pada gambar 3.

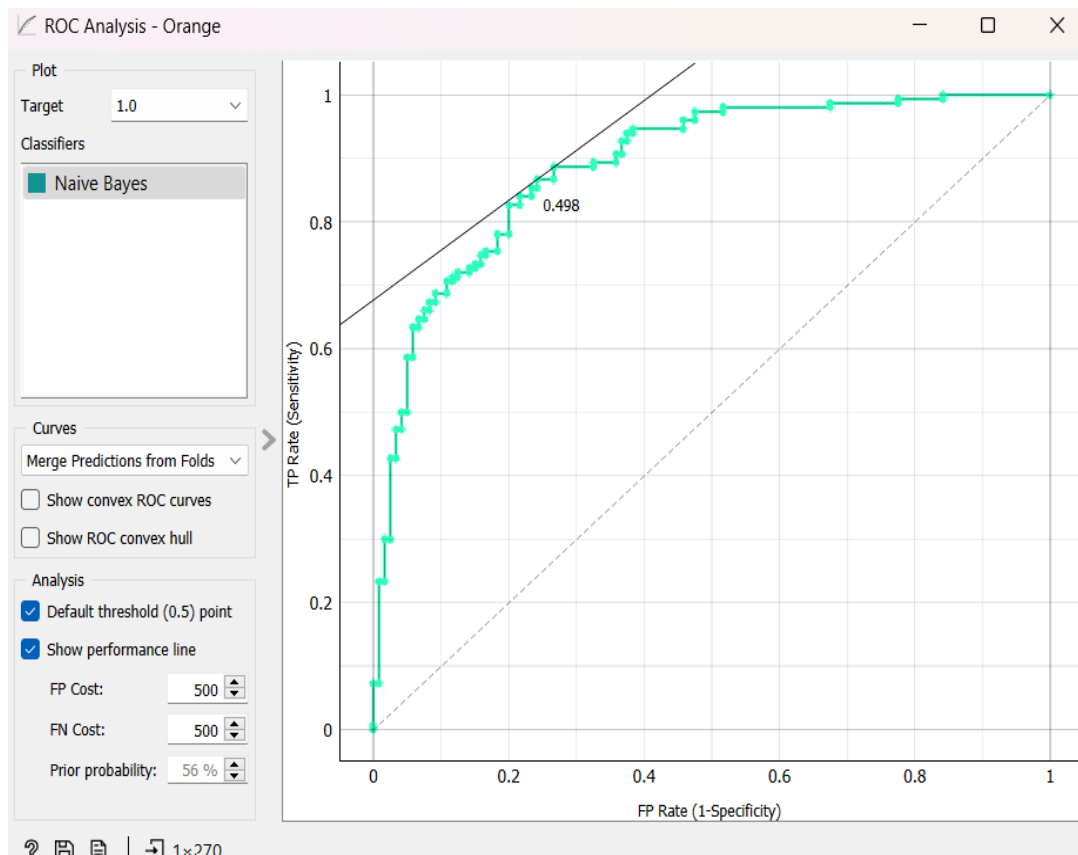


Gambar 3. Hasil widget test and score

Berdasarkan data yang telah diuji dengan klasifikasi model Naïve Bayes diperoleh hasil perhitungan Precision 0.815, Recall 0.815, Accuracy 0.890.



Gambar 4. Hasil Nilai Confusion Matrix dengan metode Naive Bayes



Gambar 5. Hasil Analisis ROC

4.3 Web / GUI

Untuk mencari nilai probabilitas dari dataset penyakit jantung dibuat dengan menggunakan bahasa pemrograman PHP. Aplikasi ini berjalan di localhost. Metode yang digunakan pada aplikasi ini adalah metode gullible bayes, dimana kita harus menentukan terlebih dahulu kemudian menghitung kelasnya, setelah itu kita membuat beberapa kasus untuk dihitung setelah diketahui hasilnya, kita hitung hasilnya, setelah diketahui hasil perhitungan pertama kita masuk ke compositions perhitungan akhir, setelah diketahui hasil akhir dimana hasil akhir perhitungan adalah nilai

Age	Sex	CP	RBP	SC	FBS	RER	MHRA	EIA	Oldpeak	STS	NV	T	Heart Disease
70	1	4	130	322	0	2	109	0	2.4	2	3	3	2
67	0	3	115	564	0	2	160	0	1.6	2	0	7	1
57	1	2	124	261	0	0	141	0	0.3	1	0	7	2
64	1	4	128	263	0	0	105	1	0.2	2	1	7	1
74	0	2	120	269	0	2	121	1	0.2	1	1	3	1
65	1	4	120	177	0	0	140	0	0.4	1	0	7	1
56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
59	1	4	110	239	0	2	142	1	1.2	2	1	7	2
60	1	4	140	293	0	2	170	0	1.2	2	2	7	2
63	0	4	150	407	0	2	154	0	4	2	3	7	2
59	1	4	135	234	0	0	161	0	0.5	2	0	7	1
53	1	4	142	226	0	2	111	1	0	1	0	7	1
44	1	3	140	235	0	2	180	0	0	1	0	3	1
61	1	1	134	234	0	0	145	0	2.6	2	2	3	2
57	0	4	128	303	0	2	159	0	0	1	1	3	1

Input Nilai

Age

Sex

Chest Pain Type

Resting Blood Pressure

Serum Cholesterol in mg/dl

Fasting Blood Sugar > 120 mg/dl

Resting Electrocardiographic Results (values 0,1,2)

Maximum Heart Rate Achieved

Exercise Induced Angina

Oldpeak

ST Segment

Number of Major Vessels

Thal

Gambar di atas adalah tampilan untuk mencari probabilitas dan pada side bar kanan berfungsi untuk menginputkan nilai

The screenshot shows a web browser window with the address bar displaying 'localhost / 127.0.0.1 | phpMyAdmin x NB Proses'. The page title is 'Proses Naive Bayes'. Below the title, there is a section 'Hasil Input untuk Diklasifikasi' containing a table with 13 columns: A, S, CP, RBP, SC, FBS, RER, MHRA, EIA, O, STS, NV, and T. The table has one row of data with values: 67, 1, 2, 115, 261, 1, 2, 109, 1, 4, 2, 3, 3. Below the table, the text 'Maka, Conditional Probabilities =' is followed by two lines of probability values: $P(X | \text{Group} = 1) = 0$ and $P(X | \text{Group} = 2) = 1.2651003435351E-7$. Below this, a line of text states: 'Jadi, Conditional Probabilities yang terpilih adalah Group 2 / Presence'. At the bottom of the page, there is a blue button labeled 'Kembali'.

A	S	CP	RBP	SC	FBS	RER	MHRA	EIA	O	STS	NV	T
67	1	2	115	261	1	2	109	1	4	2	3	3

Maka, Conditional Probabilities =

$P(X | \text{Group} = 1) = 0$

$P(X | \text{Group} = 2) = 1.2651003435351E-7$

Jadi, Conditional Probabilities yang terpilih adalah Group 2 / Presence

[Kembali](#)

Setelah memasukan nilai maka akan tampil Conditional probabilitas beserta dengan keterangan pengelompokannya

1

4.4 Evaluasi

Kesimpulan hasil dari penelitian dijelaskan dan dilakukan pada sistem klasifikasi penyakit jantung menggunakan naïve bayes. Klasifikasi sistem ini disimpulkan nilai hasil AUC 0.890, hasil nilai Precision 0.815 dan nilai Recall 0.815 dengan konfirmasi data yang terdapat pada UCI Machine Learning yang berisi klasifikasi dan 13 atribut dengan jumlah 270 data.

BAB V

KESIMPULAN

5.1 Menjawab pertanyaan masalah bab. 1

Algoritma Naive Bayes adalah teknik klasifikasi berdasarkan teori Bayes dengan asumsi bahwa setiap predikat independen adalah sama. Dengan individualized organization sederhana, asumsinya adalah bahwa keberadaan fitur di kelas tidak bergantung pada keberadaan fitur lain di kelas yang sama. Secara keliru, saran dapat digunakan jika Anda memiliki tempat tidur yang buruk, kemampuan internet, kamera yang buruk, dll. Terlepas dari kenyataan bahwa fitur khusus ini sedang digunakan, pengguna tetap harus berkontribusi pada pengetahuan bahwa perangkat yang disarankan adalah smartphone.

Metode ini digunakan oleh Naive Bayes Keuntungan dan membutuhkan penggunaan data khusus untuk mengestimasi parameter yang digunakan dalam proses klasifikasi.

5.2 Jelaskan skenario mana yg terbaik akurasi tertinggi

Hasil penelitian ini menunjukkan bahwa menggunakan metode Naive Bayes untuk mengklasifikasi $x = (\text{Age} = 35 | \text{Sex} = 0 | \text{CPT} = 1 | \text{RBP} = 112 | \text{SC} = 149 | \text{FBS} = 0 | \text{RER} = 0 | \text{MHRA} = 125 | \text{EIA} = 0 | \text{Oldpeak} = 1 | \text{SOP} = 1 | \text{NMV} = 0 | \text{Thal} = 3)$, Maka = untuk "Conditional probabilitas" = $P(X | \text{Group} = 1) = 8.48645\text{E-}06$ $P(X | \text{Group} = 2) = 3.04437\text{E-}08$, Akurasi 0.890, hasil nilai Precision 0.815 dan nilai Recall 0.815 dengan konfirmasi data yang terdapat pada UCI Machine Learning yang berisi klasifikasi dan 13 atribut dengan jumlah 270 data.

Daftar pus taka

([UCI Machine Learning Repository: Statlog \(Heart\) Data Set](#))