**BAB III**

**METODE PENELITIAN**

1. **Data**

Data yang akan digunakan adalah data Penyakit Jantung yang diperoleh dari situs kaggle.com. Dari data penyakit jantung diperoleh 918 data record dan terdapat 11 atribut dan 1 kelas yang digunakan sebagai input dalam perhitungan metode klasifikasi Correlate Naïve Bayes. Atribut-atribut tersebut antara lain yang berhubungan dengan ciri-ciri terkena penyakit jantung seperti kolesterol, gula darah, dan nyeri dada. Atribut yang berhubungan dengan kondisi fisik seorang yang memiliki penyakit jantung diperhatikan berapa angka cek yang dilakukan setiap waktunya sesuai dengan atribut yang digunakan.

|  |  |
| --- | --- |
| Atribut | Keterangan |
| Age | Usia |
| Sex | Gender/jenis kelamin |
| ChestPain Type | Jenis nyeri dada |
| RestingBP | Tekanan darah istirahat, diambil setelah duduk sekitar kurang dari 10 menit |
| Cholesterol | Produksi lemak oleh berbagai sel dalam tubuh |
| FastingBS | Mengukur gula darah setelah puasa selama 8 jam, sering disebut sebagai pemeriksaan gula darah puasa |
| RestingECG | Pemeriksaaan EKG yang dilakukan pada saat pasien dalam kondisi istirahat (dalam kondisi berbaring) |
| MaxHR | Detak jantung maximal |
| ExerciseAngina |  |
| Oldpeak |  |
| ST\_Slope |  |
| HeartDisease | Keterangan penyakit jantung |

Penjelasan mengenai rincian atribut :

Contoh data awal

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Age | Sex | ChestPainType | RestingBP | Cholesterol | FastingBS | RestingECG | MaxHR | ExerciseAngina | Oldpeak | ST\_Slope | HeartDisease |
| 1 | 40 | M | ATA | 140 | 289 | 0 | Normal | 172 | N | 0 | Up | 0 |
| 2 | 49 | F | NAP | 160 | 180 | 0 | Normal | 156 | N | 1 | Flat | 1 |
| 3 | 37 | M | ATA | 130 | 283 | 1 | ST | 98 | N | 0 | Up | 0 |
| 4 | 48 | F | ASY | 138 | 214 | 0 | Normal | 108 | Y | 1.5 | Flat | 1 |
| 5 | 54 | M | NAP | 150 | 195 | 0 | Normal | 122 | N | 0 | Up | 0 |
| 6 | 39 | M | NAP | 120 | 339 | 1 | Normal | 170 | N | 0 | Up | 0 |
| 7 | 45 | F | ATA | 130 | 237 | 0 | Normal | 170 | N | 0 | Up | 0 |
| 8 | 54 | M | ATA | 110 | 208 | 0 | Normal | 142 | N | 0 | Up | 0 |
| 9 | 37 | M | ASY | 140 | 207 | 0 | Normal | 130 | Y | 1.5 | Flat | 1 |
| 10 | 48 | F | ATA | 120 | 284 | 1 | Normal | 120 | N | 0 | Up | 0 |

1. ***Synthethic Minority Oversampling Technique (SMOTE)***

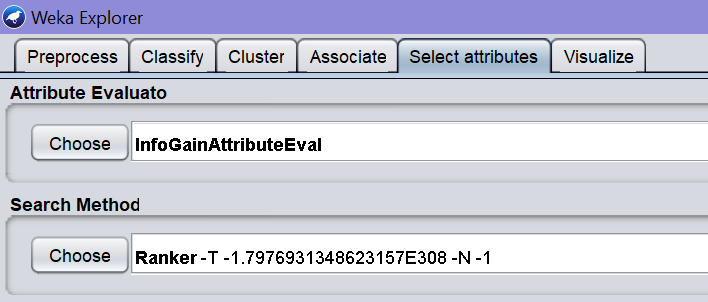
Penelitian ini tidak menggunakan *Synthethic Minority Oversampling Technique* (SMOTE) karena data yang digunakan dalam penelitian sudah *relative* seimbang dimana data yang digunakan bernilai 40% untuk ya dan 60% untuk tidak dimana data ya berjumlah 480 data dan tidak berjumlah 510 data. Hal tersebut membuat penelitian ini tidak menggunakan SMOTE karena data yang digunakan sudah relative seimbang.

1. **Cleaning Data**

Tahap data *cleaning* merupakan proses untuk membersihkan data yang tidak lengkap, mengandung *error*, dan tidak konsisten serta *missing value* atau kondisi adanya data yang hilang atau tidak lengkap. Data yang digunakan pada penelitian ini tidak memiliki *missing value*.

1. **Data Selection**

Tahap data selection merupakan proses pemilihan atribut yang relevan dengan rangking atau urutan bobot pada data menggunakan metode *information gain*. Untuk memilih data atau atribut yang relevan, maka proses seleksi atribut akan menggunakan *Waikoto Environment for Knowledge Analysis* (WEKA) tools versi 3.9.4. Data yang melewati tahap data selection berjumlah 918 dengan 11 atribut dan 1 label, atribut tersebut merupakan atribut dari ciri-ciri terkena penyakit jantung seperti umur, jenis kelamin, tipe nyeri dada, tekanan darah, dan kolesterol. Nama-nama atribut meliputi age, sex, chest pain type, resting bp, cholesterol, fasting bs, resting ecg, max hr, exercise angina, old peak, st slope, *heart disease* sebagai label.



Pemilihan atribut dengan menggunakan weka menghasilkan ranked yang dimana hasil dari ranked pada penelitian ini yang diambil adalah yang nilai ranked nya lebih dari 0.01 sedangkan nilai yang kurang dari 0.01 tidak digunakan. Hasil dari pemeringkatan atribut, berikut ini :

|  |  |  |
| --- | --- | --- |
| No | Ranked | Attributes |
| 1 | 0.29932 | ST\_Slope |
| 2 | 0.22504 | ChestPainType |
| 3 | 0.18997 | ExerciseAngina |
| 4 | 0.16143 | Oldpeak |
| 5 | 0.12748 | MaxHR |
| 6 | 0.08337 | Cholesterol |
| 7 | 0.0696 | Age |
| 8 | 0.06849 | Sex |
| 9 | 0.05488 | FastingBS |
| 10 | 0.01552 | RestingBP |
| 11 | 0.00872 | RestingECG |

1. **Data Transformation**

Tahap ini akan dilakuakn perubahan tipe data pada data yang melekat pada atribut untuk mempermudah proses penambangan data. Berikut merupakan proses dari *transformasi* data yang berjalan:

1. Transformasi atribut Sex

Proses ini melakukan transformasi data *sex,* data yang berkaitan dengan jenis kelamin akan dikelompokkan berdasarkan ketentuan sebagai berikut:

M (Male) : 1

F (Female) : 2

1. Transformasi atribut Chest Pain Type

Proses ini melakukan transformasi data *Chest Pain Type,* data yang berkaitan dengan tipe nyeri dada akan dikelompokkan berdasarkan ketentuan sebagai berikut:

ATA : 1

NAP : 2

ASY : 3

1. Transformasi atribut Exercise Angina

Proses ini melakukan transformasi data *exercise angina,* data akan dikelompokkan berdasarkan ketentuan sebagai berikut:

N : 2

Y : 1

1. Transformasi atribut ST\_Slope

Proses ini melakukan transformasi dari st\_slope, data akan dikelompokkan berdasarkan ketentuan sebagai berikut:

UP : 1

FLAT : 2

1. Normalisasi min-max attribute Age

Proses ini melakukan normalisasi min-max data age, data tersebut terdiri dari usia termuda 36 dan usia tertua 60. Nilai tersebut akan dihitung menggunakan normalisasi min-max berdasarkan rumus sebegai berikut

Age = 40

1. Normalisasi min-max attribute Resting BP

Proses ini melakukan normalisasi min-max data resting bp, data tersebut terdiri dari data terkecil 100 dan data tertinggi 160. Nilai tersebut akan dihitung menggunakan normalisasi min-max berdasarkan rumus sebegai berikut:

Resting BP = 140

1. Normalisasi min-max attribute Cholesterol

Proses ini melakukan normalisasi min-max data cholesterol, data tersebut terdiri dari data terkecil 164 dan data tertinggi 339. Nilai tersebut akan dihitung menggunakan normalisasi min-max berdasarkan rumus sebegai berikut:

Cholesterol = 289

1. Normalisasi min-max attribute Max HR

Proses ini melakukan normalisasi min-max data cholesterol, data tersebut terdiri dari data terkecil 98 dan data tertinggi 172. Nilai tersebut akan dihitung menggunakan normalisasi min-max berdasarkan rumus sebegai berikut:

Max HR = 172

Transformasi data yang dilakukan meliputi perubahan data kategorial menjadi numeric dan menggunakan fungsi min-max untuk memperkecil rentang data antar atribut. Hasil transformasi yang dilakukan pada data tabel:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Age | Sex | ChestPainType | RestingBP | Cholesterol | FastingBS | MaxHR | ExerciseAngina | Oldpeak | ST\_Slope | HeartDisease |
| 1 | 0,176470588 | 1 | 1 | 0,6 | 0,685534591 | 0 | 1 | 2 | 0 | 1 | 2 |
| 2 | 0,705882353 | 2 | 2 | 1 | 0 | 0 | 0,783783784 | 2 | 1 | 2 | 1 |
| 3 | 0 | 1 | 1 | 0,4 | 0,647798742 | 1 | 0 | 2 | 0 | 1 | 2 |
| 4 | 0,647058824 | 2 | 3 | 0,56 | 0,213836478 | 0 | 0,135135135 | 1 | 1.5 | 2 | 1 |
| 5 | 1 | 1 | 2 | 0,8 | 0,094339623 | 0 | 0,324324324 | 2 | 0 | 1 | 2 |
| 6 | 0,117647059 | 1 | 2 | 0,2 | 1 | 1 | 0,972972973 | 2 | 0 | 1 | 2 |
| 7 | 0,470588235 | 2 | 1 | 0,4 | 0,358490566 | 0 | 0,972972973 | 2 | 0 | 1 | 1 |
| 8 | 1 | 1 | 1 | 0 | 0,176100629 | 0 | 0,594594595 | 2 | 0 | 1 | 2 |
| 9 | 0 | 1 | 3 | 0,6 | 0,169811321 | 0 | 0,432432432 | 1 | 1.5 | 2 | 1 |
| 10 | 0,647058824 | 2 | 1 | 0,2 | 0,65408805 | 1 | 0,297297297 | 2 | 0 | 1 | 2 |

Data hasil transformasi

1. ***Modelling Correlated Naïve Bayes***

Tahap pembentukan model, data telah melalui tahap *preprocessing* akan dibentuk modelnya terlebih dengan menggunakan algoritma *Correlated Naïve Bayes*. Sebelum masuk dalam tahap perhitungan, data akan dibagi menjadi beberapa bagian menggunakan *K-fold cross validation*. Data yang digunakan dalam perhitungan menggunakan data hasil transformasi yang akan dibagi menjadi data *training* dan data *testing*. Untuk permodelan metode Correlated Naïve Bayes akan digunakan data pada atribut *maxHR* Dengan data testing yang digunakan adalah data pada baris pertama. Perhitungan dimulai dengan mencari nilai korelasi (R-Square).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MaxHR  (X) | HeartDisease  (Y) |  |  |  |
| 1 | 2 | 1 | 4 | 2 |
| 0,783783784 | 1 | 0,61431702 | 1 | 0,783783784 |
| 0 | 2 | 0 | 4 | 0 |
| 0,135135135 | 1 | 0,018261505 | 1 | 0,135135135 |
| 0,324324324 | 2 | 0,105186267 | 4 | 0,648648648 |
| 0,972972973 | 2 | 0,946676406 | 4 | 1,945945946 |
| 0,972972973 | 1 | 0,946676406 | 1 | 0,972972973 |
| 0,594594595 | 2 | 0,353542732 | 4 | 1,18918919 |
| 0,432432432 | 1 | 0,186997808 | 1 | 0,432432432 |
| 0,297297297 | 2 | 0,088385683 | 4 | 0,594594594 |
| () 5,513513513 | ()  16 | () 4,260043828 | ()  28 | ()  8,702702702 |

1. **Akurasi**
2. **Spesifikasi Sistem**
3. **Desain Interface**