BAB III
METODOLOGI PENELITIAN

Bab ini membahas tentang deskripsi masalah, tahap penelitian, dan metode yang digunakan untuk memprediksi apakah seseorang memiliki penyakit jantung atau tidak.

3.1 Deskripsi Masalah

Penelitian ini bertujuan untuk memprediksi apakah seseorang menderita penyakit jantung atau tidak berdasarkan sejumlah indikasi. Indikasi-indikasi tersebut adalah tekanan darah, kolesterol, kebiasaan merokok, riwayat penyakit *Stroke*, usia, aktivitas fisik, konsumsi alkohol, dan lainnya. Informasi tersebut akan digunakan sebagai basis pengetahuan untuk prediksi.

Pada penelitian ini, prediksi dilakukan dengan menggunakan dua metode, yaitu pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*. Algoritma *Particle Swarm Optimization* (PSO) akan digunakan untuk menentukan bobot optimal dalam pemilihan atribut pada kedua metode. Selanjutnya, hasil prediksi kedua metode akan dibandingkan untuk mengetahui metode mana yang paling akurat.

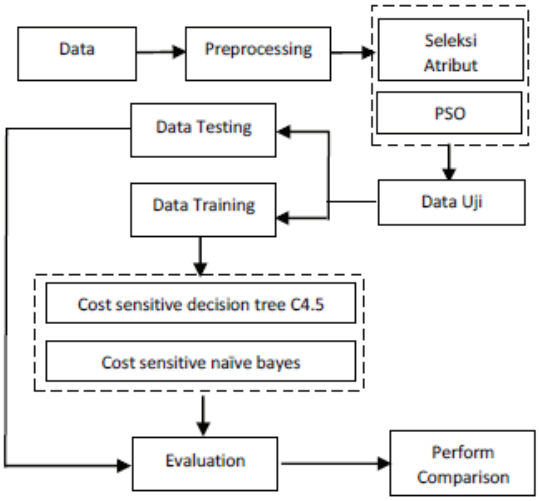
Untuk melakukan prediksi diperlukan kumpulan data. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil melalui website bersumber dari web *Kaggle* (https://www.kaggle.com/). Data tersebut merupakan hasil wawancara lembaga CDC (*Centers of Disease Control and Prevention*) dengan judul "*The underlying uncleaned data comes from the CDC's BRFSS* 2015". Kumpulan data yang telah diolah selanjutnya disebut sebagai *dataset*. Pada *dataset* terdapat indikator-indikator yang memungkinkan seseorang menderita penyakit jantung seperti tekanan darah, kolesterol, kebiasaan merokok, riwayat penyakit *Stroke*, usia, aktivitas fisik, konsumsi alkohol, dan lainnya. Indikator-indikator yang ada pada *dataset* nantinya akan digunakan sebagai atribut yang akan memprediksi seseorang menderita penyakit jantung atau tidak.

Pemrosesan awal *dataset* dilakukan dengan menghapus data yang nilainya berbeda dengan data lainnya (*outlier*) dan mengisi data yang tidak lengkap (*missing value*). *Missing value* dilakukan dengan mengisi nilai rata-rata dari atribut tersebut.

Setiap atribut yang ada, nantinya akan dioptimalkan menggunakan *Particle Sawrm Optimization* (PSO) agar dapat terpilih atribut yang paling berpengaruh.

3.2 Tahapan Penelitian

Pada tahap ini akan dijelaskan secara singkat tahap-tahap penelitian pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes* yang akan dioptimalkan oleh *Particle Swarm Optimization* (PSO). Tahap tersebut digambarkan dalam Gambar 3.1.



Gambar 3. 1 Proses Klasifikasi.

Penjelasan mengenai Gambar 3.1 adalah sebagai berikut.

1. Pengumpulan data.

   Pada tahap ini pengumpulan data dilakukan dengan memilih data yang akan dijadikan sebagai penelitian. Data yang digunakan berjudul "*Heart Disease Health Indicators Dataset"*.

2. *Preprocessing.*

   Setelah mendapatkan data, data tersebut diolah agar dapat menjadi data siap pakai. Proses pengolahan data dilakukan dengan beberapa tahap yaitu pemahaman data, pembersihan data, dan penyeimbangan data.

3. Seleksi atribut.

   Pada tahap ini akan dilakukan penyeleksian atribut yang paling berpengaruh. Penyeleksian atribut dilakukan menggunakan algoritma *Particle Swarm Optimization* (PSO). Algoritma PSO bekerja dengan cara memberikan bobot

setiap atribut yang ada, atribut yang bobotnya terendah akan dieliminasi sehingga menyisakan atribut terpilih.

4. Pembagian Data Uji.

   Data dengan atribut terpilih selanjutnya akan dibagi menjadi 2 jenis data, yaitu data *testing* dan data *training*. Tujuan dari pembagian tersebut ialah untuk menguji performa dari metode yang akan digunakan yaitu pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*. Pada tahap ini data akan dibagi menjadi 30% data *testing* dan 70% data *training*.

5. Data *Training.*

   Pada tahap ini akan dilakukan proses klasifikasi terhadap 70% dari total seluruh data dengan atribut terpilih menggunakan metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*.

6. Data *Testing.*

   Selanjutnya 30% data akan dijadikan data *testing*. Data *testing* merupakan data asli yang tidak melalui proses klasifikasi, tujuannya untuk membandingkan hasil dari proses klasifikasi dan data aslinya.

7. *Evaluation.*

   Setelah data *training* melalui proses klasifikasi menggunakan metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes* maka hasil klasifikasi akan dibandingkan dengan data *testing*. Pada proses ini digunakan tabel *confusion matrix* untuk mendapatkan nilai akurasi, presisi, dan *recall* dari tiap-tiap algoritma.

8. *Perform Comparison.*

   Di tahap ini akan dibandingkan nilai akurasi, presisi, dan juga nilai *recall* dari metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*.

3.3 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data dengan judul "*Heart Disease Health Indicators Dataset"* yang diakses pada tanggal 11 November 2023. Data ini merupakan data wawancara yang dilakukan oleh CDC (*Centers of Disease Control and Prevention*), dengan judul "*The underlying uncleaned data comes from the CDC's BRFSS* 2015". Pada penelitian ini terdapat 2 macam atribut, yaitu atribut independen dan dependen. Atribut Independen

merupakan atribut yang memengaruhi variabel dependen, sedangkan atribut dependen merupakan atribut yang dipengaruhi oleh atribut independen atau sering disebut atribut hasil. Penelitian ini menggunakan 21 atribut independen dan 1 atribut dependen. Atribut yang digunakan dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Atribut-Atribut *Dataset*.

No	Atribut	Jenis	Deskripsi
1.	*Blood pressure (high)*	Independen	Responden memiliki riwayat tekanan darah tinggi
2.	*Cholesterol (high)*	Independen	Responden memiliki riwayat kolesterol tinggi selama lima tahun terakhir
3.	*Cholesterol check*	Independen	Responden melakukan pemeriksaan kolesterol dalam lima tahun terakhir
4.	*BMI*	Independen	*Body Mass Index* (BMI)
5.	*Smoking*	Independen	Pernah merokok setidaknya 100 batang selama hidup
6.	*Stroke*	Independen	Responden memiliki riwayat *Stroke*
7.	*Diabetes*	Independen	Responden memiliki riwayat diabetes
8.	*Physical activity*	Independen	Responden melakukan aktivitas fisik atau olahraga selama 30 hari terakhir
9.	*Fruit*	Independen	Mengonsumsi buah setiap hari
10.	*Vegetables*	Independen	Mengonsumsi sayuran setiap hari
11.	*Alcohol consumption*	Independen	Mengonsumsi minuman beralkohol lebih dari 7 botol perminggu
12	*Health care coverage*	Independen	Responden memiliki asuransi kesehatan
13.	*No doc because cost*	Independen	Apakah selama satu tahun terakhir responden membutuhkan konsultasi ke dokter tetapi terkendala biaya
14.	*General health*	Independen	Kondisi kesehatan responden secara menyeluruh

No	Atribut	Jenis	Deskripsi
15.	*Mental health*	Independen	Kondisi mental responden selama 30 hari terakhir (stress, depresi, atau masalah emosi lainnya).
16.	*Physical health*	Independen	Kondisi kesehatan fisik responden, termasuk penyakit fisik dan cedera yang dialami selama 30 hari terakhir
17.	*Diff walk*	Independen	Responden mengalami kesulitan saat berjalan dan menaiki tangga
18.	*Sex*	Independen	Jenis kelamin responden
19.	*Age*	Independen	Usia responden
20.	*Education*	Independen	Riwayat pendidikan terakhir
21.	*Household income*	Independen	Pendapatan per tahun
22.	*Hearth disease or attack*	Dependen	Responden terdata memiliki penyakit jantung

3.4 Pengolahan Data

Pengolahan data dilakukan menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*). CRISP-DM merupakan metode yang digunakan sebagai proses pengolahan data. Metode ini bertujuan untuk menggali informasi yang terkandung dalam data agar dapat digunakan sebagai data penelitian (Pambudi dkk., 2023). Pada penelitian ini informasi yang ada dalam data adalah indikasi-indikasi yang dapat menyebabkan seseorang terkena penyakit jantung.

Menurut Vanegas dkk (2023), proses data *mining* dengan metode CRISP-DM yang terdiri dari 6 fase berikut.

1. Fase Pemahaman Bisnis (*Business Understanding*).

   Tahapan ini berisi pemahaman terhadap permasalahan yang akan diselesaikan. Penelitian ini fokus pada pendeteksian penyakit jantung dengan menggunakan perbandingan 2 algoritma klasifikasi data mining, yaitu pohon keputusan dan *Naïve Bayes*.

2. Fase Pemahaman Data (*Data Understanding).*

   Fase ini berfokus pada pemahaman data, yaitu mengenal lebih dalam data yang akan digunakan. Proses ini juga melakukan identifikasi kualitas data, data yang berkualitas ditentukan berdasarkan kesesuaian atribut prediksi dengan atribut hasil. Data yang digunakan dalam penelitian ini adalah *dataset* indikator penyakit jantung judul *Heart Disease Health Indicators Dataset* ditulis oleh CDC (*Centers of Disease Control and Prevention*), "*The underlying uncleaned data comes from the CDC's BRFSS* 2015". Data yang ada berisi indikasi-indikasi yang dapat menyebabkan seseorang terkena penyakit jantung. Data tersebut terdiri dari 21 indikasi yang menyebabkan penyakit jantung yang selanjutnya disebut atribut independen dan berisi 1 atribut hasil yaitu menderita penyakit jantung atau tidak. Data berjumlah 253.680 data dengan 23.893 merupakan penderita penyakit jantung dan 229.787 bukan penderita penyakit jantung. Data tersebut cocok digunakan pada penelitian ini karena atribut yang ada berisikan indikasi yang dapat menyebabkan seseorang terkena penyakit jantung.

3. Fase Persiapan data (*Data Preparation).*

   Pada fase ini dilakukan penyiapan data agar data dapat digunakan dengan baik. Pada tahap ini dilakukan data *preprocessing*, yaitu proses pengolahan data melalui beberapa tahap.

   1) *Select data.*

      Proses pemilihan data, semua indikasi yang memungkinkan seseorang terkena penyakit jantung dipilih sebagai atribut. Pada data ini dipilih sebanyak 21 indikasi yang menyebabkan seseorang terkena penyakit jantung.

   2) *Data Cleaning.*

      Pada tahapan ini dilakukan pembersihan data, yaitu data-data yang mempunyai nilai jauh dari rata-rata dilakukan penghapusan.

   3) *Missing value.*

      Missing value dapat diartikan data yang nilainya kosong. Penanganan *missing value* dapat dilakukan dengan 2 cara yaitu dengan mengsisi data tersebut sesuai rata-rata data atau menghapus data tersebut.

4. *Modelling.*

   Fase ini berfokus pada pemilihan dan penerapan metode yang digunakan. Data yang ada merupakan data klasifikasi penyakit, maka penelitian ini menggunakan metode pohon keputusan dengan penerapan algoritma C4.5 dan *Naïve Bayes*. Pengklasifikasian penyakit juga dikombinasikan dengan algoritma PSO. Penggunaan algoritma PSO yaitu dengan memilih atribut-atribut yang ada seusai bobot tiap atributnya. Proses pembobotan atribut PSO, pohon keputusan dengan penerapan algoritma C4.5 dan *Naïve Bayes* menggunakan bantuan *Software Jupyter Notebook* dengan bahasa pemrograman *Python*.

5. Fase Evaluasi.

   Tahap ini akan dilakukan evaluasi terhadap metode yang dipilih apakah sudah sesuai dengan tujuan penelitian. Fase ini juga memastikan proses pengolahan data tidak ada yang terlewat.

6. *Deployment.*

   Pada tahap ini akan dilakukan penyajian pengetahuan data yang diperoleh. Setelah informasi yang ada dalam data sudah diketahui, maka data sudah dapat digunakan untuk proses selanjutnya.

3.4 Algoritma PSO untuk Pembobotan Atribut

Algoritma PSO digunakan dalam mencari atribut optimal untuk memprediksi penyakit jantung. Atribut terpilih nantinya akan dilakukan perhitungan oleh metode pohon keputusan C4.5 dan *Naïve Bayes*. Melalui PSO, data yang akan diproses diberikan bobot untuk mendukung perhitungan. Penetapan bobot ini dilakukan secara acak atau *random* dengan menetapkan rentang nilai antara 0 hingga 1, (Dwiasnati & Devianto, 2019).

Selanjutnya, dilakukan pengelompokan data mengenai jumlah kasus pada setiap atribut dan kategori atribut terhadap kedua kelas, yakni menderita penyakit jantung atau tidak. Tujuannya adalah untuk mempermudah perhitungan nilai kecocokan. Proses pengelompokan tersebut tergambar seperti pada contoh Tabel 3.2.

Tabel 3. 2 Pengelompokan Data Berdasarkan PSO.

Atribut	Kategori	Jumlah data	Bobot	Kelas	
				Jantung	Tidak jantung
Total data		**253.680**		**23.893**	**229.787**
Blood pressure (high)	Tinggi	108.829	0,4	17.928	90.901
	Normal	144.851	0,6	5.965	138.886
Cholesterol (high)	Tinggi	107.591	0,4	16.753	90.838
	Normal	146.089	0,6	7.140	138.949
Cholesterol check	Melakukan pemeriksaan	244.210	0,8	23.622	220.588
	Tidak melakukan pemeriksaan	9.470	0,2	271	9.199
BMI	Underweight (<18,5)	3.127	0,1	331	2.796
	Normal (>18,5 dan <25)	68.953	0,2	4.719	64.234
	Overweight (>25 dan <30)	93.749	0,4	8.714	85.035
	Obese (>30)	87.851	0,3	10.129	77.722
Smoking	Merokok	112.423	0,5	14.801	97.622
	Tidak merokok	141.257	0,5	9.092	132.165
Stroke	Stroke	10.292	0,1	3.937	6.355
	Tidak Stroke	243.388	0,9	19.956	223.432
Diabetes	Diabetes	4.631	0,2	664	3.967
	Tidak diabetes	213.703	0,5	153.351	198.352
	Diabetes (hanya saat hamil)	35.346	0,3	7.878	27.468
Physical activity	Melakukan	191.920	0,8	15.300	176.620

Atribut	Kategori	Jumlah data	Bobot	Kelas	
				Jantung	Tidak jantung
Total data		**253.680**		**23.893**	**229.787**
	Tidak melakukan	61.760	0,2	8.593	53.167
Fruit	Mengonsumsi	160.898	0,7	14.448	146.450
	Tidak mengonsumsi	92.782	0,3	9.445	83.337
Vegetables	Mengonsumsi	205.841	0,8	18.252	187.589
	Tidak mengonsumsi	47.839	0,2	5.641	42.198
Alcohol consumption	Mengonsumsi	14.256	0,1	848	13.408
	Tidak mengonsumsi	239.424	0,9	23.045	216.379
Health care coverage	Memiliki asuransi	241.263	0,95	23.023	218.240
	Tidak memiliki asuransi	12.417	0,05	870	11.547
No doc because cost	Iya	21.354	0,1	2.649	18.705
	Tidak	232.326	0,9	21.244	211.082
General health	Sempurna	45.299	0,1	1.016	44.283
	Sangat baik	89.084	0,4	4.128	84.956
	Baik	75.646	0,3	7.914	67.732
	Cukup baik	31.570	0,15	6.728	24.842
	Kurang baik	12.081	0,05	4.107	7.974
Mental health	Terganggu	78.000	0,4	8.089	69.911
	Tidak terganggu	175.680	0,6	15.804	159.876
Physical health	Terganggu	93.628	0,35	13.343	80.285

Atribut	Kategori	Jumlah data	Bobot	Kelas	
				Jantung	Tidak jantung
Total data		**253.680**		**23.893**	**229.787**
	Tidak terganggu	160.052	0,65	10.550	149.502
Diff walk	Kesulitan	42.675	0,15	9.915	32.760
	Tidak kesulitan	211.005	0,85	13.978	197.027
Sex	Pria	141.974	0,5	10.205	131.769
	Wanita	111.706	0,5	13.688	98.018
Age	Kategori 1 (18-24 tahun)	5.700	0,02	29	5.671
	Kategori 2 (25-29 tahun)	7.598	0,03	54	7.544
	Kategori 3 (30-34 tahun)	11.123	0,04	126	10.997
	Kategori 4 (35-39 tahun)	13.823	0,05	193	13.630
	Kategori 5 (40-44 tahun)	16.157	0,06	351	15.806
	Kategori 6 (45-49 tahun)	19.819	0,11	712	19.107
	Kategori 7 (50-54 tahun)	26.314	0,1	1.425	24.889
	Kategori 8 (55-59 tahun)	30.832	0,12	2.253	28.579
	Kategori 9 (60-64 tahun)	33.244	0,13	3.358	29.886
	Kategori 10 (65-69 tahun)	32.194	0,12	4.193	28.001
	Kategori 11	23.533	0,09	3.947	19.586

Atribut	Kategori	Jumlah data	Bobot	Kelas	
				Jantung	Tidak jantung
Total data		**253.680**		**23.893**	**229.787**
	(70-74 tahun)				
	Kategori 12 (75-79 tahun)	15.980	0,06	3.093	12.887
	Kategori 13 (>80 tahun)	17.363	0,07	4.159	13.204
Education	Tidak bersekolah	174	0,1	29	145
	SMP	4.043	0,1	778	3.265
	Belum lulus SMA	9.478	0,2	1.618	7.860
	Lulus SMA	62.750	0,2	7.467	55.283
	Belum lulus PT	69.910	0,2	6.918	62.992
	Lulus PT	107.325	0,4	7.083	100.242
Household income	< $10,000	9.811	0,05	1.553	8.258
	< $15,000	11.783	0,1	2.197	9.586
	< $20,000	15.994	0,15	2.519	13.475
	< $25,000	20.135	0,05	2.828	17.307
	< $35,000	25.883	0,15	3.161	22.722
	< $50,000	36.470	0,15	3.646	32.824
	< $75,000	43.219	0,15	3.404	39.815
	>$ 75,000	90.385	0,2	4.585	85.800

Setelah pengelompokkan data, selanjutnya akan dilakukan penerapan algoritma PSO dengan langkah-langkah sebagai berikut:

1. Menentukan data.
2. Tetapkan kecepatan awal partikel, posisi awal partikel, bobot untuk setiap partikel, dan jumlah iterasi maksimal. Kecepatan awal partikel ditetapkan sebesar 0. Posisi awal partikel merupakan nilai dari tiap dimensi pada partikel

A_1, A_2, sampai A_n. Bobot setiap partikel ditentukan secara acak dengan memberikan nilai random [0,1].

3. Hitung bobot inersia.

   Pada langkah ini bobot inersia dihitung dengan penentuan nilai ω_{max} dan ω_{min} . Bobot inersia ω dihitung menggunakan persamaan:

$$\omega = \omega_{max} - \left(\frac{\omega_{max}-\omega_{min}}{i \text{ maksimal}}\right) i \,, \tag{3.1}$$

   di mana

   ω : bobot inersia; dan

   i : iterasi ke-*i*.

4. Tentukan P_{best}; yaitu posisi partikel A_1, A_2, dan seterusnya.
5. Hitung nilai *cost* yang ditentukan dengan dengan mencari jumlah nilai bobot terkecil pada setiap data latih PSO. Pada setiap data *training*, hitung jumlah bobot yang di bawah 0,5. Jumlah bobot yang bernilai di bawah 0,5 ditentukan sebagai nilai *cost*.
6. Hitung nilai *fitness* setiap data dengan rumus:

$$fitness = \frac{\text{data } training - cost}{\text{banyak data } training} \,. \tag{3.2}$$

7. Tentukan G_{best} dengan mengambil nilai *fitness* terbaik.
8. *Update* kecepatan setiap partikel menggunakan perasamaan:

$$v_{i,j}^{t+1} = w.v_{i,j}^{t} + c_1 \times r_1(P_{best}{}_{i,j}^{t} - x_{i,j}^{t}) + c_2 \times r_2(G_{best}{}_{i,j}^{t} - x_{i,j}^{t}) \,, \tag{3.3}$$

   di mana

   w : mengontrol pengaruh kecepatan sebelumnya dikecepatan sekarang dengan *range* w = 0,4 – 1,4;

   c_1 dan c_2 : *learning rate* untuk kemampuan individu (kognitif) dan pengaruh hasil;

   $P_{best}{}_{i,j}^{t}$: posisi terbaik partikel *i,j;*

   $G_{best}{}_{i,j}^{t}$: posisi terbaik global *i,j*;

   $x_{i,j}^{t}$: posisi partikel *i* saat iterasi *t*;

   $v_{i,j}^{t}$: kecepatan partikel *i* saat iterasi *t*; dan

   r_1 dan r_2 : bilangan random yang berdistribusi uniformal dalam interval dan 1.

Untuk setiap kecepatan, yang di*update* adalah kecepatan dari setiap dimensinya.

9. U*pdate* posisi setiap partikel menggunakan persamaan:

$$x_{i,j}^{t+1} = x_{i,j}^{t} + v_{i,j}^{t+1} \quad , \tag{3.4}$$

di mana

$x_{i,j}^{t+1}$: posisi baru yang dicari;

$x_{i,j}^{t}$: posisi saat ini; dan

$v_{i,j}^{t+1}$: arah pencarian.

10. Perbarui kembali nilai *fitness* dari setiap data menggunakan Persamaan 3.3.
11. Ulangi langkah 4 hingga langkah 9 sampai sampai jumlah iterasi terpenuhi.
12. Selanjutnya membandingkan *fitness* P_{best} dengan partikel baru, untuk menentukan bobot nilai akhir dari setiap partikel A1, A2, dan seterusnya dengan mencari nilai *fitness* tertinggi pada setiap atribut. Atribut yang memiliki nilai terendah tidak akan dipilih menjadi atribut.

Atribut terpilih selanjutnya akan dijadikan atribut untuk proses klasifikasi pada metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*.

3.5 Metode Pohon Keputusan dengan Algoritma C4.5

Pada tahap ini, dilakukan proses klasifikasi menggunakan metode pohon keputusan dengan algoritma C4.5. Data yang digunakan merupakan data yang atributnya telah dipilih menggunakan PSO. Konsep yang digunakan dalam pohon keputusan adalah membuat pohon keputusan berdasarkan hubungan sebab akibat antara atribut yang ada (Muflikhah dkk., 2018). Algoritma yang digunakan pada tahap ini adalah algoritma C4.5 merupakan salah satu algoritma yang ada pada pohon keputusan. Berikut merupakan langkah-langkah algoritma C4.5 (Nofriansyah, 2014).

1. Persiapkan data.

   Data yang digunakan pada perhitungan ini merupakan data *training*.

2. Hitung nilai *entropy* total.

   Perhitungan *entropy* total dilakukan dengan menghitung kasus yang mengalami penyakit jantung dan tidak mengalami penyakit jantung. Nilai *entropy* total dihitung menggunakan rumus berikut.

$$Entropy(S) = \sum_{i=1}^{n} - p_i \times log_2\ p_i\ , \tag{3.7}$$

di mana

S : total kasus;

n : jumlah partisi S;

p_i : proporsi dari total kasus pada partisi ke-i terhadap S.

3. Hitung nilai *entropy* tiap atribut.

   Setelah menghitung nilai *entropy* total selanjutnya tiap atribut akan dihitung nilai *entropy* nya menggunakan rumus berikut.

$$Entropy(S_i) = \sum_{i=1}^{n} -\ p_i \times log_2\ p_i \tag{3.8}$$

   di mana

   S_i : total kasus pada partisi ke-i;

   n : jumlah partisi S; dan

   p_i : proporsi dari S_i terhadap S.

4. Hitung nilai *gain.*

   Setelah mendapatkan nilai *entropy* total dan nilai *entropy* tiap atribut selanjutnya adalah menentukan nilai *gain* dari tiap atribut. Nilai *gain* diperoleh dengan rumus sebagai berikut.

$$Gain(S, A) = Entropy(S) - \left(\sum_{i=1}^{n} \frac{S_i}{|S|} \times Entropy(S_i) \right)\ , \tag{3.9}$$

   di mana

   S : jumlah kasus;

   A : atribut;

   n : jumlah partisi atribut A;

   S_i : total kasus pada partisi ke-i; dan

   $|S|$: jumlah kasus dalam S.

5. Tentukan akar pertama.

   Untuk memilih akar pertama perlu dilakukan perhitungan nilai *gain* untuk semua atribut. Setelah semua atribut dihitung, atribut yang memiliki nilai *gain* paling tinggi dijadikan akar pertama.

6. Tentukan cabang.

   Setelah mendapatkan atribut sebagai akar pertama, bagi atribut tersebut sesuai kasus yang ada pada tiap atribut. Setelah dibagi ke dalam kasus nantinya cabang tersebut akan mengklasifikasikan dirinya ke dalam kelas apakah cabang tersebut merupakan kelas menderita penyakit jantung atau tidak menderita penyakit jantung.
7. Ulangi Langkah 3 sampai Langkah 6 hingga syarat terpenuhi.

   Saat nilai *gain* dari tiap cabang lebih dari nol, maka proses terus dilanjutkan. Tetapi jika semua nilai *gain* pada tiap cabang sama dengan nol maka proses tersebut berhenti

3.6 Algoritma *Naïve Bayes*

Perhitungan metode *Naive Bayes* dilakukan pada *dataset* yang terdiri dari atribut-atribut hasil pembobotan dengan PSO. Metode *Naïve Bayes* menggunakan konsep peluang untuk melakukan klasifikasi pada tiap-tiap atribut. Adapun langkah-langkah perhitungan *Naïve Bayes* sebagai berikut (Satya dkk, 2018)

1. Persiapkan data.

   Data yang digunakan pada tahap ini adalah data *training.*
2. Tentukan nilai *prior.*

   Penentuan nilai *prior* dilakukan dengan membanding banyak anggota kelas dengan seluruh jumlah data sampel. Pada tahap ini terdiri dari 2 kelas yaitu jumlah yang mengalami penyakit jantung dan tidak mengalami penyakit jantung. Rumus yang digunakan adalah.

$$P(H) = \frac{x}{A}, \tag{3.10}$$

   di mana

   P(H) : nilai *prior*;

   X : jumlah data tiap kelas; dan

   A : jumlah data seluruh kelas.
3. Tentukan nilai *likelihood.*

   Nilai *likelihood* ditentukan dengan mencari nilai peluang dari tiap atribut terhadap kelasnya. Untuk menentukan nilai *likelihood* menggunakan persamaan sebagai berikut.

$$P(E|H) = \frac{F}{B} \ , \quad (3.11)$$

di mana

P(E|H) : nilai *likelihood*;

F : jumlah data fitur tiap kelas; dan

B : jumlah seluruh data tiap kelas.

4. Tentukan nilai *posterior.*

   Penentuan nilai *posterior* dilakukan dengan cara mengalikan kemungkinan atribut dengan kelas. Rumus untuk menghitung nilai *posterior* sebagai berikut.

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)} \ , \quad (3.12)$$

   di mana

   P(H|E) : nilai *posterior*;

   P(H) : nilai *prior*;

   P(E|H) : nilai *likelihood*; dan

   P(E) : peluang total kasus secara keseluruhan.

   Perhitungan nilai *P(E)* dapat menggunakan rumus sebagai berikut.

$$P(E) = \frac{n(A)}{n(S)},$$

   di mana

   n(A) : total pada kelas *A*; dan

   n(S) : total seluruh kasus.

5. Bandingkan hasil nilai *posterior*

   Setelah mendapatkan nilai *posterior* untuk setiap kelas, maka akan dibandingkan hasilnya. Kelas yang memiliki nilai *posterior* tertinggi akan dipilih menjadi hasil klasifikasi.

3.7 *Evaluation*

Proses evaluasi dilakukan untuk menguji keakuratan dan *performance* dari metode yang telah digunakan. Proses ini menggunakan *Cross Validation* dengan 10 uji. Pada *Cross Validation* ini akan dievaluasi kinerja dari algoritma pohon keputusan dan *Naïve Bayes* berbasis PSO di mana nantinya data akan dibagi sebanyak 10 data menjadi 70 % data *training* dan 30% data *testing.* Nilai k diambil sebanyak 10-*fold* sehingga dari sehingga dari 253.680 data, akan terbagi menjadi

10 bagian dengan jumlah data yang sama, sekitar 25.368 data pada setiap bagian. Dari masing-masing 10 bagian tersebut, 177.576 data akan menjadi data *training*.

Data yang telah melalui proses pohon keputusan dan *Naïve Bayes* nantinya akan di cari nilai akurasi, presisi, dan *recall* melalui tabel *confusion matrix* yang ada pada Gambar 2.1.

3.8 *Perform Comparison*

Perform comparison dilakukan dengan membandingkan seluruh hasil akurasi, presisi, dan *recall* dari setiap metode. Metode yang dibandingkan adalah metode pohon keputusan dengan algoritma C4.5 dan juga *Naïve Bayes* yang sudah dioptimalkan dengan PSO. Metode yang memiliki nilai akurasi, presisi, dan *recall* tertinggi merupakan metode yang akurat dalam memprediksi penyakit jantung.

3.9 Contoh Kasus

Pada sub bab ini akan diberikan contoh kasus sederhana dari penggunaan *Particle Swarm Optimization* (PSO), metode pohon keputusan dengan algoritma C4.5, dan *Naïve Bayes*. Pada contoh kasus, digunakan data kecil yang terdiri dari 8 atribut dengan '*Heart Disease or Attack'* sebagai variabel dependen dan 7 atribut lainnya sebagai variabel independen. Data tersebut akan disajikan ke dalam Tabel 3.3.

Tabel 3. 3 Contoh Kasus PSO.

No	A1	A2	A3	A4	A5	A6	A7	*Heart Disease or Attack*
Partikel 1	1	1	1	0	0	10	6	1
Partikel 2	1	1	0	0	0	13	5	0
Partikel 3	0	1	0	0	0	9	6	0
Partikel 4	1	1	1	0	0	9	6	1
Partikel 5	1	1	1	0	1	12	6	1

Keterangan Tabel

A1: *Blood Preassure*
A2: *Cholesterol*
A3: *Smoking*
A4: *Stroke*
A5: *Difficult walk*
A6: *Age*
A7: *Education*

3.9.1 Perhitungan Algoritma PSO

Berikut adalah langkah-langkah pengerjaan PSO:

1. Tentukan banyaknya data, pada percobaan ini data yang digunakan untuk contoh perhitungan berjumlah 5 data. Bobot dari setiap partikel, disajikan pada Tabel 3.4.

Tabel 3. 4 Sampel Data.

	A1	A2	A3	A4	A5	A6	A7
1	0,4	0,4	0,5	0,9	0,85	0,12	0,4
2	0,4	0,4	0,5	0,9	0,85	0,07	0,2
3	0,6	0,4	0,5	0,9	0,85	0,13	0,4
4	0,4	0,4	0,5	0,9	0,85	0,13	0,4
5	0,4	0,4	0,5	0,9	0,15	0,06	0,4

2. Tentukan kecepatan awal, kecepatan awal partikel adalah 0. Posisi awal partikel merupakan nilai – nilai dari tiap dimensi pada partikel A_1, A_2, dan seterusnya. Selanjutnya tentukan jumlah iterasi maksimal yatu 2.
3. Tentukan bobot inersia awal dengan menggunakan bobot inersia ω_{max} dan ω_{min}. Misalkan ditetapkan ω_{max} = 0,9 dan ω_{min} = 0,4. Untuk menentukan ω digunakan persamaan 3.1.

$$\omega = 0{,}9 - \left(\frac{0{,}9 - 0{,}4}{2}\right) 2 = 0{,}4 .$$

   Pada percobaan ini iterasi maksimal yang digunakan adalah 2 iterasi dan untuk nilai dari $c_1 = 0{,}9$, $c_2 = 0{,}5$, $r_1 = 0{,}5$, $r_2 = 0{,}1$.
4. Tentukan P_{best} awal; P_{best} awal adalah posisi awal dari partikel A_1,A_2 dan seterusnya. P_{best} awal dapat dilihat melalui Tabel 3.4.
5. Tentukan nilai *cost* dengan menghitung jumlah bobot yang bernilai di bawah 0,5 pada tiap data *training* Tabel 3.5 berisikan *cost* dari setiap data *training.*

Tabel 3. 5 Nilai *Cost*.

	A1	A2	A3	A4	A5	A6	A7	*cost*
1	0,4	0,4	0,5	0,9	0,85	0,12	0,4	4
2	0,4	0,4	0,5	0,9	0,85	0,07	0,2	4
3	0,6	0,4	0,5	0,9	0,85	0,13	0,4	3
4	0,4	0,4	0,5	0,9	0,85	0,13	0,4	4
5	0,4	0,4	0,5	0,9	0,15	0,06	0,4	5

6. Setelah mendapatkan jumlah *cost* tiap data maka untuk hitung nilai *fitness* menggunakan persamaan 3.2.

$$fitness\ 1 = \frac{5-4}{5} = 0{,}2\ ;$$

sehingga didapat nilai *fitness* untuk setiap data

fitness 2 = 0,2; *fitness* 4 = 0,2; dan

fitness 3 = 0,4; *fitness* 5 = 0.

7. Setelah menghitung nilai *fitness* tiap data, untuk tentukan G_{best} adalah mencari nilai *fitness* terbaik, nilai *fitness* terbaik berada pada data *fitness* 3, dengan begitu Data ke 3 merupakan G_{best} awal dengan nilai 0,4.
8. Selanjutnya, untuk iterasi pertama lakukan *update* kecepatan setiap partikel, untuk setiap partikel kecepatan yang di*update* adalah kecepatan dari tiap dimensinya, sebagai contoh pada partikel A_1 data pertama. *Update* kecepatan menggunakan persamaan 3.3.

   Partikel A_1

$$v_{(1,1)}^{t+1} = 0{,}4 \times 0 + 0{,}9 \times 0{,}5(0{,}4 - 0{,}4) + 0{,}5 \times 0{,}1(0{,}4 - 0{,}4)$$
$$= 0.$$

   Untuk V(1,2) hingga V(5,7) menggunakan cara perhitungan yang sama seperti V(1,1).

Tabel 3. 6 *Update* Kecepatan Iterasi 1.

	A1	A2	A3	A4	A5	A6	A7
1	0	0	-0,005	-0,025	-0,0225	0,014	0
2	0	0	-0,005	-0,025	-0,0225	0,0165	0,01
3	-0,01	0	-0,005	-0,025	-0,0225	0,0135	0
4	0	0	-0,005	-0,025	-0,0225	0,0135	0
5	0	0	-0,005	-0,025	0,0125	0,017	0

9. Selanjutnya *update* posisi setiap partikel menggunakan persamaan 3.4. *Update* posisi iterasi pertama disajikan dalam Tabel 3.7. Sebagai contoh, *update* posisi partikel A_1 pada data pertama.

$$x(1,1) = 0{,}4 + 0 = 0{,}4\ .$$

Tabel 3. 7 *Update* Posisi Iterasi 1.

	A1	A2	A3	A4	A5	A6	A7	*cost*
1	0,4	0,4	0,495	0,875	0,8275	0,134	0,4	5
2	0,4	0,4	0,495	0,875	0,8275	0,0865	0,21	4
3	0,59	0,4	0,495	0,875	0,8275	0,1435	0,4	4
4	0,4	0,4	0,495	0,875	0,8275	0,1435	0,4	5
5	0,4	0,4	0,495	0,875	0,1625	0,077	0,4	6

10. Setelah itu perbarui kembali nilai *fitness* dari setiap data dan didapatkan hasil *fitness* disajikan pada Tabel 3.8.

Tabel 3. 8 *Update* Nilai *Fitness*.

Data	*Cost*	*Fitness*
1	5	0
2	4	0,2
3	4	0,2
4	5	0
5	6	-0,2

Dari hasil perhitungan diatas diketahui bahwa nilai *fitness* tertinggi ada pada data ke 2 dan 3, sehingga didapatkan nilai *fitness* tertinggi ada pada data ke 2 dan 3 di mana dapat dipilih 0,2 sebagai G_{best} awal untuk iterasi selanjutnya.

11. Untuk iterasi kedua lakukan langkah 4 hingga 11 dengan nilai $\omega = 0{,}4$, $c_1 = 0{,}4$, $c_2 = 0{,}2$, $r_1 = 0{,}3$, $r_2 = 0{,}1$.

    Update Kecepatan untuk iterasi ke 2 menggunakan persamaan 3.3. Partikel A_1.

$$v_{(1,1)}^{t+1} = 0{,}4 \times 0 + 0{,}4 \times 0{,}3(0{,}4 - 0{,}4) + 0{,}2 \times 0{,}1(0{,}2 - 0{,}4)$$
$$= 0{,}2 .$$

Untuk V(1,2) hingga V(5,8) menggunakan cara perhitungan yang sama seperti V(1,1). *Update* kecepatan iterasi 2 disajikan pada Tabel 3.9.

Tabel 3. 9 *Update* Kecepatan Iterasi 2.

	A1	A2	A3	A4	A5	A6	A7
1	0,2	0,2	0,203	0,215	0,2135	0,1916	0,2
2	0,2	0,2	0,203	0,215	0,2135	0,1901	0,194
3	0,206	0,2	0,203	0,215	0,2135	0,1919	0,2
4	0,2	0,2	0,203	0,215	0,2135	0,1919	0,2
5	0,2	0,2	0,203	0,215	0,1925	0,1898	0,2

update posisi untuk iterasi ke 2 menggunakan persamaan 3.4. Hasil *update* posisi iterasi 2 disajikan pada Tabel 3.10.

Sebagai contoh, *update* posisi partikel A_1 pada data pertama.

$$x(1,1) = 0{,}4 + 0{,}2 = 0{,}6\,.$$

Tabel 3. 10 *Update* Posisi Iterasi 2.

	A1	A2	A3	A4	A5	A6	A7	Cost
1	0,6	0,6	0,698	1,09	1,041	0,3256	0,6	3
2	0,6	0,6	0,698	1,09	1,041	0,2766	0,404	4
3	0,796	0,6	0,698	1,09	1,041	0,3354	0,6	3
4	0,6	0,6	0,698	1,09	1,041	0,3354	0,6	3
5	0,6	0,6	0,698	1,09	0,355	0,2668	0,6	3

Didapatkan nilai *fitness* untuk partikel baru disajikan pada Tabel 3.11.

Tabel 3. 11*Update* Nilai *Fitness* Iterasi 2.

Data	*Cost*	*Fitness*
1	3	0,4
2	4	0,2
3	3	0,4
4	3	0,4
5	3	0,4

12. Selanjutnya membandingkan *fitness* P_{best} lama dengan partikel baru. Untuk menentukan bobot nilai akhir dari setiap partikel A_1, A_2 dan seterusnya dengan mencari nilai *fitness* tertinggi pada setiap dataset. Nilai *fitness* P_{best} lama dan P_{best} baru disajikan pada Tabel 3.12.

Tabel 3. 12 Perbandingan Nilai *Fitness* Lama dan Baru.

Data	*Fitness* P_{best} lama	*Fitness* P_{best} baru	P_{best} terbaru
1	0	0,4	Iterasi 1 & 2
2	0,2	0,2	Iterasi 1 & 2
3	0,2	0,4	Iterasi 1 & 2
4	0	0,4	Iterasi 1 & 2
5	-0,2	0,4	Iterasi 1 & 2

Dikarenakan nilai *fitness* tertinggi ada pada data 1, 3, 4, dan 5 maka perlu dilihat *history* perhitungan sebelumnya dimana nilai keseluruhan atribut di masing-masing data terbesar ada pada data ke-3 yaitu di iterasi

pertama dan kedua sehingga pada percobaan ini data ke-3 dipilih sebagai acuan untuk penentuan bobot akhir dari setiap atribut. Sehingga dihasilkan bobot akhir masing-masing partikel seperti pada tabel hasil seleksi partikel pada Tabel 3.13.

Tabel 3. 13 Bobot Akhir Tiap Atribut.

Kode	Atribut	Bobot Akhir
A1	*Blood pressure (high)*	0,796
A2	*Cholesterol (high)*	0,6
A3	*Smoking*	0,698
A4	*Stroke*	1,09
A5	*Diff walk*	1,041
A6	*Age*	0,3354
A7	*Education*	0,6

didapatkan 1 dari 7 atribut yang terseleksi yaitu ada pada *age* di mana nilai bobot akhirnya terendah. Pada penelitian ini atribut yang memiliki bobot nilai terendah dapat dihilangkan karena tidak memiliki pengaruh pada akurasi prediksi penyakit jantung. Sehingga didapatkan atribut terpilih yaitu *Blood pressure (high), Cholesterol (high), Smoking, Stroke, Diff walk,* dan *Education*.

3.9.2 Pohon Keputusan dengan Algoritma C4.5

Berdasarkan hasil perhitungan seleksi partikel menggunakan PSO didapatkan atribut yang akan dilakukan proses perhitungan klasifikasi metode pohon keputusan dengan algoritma C4.5 adalah *Blood pressure (high), Cholesterol (high), Smoking, Stroke, Diff walk,* dan *Education*. Data dengan atribut terpilih disajikan dalam Tabel 3.14.

Tabel 3. 14 Sampel Data dengan Atribut Terpilih.

Data	A1	A2	A3	A4	A5	A7	*Heart Disease or Attack*
1	1	1	1	0	0	6	1
2	1	1	0	0	0	5	0
3	0	1	0	0	0	6	0
4	1	1	1	0	0	6	1
5	1	1	1	0	1	6	1

Berikut merupakan langkah perhitungan nilai *Entropy* dan *Gain* pada setiap atribut yang memiliki label *Heart Disease or attack.*

1. Perhitungan nilai *entropy* total.

   Langkah awal Algoritma C4.5 adalah mencari nilai *entropy*. Pertama, tentukan terlebih dahulu nilai *entropy* total dalam kasus menggunakan persamaan 3.7.

   Diketahui:

   total kasus: 5;

   jumlah penderita penyakit jantung: 3; dan

   jumlah yang tidak menderita penyakit jantung: 2.

$$Entropy(S) = (-(\frac{jumlah\ penderita}{total\ kasus} * \log_2 \frac{jumlah\ penderita}{total\ kasus}) + (-(\frac{jumlah\ yang\ tidak\ menderita}{total\ kasus} * \log_2 \frac{jumlah\ yang\ tidak\ menderita}{total\ kasus})$$

$$= (-\tfrac{3}{5} * \log_2(\tfrac{3}{5})) + (-\tfrac{2}{5} * \log_2(\tfrac{2}{5}))$$

$$= 0{,}970951.$$

2. Perhitungan nilai *gain*

   Setelah menghitung semua nilai *entropy*, selanjutnya adalah menghitung nilai *gain* menggunakan persamaan 3.9.

$$Gain(Total, blood\ pressure) = 0{,}970951 - \left(\left(\frac{4}{5} \times 0{,}811278\right) + \left(\frac{1}{5} \times 0\right)\right)$$

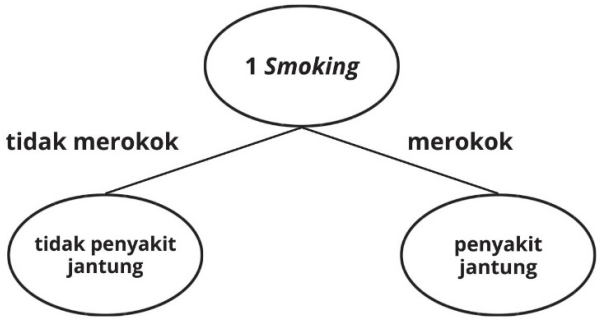
$$= 0{,}321928.$$

   Perhitungan nilai *gain* dilakukan pada setiap atribut yang ada. Hasil perhitungan nilai *gain* disajikan pada Tabel 3.15.

Tabel 3. 15 Perhitungan Nilai *Gain.*

Atribut	Kategori	Total (Nilai)	Total (penderita penyakit jantung)	Total (tidak menderita penyakit jantung)	*Entrophy*	*Gain*
Total data		**5**	**3**	**2**	0,970951	
Blood pressure (high)	Tinggi	4	3	1	0,811278	0,321928
	Normal	1	0	1	0	
Cholesterol (high)	Tinggi	5	3	2	0,970951	0
	Normal	0	0	0	0	
Smoking	Merokok	3	3	0	0	0,970950
	Tidak merokok	2	0	2	0	
Stroke	Stroke	0	0	0	0	0
	Tidak Stroke	5	3	2	0,970951	
Diff walk	Kesulitan	1	1	0	0	0,170950
	Tidak kesulitan	4	2	2	1,000000	
Education	Belum Lulus PT	1	0	1	0	0,321928
	Lulus PT	4	3	1	0,811278	

Dari Tabel 3.15 dapat diketahui bahwa atribut tertinggi dengan *gain* tertinggi adalah *Smoking* yaitu sebesar 0,970950. Dengan demikian *Smoking* menjadi *node* akar. Ada 2 nilai atribut dari *Smoking* yaitu merokok dan tidak merokok. Dari kedua nilai atribut merokok sudah mengklasifisikasikan kasus menjadi 1 keputusan yaitu menderita penyakit jantung, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Begitupun kasus tidak merokok sudah mengklasifikasikan kasus menjadi satu keputusan yaitu tidak menderita penyakit jantung. Atribut merokok menjadi satu aturan *rule* yang terbentuk. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan pada Gambar 3.2.



Gambar 3. 2 Pohon Keputusan.

Hasil pohon keputusan yang terbentuk menghasilkan 1 aturan keputusan dari target yang ingin dicapai yaitu memiliki peyakit jantung atau tidak memiliki penyakit jantung. Adapun *rule* tersebut:

1. Jika merokok maka memiliki penyakit jantung.
2. Jika tidak merokok maka tidak memiliki penyakit jantung.

Aturan keputusan tersebut dibentuk berdasarkan data yang digunakan untuk melakukan perhitungan. Perbedaan jumlah data dan atribut yang digunakan akan menghasilkan pohon keputusan yang berbeda.

3.9.3 *Naïve Bayes*

Berdasarkan hasil perhitungan seleksi partikel menggunakan PSO didapatkan atribut yang akan dilakukan ke proses perhitungan prediksi *Naïve Bayes* yaitu *Blood pressure (high), Cholesterol (high), Smoking, Stroke, Diff walk,* dan *Education.* Selanjutnya pada *Naïve Bayes* dilakukan perhitungan nilai *prior, likelihood* dan *posterior.* Pada Tabel 3.16 disajikan contoh data *testing*.

Tabel 3. 16 Contoh Data *Testing*

Data	A1	A2	A3	A4	A5	A7	*Heart Disease or Attack*
1	1	1	1	0	0	6	1

1. Menghitung nilai *prior.*

   Nilai *prior* didapatkan dari hasil masing-masing data dengan nilai kelas yang sama dibagi dengan total keseluruhan data *training* yang berjumlah 5 data.

P (Jantung) $= \frac{3}{5}$ = 0,6.

P (Tidak Jantung) $= \frac{2}{5}$ = 0,4.

2. Menghitung *likelihood*

Nilai *likelihood* didapatkan dari hasil menghitung jumlah kasus perkelas.

P (BP = Tinggi |J) $= \frac{3}{4}$ = 0,75.

P (BP = Tinggi |TJ) $= \frac{1}{4}$ = 0,25.

P (CL = Tinggi |J) $= \frac{3}{5}$ = 0,6.

P (CL = Tinggi |TJ) $= \frac{2}{5}$ = 0,4.

P (SM = M |J) $= \frac{3}{3}$ = 1.

P (SM = M |TJ) $= \frac{0}{3}$ = 0.

P (STR = Tidak |J) $= \frac{3}{5}$ = 0,6.

P (STR = Tidak |TJ) $= \frac{2}{5}$ = 0,4.

P (DW = Tidak Kesulitan|J) $= \frac{2}{4}$ = 0,5.

P (DW = Tidak Kesulitan|TJ) $= \frac{2}{4}$ = 0,5.

P (EDU = Lulus PT |J) $= \frac{3}{4}$ = 0,75.

P (EDU = Lulus PT|TJ) $= \frac{1}{4}$ = 0,25.

3. Menghitung *Postrior*

Nilai *postrior* didapatkan dari mengalikan kemungkinan atribut masukan dengan kelas, yaitu;

$$P(\text{Penyakit}|\text{Jantung}) = P(\text{BP: Tinggi}|\text{J}) \times P(\text{CL: Tinggi}|\text{J}) \times P(\text{SM: M}|\text{J}) \times P(\text{STR: Tidak}|\text{J}) \times P(\text{DW: TidakKesulitan}|\text{J}) \times P(\text{EDU: Lulus PT}|\text{J})$$

$$= 0{,}75 \times 0{,}6 \times 1 \times 0{,}6 \times 0{,}5 \times 0{,}75$$
$$= 0{,}10125.$$

$$P(\text{Penyakit}|\text{Tidak Jantung}) = P(\text{BP: Tinggi}|\text{TJ}) \times P(\text{CL: Tinggi}|\text{TJ}) \times P(\text{SM: M}|\text{TJ}) \times P(\text{STR: Tidak}|\text{TJ}) \times P(\text{DW: TidakKesulitan}|\text{TJ}) \times P(\text{EDU: Lulus PT}|\text{TJ})$$
$$= 0{,}25 \times 0{,}4 \times 0 \times 0{,}4 \times 0{,}5 \times 0{,}25$$
$$= 0.$$

Selanjutnya dilakukan perhitungan peluang keseluruhan dengan membagi total yang ada pada kelas tertentu dibagi dengan total keseluruhan seperti berikut.

$$P(\text{Penyakit Jantung}) = \frac{4}{5} \times \frac{5}{5} \times \frac{3}{5} \times \frac{5}{5} \times \frac{4}{5} \times \frac{4}{5}$$
$$= 0{,}3072.$$

Dari nilai peluang keseluruan didapatkan prediksi penyakit sebesar

$$\text{Prediksi Penyakit Jantung} = \frac{0{,}10125}{0{,}3072} = 0{,}32959, dan$$

$$\text{Prediksi tidak Penyakit Jantung} = \frac{0}{0{,}3072} = 0.$$

Dari perhitungan di atas, didapatkan hasil nilai peluang prediksi terkena penyakit jantung sebesar 0,32959 sedangkan peluang prediksi tidak terkena penyakit jantung sebesar 0, maka dapat disimpulkan data *testing* pada Tabel 3.16 menderita penyakit jantung.