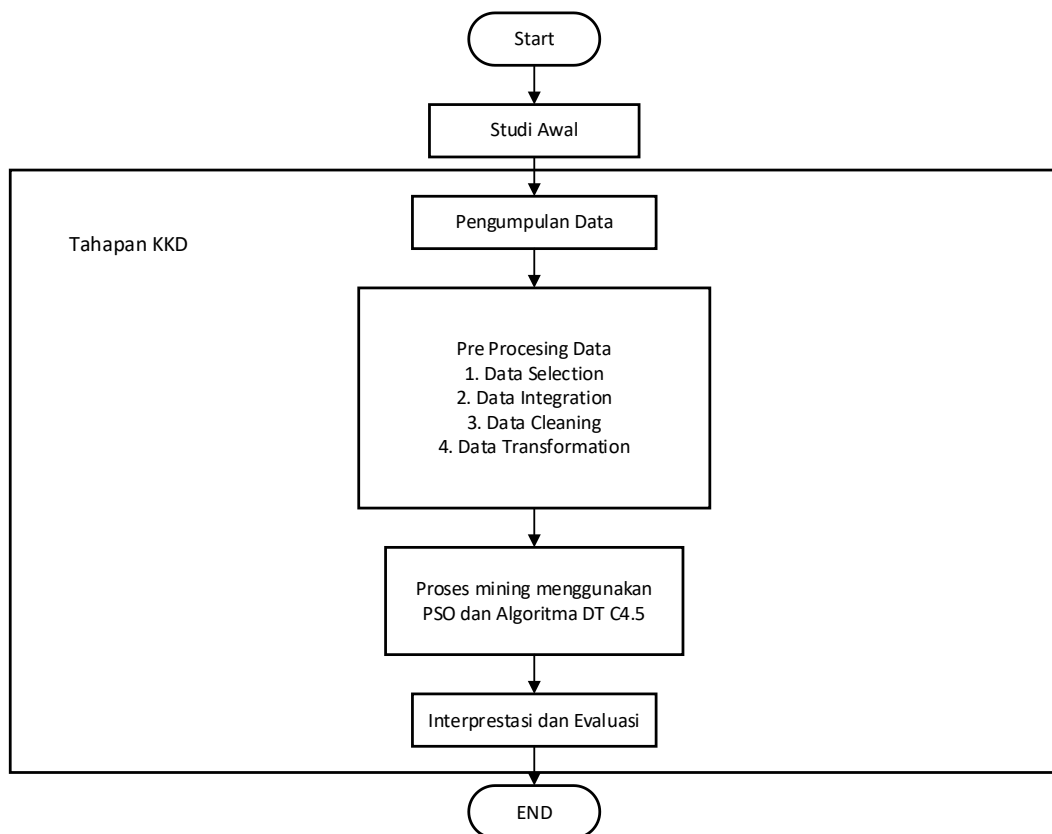


## BAB III METODE PENELITIAN

### 3.1 Metode Penelitian

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur dalam tahapan penelitian yang akan dilakukan berikut :



Gambar 3. 1 Alur dalam Tahapan Penelitian

Tahapan pada gambar 3.1 . Adalah proses data mining pada penelitian ini :

1. Tahapan studi awal merupakan tahap awal dalam melakukan penelitian. Pada tahap ini, peneliti melakukan observasi atau pemahaman penelitian yang meliputi tujuan dan persyaratan proyek dengan jelas dalam hal bisnis atau unit penelitian secara keseluruhan, menterjemahkan tujuan dan batasan ke dalam perumusan definisi masalah data mining, menyiapkan strategi awal untuk mencapai tujuan tersebut.
2. Tahap pengumpulan data merupakan suatu tahapan pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal dan mengevaluasi kualitas data. Data penelitian yang digunakan diperoleh dari situs kaggle.com (<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>).
3. *Preprocessing* data merupakan proses pengolahan data mentah kedalam bentuk yang lebih mudah dipahami, proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur.

a) *Data Selection*

Pada proses seleksi ini, dipilih dari keseluruhan data dari dataset Kaggle.com yaitu data prediksi stroke. Dari hasil seleksi akan digunakan sebanyak 5.110 data pasien untuk perhitungan data mining.

b) *Data Integration*

Proses ini dilakukan karena perlu adanya perubahan skema struktur

data dari data mentah ke data yang siap akan dilakukan proses *preprocessing* lebih lanjut dengan tahap *cleaning*.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke	
2	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1	
3	51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1	
4	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1	
5	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1	
6	1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1	
7	56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1	
8	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1	
9	10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1	
10	27419	Female	59	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1	
11	60491	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1	
12	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1	
13	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1	
14	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1	
15	8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1	
16	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1	
17	58202	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1	
18	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1	
19	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1	
20	27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1	

Gambar 3. 2 Data Sebelum *Integration*

Data akan berubah skema atau struktur menjadi seperti pada Gambar 3.3 yang mudah untuk diproses. Data tabel didefinisikan memanjang ke samping kanan dengan penggunaan kolom yang lebih banyak dan sesuai secara rapi serta type file yang sebelumnya .csv di export menjadi type data excel sehingga memudahkan untuk merubah data.

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
2	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
3	51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
4	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
5	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
6	1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
7	56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
8	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
9	10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
10	27419	Female	59	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1
11	60491	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
12	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
13	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
14	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
15	8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1
16	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
17	58202	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
18	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
19	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
20	27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
21	25226	Male	57	0	1	No	Govt_job	Urban	217.08	N/A	Unknown	1

Gambar 3. 3 Data Sesudah *Integration*

### c) Data Cleaning

Pada proses ini data yang tidak perlu dipakai pada penggunaan pengolahan data dan data yang akan menyebabkan kekacauan data sehingga data menyimpang atau tidak konsisten.

	A	B	C	D	E	F	G	H	I	J	K	L
	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
1	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
2	51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
3	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
4	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
5	1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
6	56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
7	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
8	10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
9	27419	Female	59	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1
10	60491	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
11	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
12	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
13	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
14	8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1
15	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
16	58202	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
17	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
18	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
19	27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
20	25226	Male	57	0	1	No	Govt_job	Urban	217.08	N/A	Unknown	1
21												

Gambar 3. 4 Data Sebelum Cleaning

Dari data *cleaning* terdapat data yang dibersihkan untuk memaksimalkan hasil akurasi dimana atribut *gender* mempunyai 1 *value* data yang tidak sesuai, kemudian atribut *bmi* terdapat 201 data *value* yang tidak sesuai serta atribut *smoking* status 1.482 data yang *valuenya* tidak sesuai sehingga dari keseluruhan dataset yang berjumlah 5.110 data setelah dilakukan proses *clenaing* data menjadi 3.425 data yang akan di proses untuk data mining.

	A	B	C	D	E	F	G	H	I	J	K	L
	id	gender	age	hypertensi	heart_disea	ever_marri	work_type	Residence_ty	avg_glucose_lex	bmi	smoking_statu	stroke
1	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
2	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
5	56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
6	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
7	10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
8	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
9	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
10	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
11	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
12	58202	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
13	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
14	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
15	27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
16	70630	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
17	13861	Female	52	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
18	68794	Female	79	0	0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1
19	4219	Male	71	0	0	Yes	Private	Urban	102.87	27.2	formerly smoked	1
20	70822	Male	80	0	0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1
21												

Gambar 3. 5 Data Sesudah *Cleaning*

#### d) Data Transformation

Proses *Transformation* merupakan perubahan data tahap akhir yang siap diolah dan dimasukkan kedalam tools atau aplikasi rapidminer.

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	gender	age	hypertensi	heart_disea	ever_marri	work_type	Residence_ty	avg_glucose_lei	bmi	smoking_statu	stroki
2	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
3	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
4	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
5	1665	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
6	56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
7	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
8	10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
9	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
10	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
11	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
12	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
13	58202	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
14	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
15	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
16	27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
17	70630	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
18	13861	Female	52	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
19	68794	Female	79	0	0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1
20	4219	Male	71	0	0	Yes	Private	Urban	102.87	27.2	formerly smoked	1
21	70822	Male	80	0	0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1

Gambar 3. 6 Dataset *Transformation*

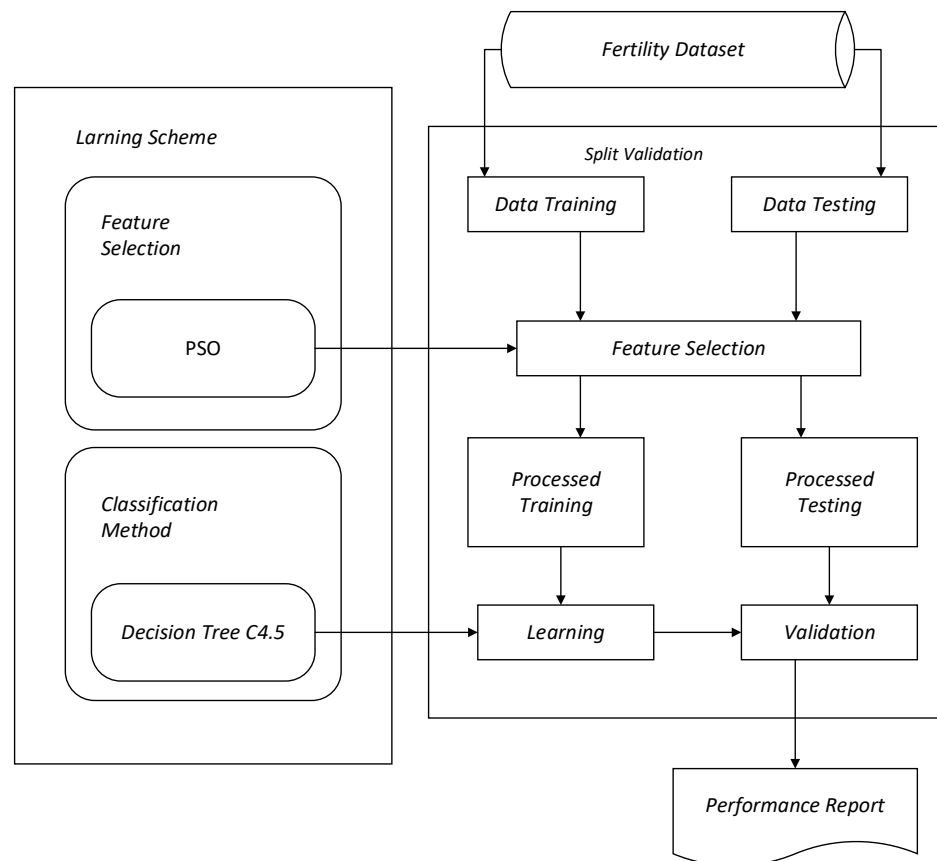
Semua atribut berjumlah 12 dimana terdapat 1 atribut id, 1 atribut label dan 10 atribut fitur. Penjelasan data dapat dilihat dibawah ini.

Tabel 3. 1 Keterangan Atribut Dataset

Atribut	Deskripsi	Kriteria
Id	Id data	84-72951
Gender	Jenis kelamin -Pria -Wanita	Male
		Female
Age	Usia	10-82
Hypertension	Riwayat hipertensi	1=Yes
		0=No
Heart_disease	Riwayat penyakit jantung	1=Yes
		0=No
Ever_married	Pernah menikah	Yes
		No
Work_type	Tipe pekerjaan	Private
		Self-employed
		Govt_job
		Never_worked
		Children
Residence_type	Tipe tempat tinggal	Urban
		Rural

Avg_glucose_level	Level kadar gula	55-271
Bmi	Indeks masa tubuh	11-92
Smoking_status	Status merokok	Formerly smoked
		Never smoked
		smoked
Stroke	Label	1=Yes
		0=No

4. Proses mining menggunakan metode *Particle Swam Optimization* (PSO) dan Algoritma *Decision Tree* C4.5 untuk meningkatkan akurasi.



Gambar 3. 7 Alur Proses PSO

- Tahapan pertama yaitu pengumpulan *dataset fertility* yang diambil dari *Kaggle repository*.
- Berikutnya pembagian data menggunakan *split validation*, yaitu membagi data menjadi *data training* dan *data testing*.

- c. Pada tahap *learning schame* menggunakan *Particle Swarm Optimization* untuk optimasi, sehingga akan mengoptimasi nilai bobot pada setiap atribut.

#### Pengujian PSO

- d. Selanjutnya proses klasifikasi menggunakan algoritma *Decision Tree* C4.5 dan masuk ketahap validasi atau melihat kebenaran dan hasil dari proses sebelumnya.

- Pengujian Algoritma *Decision Tree* C4.5

Pengujian data yang digunakan dalam algoritma C4.5 adalah dataset prediksi stroke yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada gambar berikut.

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
52282	Male	57	0	0	Yes	Private	Rural	197	34	formerly smoked	1
45535	Male	68	0	0	Yes	Private	Rural	233	42	never smoked	1
40460	Female	68	1	1	Yes	Private	Urban	247	40	formerly smoked	1
27153	Female	75	0	0	Yes	Self-employed	Rural	78	29	formerly smoked	1
30468	Male	58	1	0	Yes	Private	Urban	87	39	never smoked	0
56543	Female	70	0	0	Yes	Private	Rural	69	35	formerly smoked	0
52800	Female	52	0	0	Yes	Private	Urban	77	17	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	243	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	77	32	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	205	54	never smoked	0

Gambar 3. 8 Dataset Prediksi Stroke

Menghitung jumlah penderita terkena stroke dan tidak terkena stroke, serta entropy dari semua kasus dan semua dibagi berdasarkan atribut gender, age, hipertensi, heart disease, ever disease, work type, residence type, avg glucose level, bmi, smoking status. Perhitungan nilai entropy total dapat dicari dengan rumus berikut.

$$\begin{aligned}
 \text{Entropy (S)} &= (-4/10) \cdot \log_2(4/10) + (-6/10) \cdot \log_2(6/10) \\
 &= (-0,4) \cdot \log_2(0,4) + (-0,6) \cdot \log_2(0,6) \\
 &= (-0,4) \cdot (-1,322) + (-0,6) \cdot (-0,737) \\
 &= 0,528 + 0,443 \\
 &= 0,9709
 \end{aligned}$$

Setelah itu dilakukan perhitungan gain untuk setiap atribut, hasil perhitungan seperti dibawah ini.

Tabel 3. 2 Perhitungan Node Akar

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
<b>Total</b>		10	4	6	0,970951	
<b>gender</b>						0,366766196
	Male	3	2	1	0	
	Female	7	2	5	0,863121	
<b>Age</b>						-0,89122012
	<=49	1	0	1	0	
	>49	19	4	5	0,98009	
<b>hypertension</b>						0,005802149
	0	7	3	4	0,985228	
	1	3	1	2	0,918296	
<b>heart_disease</b>						0,007403392
	0	8	3	5	0,954434	
	1	2	1	1	1	
<b>ever_married</b>						0
	Yes	10	4	6	0,970951	
	No	0	0	0	0	
<b>work_type</b>						0,005802149
	Private	7	3	4	0,985228	
	Self-employed	3	1	2	0,918296	
	Govt_job	0	0	0	0	
	Never_worked	0	0	0	0	
	children	0	0	0	0	
<b>Residence_type</b>						0,04643934
	Urban	4	1	3	0,811278	
	Rural	6	3	3	1	
<b>avg_glucose_level</b>						-0,770392905
	<=108	13	3	10	0,77935	
	>108	7	3	2	1,04027	
<b>bmi</b>						0,005802149
	<=30	3	1	2	0,918296	
	>30	7	3	4	0,985228	
<b>smoking_status</b>						0,485475297
	formerly smoked	5	3	2	0,970951	
	never smoked	4	1	3	0	
	smokes	1	0	1	0	



Dari tabel data diatas ditemukan nilai gain tertinggi yang akan menjadi node akar selanjutnya untuk mencari cabang yang lain kita harus melakukan perhitungan entropy dan gain Kembali dengan melihat atribut *gain* yang tertinggi seperti gambar dibawah ini.

Tabel 3. 3 Perhitungan Node Cabang 1.1

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
<b>Total</b>		10	4	6	0,970951	
<b>gender</b>						0,570950594
	Male	1	1	0	0	
	Female	4	2	2	1	
<b>Age</b>						- 0,014524703
	<=49	0	0	0	0	
	>49	10	3	2	0,985475	
<b>hypertension</b>						0,570950594
	0	4	2	2	1	
	1	1	1	0	0	
<b>heart_disease</b>						0,570950594
	0	4	2	2	1	
	1	1	1	0	0	
<b>ever_married</b>						0,485475297
	Yes	5	3	2	0,970951	
	No	0	0	0	0	
<b>work_type</b>						0,570950594
	Private	4	2	2	1	
	Self-employed	1	1	0	0	
	Govt_job	0	0	0	0	
	Never_worked	0	0	0	0	
	children	0	0	0	0	
<b>Residence_type</b>						0,49546184
	Urban	2	1	1	1	
	Rural	3	2	1	0,918296	
<b>avg_glucose_level</b>						0,321928095
	<=108	8	2	6	0,811278	
	>108	2	2	0	0	
<b>bmi</b>						0,495461844
	<=30	2	1	1	1	
	>30	3	2	1	0,918296	

Tabel 3. 4 Perhitungan Node Cabang 1.2

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
<b>Total</b>		10	4	6	0,970951	
<b>gender</b>						0,695461844
	Male	0	0	0	0	
	Female	3	1	2	0,918296	
<b>Age</b>						0,474372166
	<=49	0	0	0	0	
	>49	5	1	2	0,993157	
<b>hypertension</b>						0,695461844
	0	3	1	2	0,918296	
	1	0	0	0	0	
<b>heart_disease</b>						0,695461844
	0	3	1	2	0,918296	
	1	0	0	0	0	
<b>ever_married</b>						0,695461844
	Yes	3	1	2	0,918296	
	No	0	0	0	0	
<b>work_type</b>						0,970950594
	Private	2	0	2	0	
	Self-employed	1	1	0	0	
	Govt_job	0	0	0	0	
	Never_worked	0	0	0	0	
	children	0	0	0	0	
<b>Residence_type</b>						0,77095059
	Urban	1	0	1	0	
	Rural	2	1	1	1	
<b>bmi</b>						0,25162917
	<=30	2	1	1	1	
	>30	1	0	1	0	

Tabel 3. 5 Perhitungan Node Cabang 1.3

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
<b>Total</b>		10	4	6	0,970951	
<b>gender</b>						0,970950594
	Male	0	0	0	0	
	Female	2	0	2	0	
<b>Age</b>						0,970950594
	<=49	0	0	0	0	
	>49	2	0	2	0	

<b>hypertension</b>						0,970950594
	0	2	0	2	0	
	1	0	0	0	0	
<b>heart_disease</b>						0,970950594
	0	2	0	2	0	
	1	0	0	0	0	
<b>ever_married</b>						0,970950594
	Yes	2	0	2	0	
	No	0	0	0	0	
<b>Residence_type</b>						0,97095059
	Urban	1	0	1	0	
	Rural	1	0	1	0	
<b>bmi</b>						0,97095059
	<=30	1	0	1	0	
	>30	1	0	1	0	

- e. Berikutnya tahapan terakhir yaitu *performance report*, pada tahap akhir ini menampilkan hasil dari pengolahan dari proses sebelumnya. Tahap ini menghasilkan *Confusion Matrix* dan kurva ROC.
5. Tahap interpretasi dan evaluasi adalah mengevaluasi satu atau lebih model yang disampaikan dalam tahap pemodelan untuk kualitas dan efektivitas sebelum menerapkannya untuk digunakan di lapangan. Pada tahap ini yang akan dilakukan adalah menentukan model sebenarnya mencapai tujuan yang ditetapkan untuk itu pada tahapan pertama, menetapkan apakah beberapa aspek penting dari masalah bisnis atau penelitian belum diperhitungkan secara memadai, mengambil keputusan tentang penggunaan hasil data mining.

### **3.2 Alat dan Bahan**

Penelitian ini menggunakan perangkat keras Laptop ASUS VivoBook dengan Processor Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz, RAM 8,00 GB, sedangkan perangkat lunak yang digunakan Microsoft Excel dan RapidMiner Studio Developer 9.6 untuk pengolahan data. Sedangkan bahan yang digunakan diambil dari (<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>).