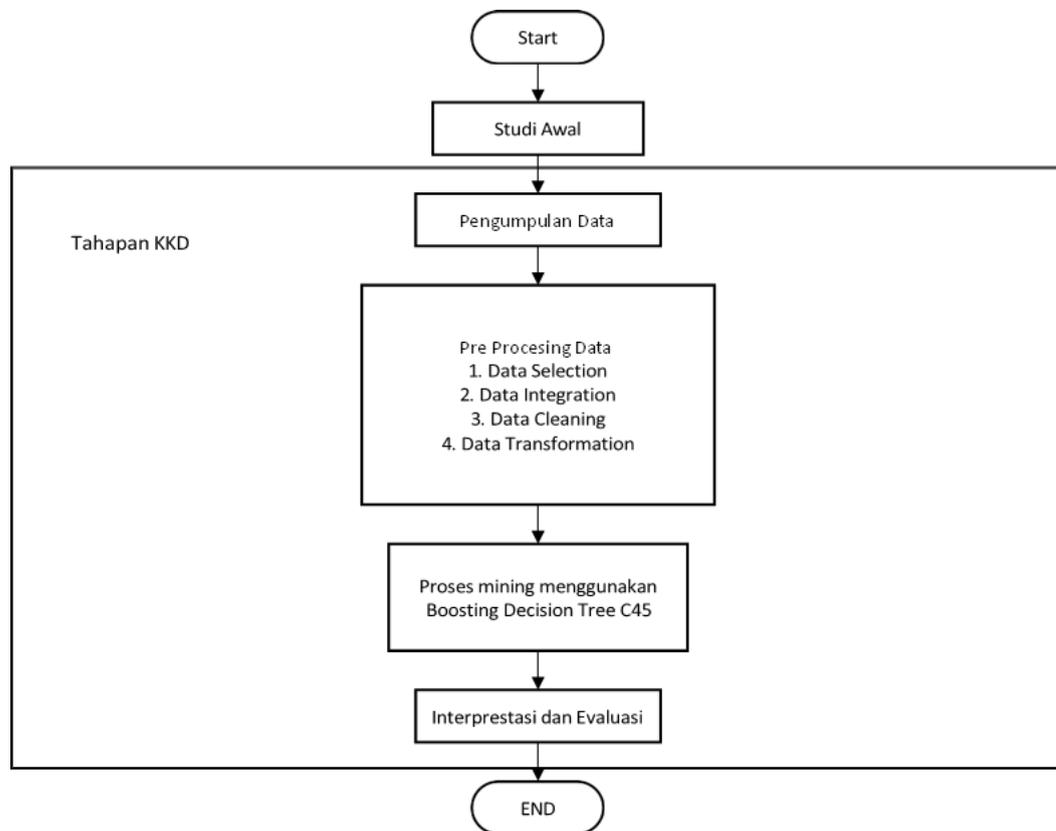


## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Metode Penelitian

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

Tahap 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 dataeksplorasi 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/code/mpwolke/obesity-levels-life-style/notebook>).

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 810 data obesitas 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*.

Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Female	21	1.62	64	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Tra	Normal_Weight
Female	21	1.52	56	yes	no	3	3	Sometimes	yes	3	yes	3	0	Sometimes	Public_Tra	Normal_Weight
Male	23	1.8	77	yes	no	2	3	Sometimes	no	2	no	2	1	Frequently	Public_Tra	Normal_Weight
Male	27	1.8	87	no	no	3	3	Sometimes	no	2	no	2	0	Frequently	Walking	Overweight_Level_I
Male	22	1.78	89.8	no	no	2	1	Sometimes	no	2	no	0	0	Sometimes	Public_Tra	Overweight_Level_II
Male	29	1.62	53	no	yes	2	3	Sometimes	no	2	no	0	0	Sometimes	Automobile	Normal_Weight
Female	23	1.5	55	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometimes	Motorbike	Normal_Weight
Male	22	1.64	53	no	no	2	3	Sometimes	no	2	no	3	0	Sometimes	Public_Tra	Normal_Weight
Male	24	1.78	64	yes	yes	3	3	Sometimes	no	2	no	1	1	Frequently	Public_Tra	Normal_Weight
Male	22	1.72	68	yes	yes	2	3	Sometimes	no	2	no	1	1	no	Public_Tra	Normal_Weight
Male	26	1.85	105	yes	yes	3	3	Frequently	no	3	no	2	2	Sometimes	Public_Tra	Obesity_Type_I
Female	21	1.72	80	yes	yes	2	3	Frequently	no	2	yes	2	1	Sometimes	Public_Tra	Overweight_Level_II
Male	22	1.65	56	no	no	3	3	Sometimes	no	3	no	2	0	Sometimes	Public_Tra	Normal_Weight
Male	41	1.8	99	no	yes	2	3	Sometimes	no	2	no	2	1	Frequently	Automobile	Obesity_Type_I
Male	23	1.77	60	yes	yes	3	1	Sometimes	no	1	no	1	1	Sometimes	Public_Tra	Normal_Weight
Female	22	1.7	66	yes	no	3	3	Always	no	2	yes	2	1	Sometimes	Public_Tra	Normal_Weight
Male	27	1.93	102	yes	yes	2	1	Sometimes	no	1	no	1	0	Sometimes	Public_Tra	Overweight_Level_II
Female	29	1.53	78	no	yes	2	1	Sometimes	no	2	no	0	0	no	Automobile	Obesity_Type_I

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.

Gender	Age	Height	Weight	family_his	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Female	21	1.62	64	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Tra	Normal_Weight
Female	21	1.52	56	yes	no	3	3	Sometimes	yes	3	yes	3	0	Sometimes	Public_Tra	Normal_Weight
Male	23	1.8	77	yes	no	2	3	Sometimes	no	2	no	2	1	Frequently	Public_Tra	Normal_Weight
Male	27	1.8	87	no	no	3	3	Sometimes	no	2	no	2	0	Frequently	Walking	Overweight_Level_I
Male	22	1.78	89.8	no	no	2	1	Sometimes	no	2	no	0	0	Sometimes	Public_Tra	Overweight_Level_II
Male	29	1.62	53	no	yes	2	3	Sometimes	no	2	no	0	0	Sometimes	Automobi	Normal_Weight
Female	23	1.5	55	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometimes	Motorbike	Normal_Weight
Male	22	1.64	53	no	no	2	3	Sometimes	no	2	no	3	0	Sometimes	Public_Tra	Normal_Weight
Male	24	1.78	64	yes	yes	3	3	Sometimes	no	2	no	1	1	Frequently	Public_Tra	Normal_Weight
Male	22	1.72	68	yes	yes	2	3	Sometimes	no	2	no	1	1	no	Public_Tra	Normal_Weight
Male	26	1.85	105	yes	yes	3	3	Frequently	no	3	no	2	2	Sometimes	Public_Tra	Obesity_Type_I
Female	21	1.72	80	yes	yes	2	3	Frequently	no	2	yes	2	1	Sometimes	Public_Tra	Overweight_Level_II
Male	22	1.65	56	no	no	3	3	Sometimes	no	3	no	2	0	Sometimes	Public_Tra	Normal_Weight
Male	41	1.8	99	no	yes	2	3	Sometimes	no	2	no	2	1	Frequently	Automobi	Obesity_Type_I
Male	23	1.77	60	yes	yes	3	1	Sometimes	no	1	no	1	1	Sometimes	Public_Tra	Normal_Weight
Female	22	1.7	66	yes	no	3	3	Always	no	2	yes	2	1	Sometimes	Public_Tra	Normal_Weight
Male	27	1.93	102	yes	yes	2	1	Sometimes	no	1	no	1	0	Sometimes	Public_Tra	Overweight_Level_II
Female	29	1.53	78	no	yes	2	1	Sometimes	no	2	no	0	0	no	Automobi	Obesity_Type_I

Gambar 3. 3 Data Sesudah *Integration*

c) Data Cleaning

Pada proses ini adalah pemilihan data yang tidak perlu digunakan seperti *missing values* dan *redundance data*. Pada tahap ini juga mengubah format data *text* menjadi *numeric* sehingga mudah dibaca oleh sistem.

Gender	Age	Height	Weight	family_his FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObesyedad	
Female	21	1.62	64	yes	no	2	3	Sometime	no	2	no	0	1	no	Public_TraNormal_Weight	
Female	21	1.52	56	yes	no	3	3	Sometime	yes	3	yes	3	0	Sometime	Public_TraNormal_Weight	
Male	23	1.8	77	yes	no	2	3	Sometime	no	2	no	2	1	Frequenth	Public_TraNormal_Weight	
Male	27	1.8	87	no	no	3	3	Sometime	no	2	no	2	0	Frequenth	Walking_Overweight_Level_I	
Male	22	1.78	89.8	no	no	2	1	Sometime	no	2	no	0	0	Sometime	Public_TraOverweight_Level_II	
Male	29	1.62	53	no	yes	2	3	Sometime	no	2	no	0	0	Sometime	AutomobiNormal_Weight	
Female	23	1.5	55	yes	yes	3	3	Sometime	no	2	no	1	0	Sometime	MotorbikeNormal_Weight	
Male	22	1.64	53	no	no	2	3	Sometime	no	2	no	3	0	Sometime	Public_TraNormal_Weight	
Male	24	1.78	64	yes	yes	3	3	Sometime	no	2	no	1	1	Frequenth	Public_TraNormal_Weight	
Male	22	1.72	68	yes	yes	2	3	Sometime	no	2	no	1	1	no	Public_TraNormal_Weight	
Male	26	1.85	105	yes	yes	3	3	Frequenth	no	3	no	2	2	Sometime	Public_TraObesity_Type_I	
Female	21	1.72	80	yes	yes	2	3	Frequenth	no	2	yes	2	2	1	Sometime	Public_TraOverweight_Level_II
Male	22	1.65	56	no	no	3	3	Sometime	no	3	no	2	0	Sometime	Public_TraNormal_Weight	
Male	41	1.8	99	no	yes	2	3	Sometime	no	2	no	2	1	Frequenth	AutomobiObesity_Type_I	
Male	23	1.77	60	yes	yes	3	1	Sometime	no	1	no	1	1	Sometime	Public_TraNormal_Weight	
Female	22	1.7	66	yes	no	3	3	Always	no	2	yes	2	1	Sometime	Public_TraNormal_Weight	
Male	27	1.93	102	yes	yes	2	1	Sometime	no	1	no	1	0	Sometime	Public_TraOverweight_Level_II	
Female	29	1.53	78	no	yes	2	1	Sometime	no	2	no	0	0	no	AutomobiObesity_Type_I	

Gambar 3. 4 Data Sebelum Cleaning

Gender	Age	family_his FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObesyedad	
Male	26	2	2	3	3	3	1	3	1	2	2	2	4	Obesity_Type_I
Male	41	1	2	2	3	2	1	2	1	2	1	3	1	Obesity_Type_I
Female	29	1	2	2	1	2	1	2	1	0	0	1	1	Obesity_Type_I
Female	52	2	2	3	1	2	2	2	1	0	0	1	1	Obesity_Type_I
Female	22	2	2	1	1	2	1	2	1	0	2	2	4	Obesity_Type_I
Male	23	2	2	2	3	4	1	2	1	0	1	3	1	Obesity_Type_I
Male	30	2	2	1	3	1	2	2	2	0	0	3	1	Obesity_Type_II
Female	23	2	2	2	3	2	1	3	1	0	1	1	4	Obesity_Type_I
Male	23	2	2	2	3	4	1	2	1	0	1	1	4	Obesity_Type_I
Female	25	1	1	3	4	4	1	1	1	2	0	1	4	Obesity_Type_II
Female	38	2	2	2	3	2	1	2	1	0	0	2	1	Obesity_Type_I
Male	22	2	1	2	3	2	1	3	1	3	2	1	5	Obesity_Type_I
Female	21	2	2	3	1	2	1	3	1	2	1	1	4	Obesity_Type_I
Female	15	2	2	3	3	2	1	1	1	3	2	1	5	Obesity_Type_I
Male	30	2	2	3	3	2	1	1	1	2	0	2	1	Obesity_Type_I
Male	44	2	1	2	3	2	2	3	1	0	0	1	2	Obesity_Type_I
Male	23	1	2	2	3	3	2	1	1	1	1	3	1	Obesity_Type_I
Female	36	2	1	3	3	2	1	1	1	0	0	2	1	Obesity_Type_I

Gambar 3. 5 Data Sesudah Cleanin

d) Data Transformation

Proses *Transformation* merupakan tahapan penyesuaian data yang siap diolah dan digunakan sesuai kriteria tools yang dibutuhkan.

Gender	Age	family_his	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObesyesdad
Male	26	2	2	3	3	3	1	3	1	2	2	2	2	4 Obesity_Type_I
Male	41	1	2	2	3	2	1	2	1	2	1	3	1	1 Obesity_Type_I
Female	29	1	2	2	1	2	1	2	1	0	0	1	1	1 Obesity_Type_I
Female	52	2	2	3	1	2	2	2	1	0	0	1	1	1 Obesity_Type_I
Female	22	2	2	1	1	2	1	2	1	0	2	2	4	4 Obesity_Type_I
Male	23	2	2	2	3	4	1	2	1	0	1	3	1	1 Obesity_Type_I
Male	30	2	2	1	3	1	2	2	2	0	0	3	1	1 Obesity_Type_I
Female	23	2	2	2	3	2	1	3	1	0	1	1	4	4 Obesity_Type_I
Male	23	2	2	2	3	4	1	2	1	0	1	1	4	4 Obesity_Type_I
Female	25	1	1	3	4	4	1	1	1	2	0	1	4	4 Obesity_Type_I
Female	38	2	2	2	3	2	1	2	1	0	0	2	1	1 Obesity_Type_I
Male	22	2	1	2	3	2	1	3	1	3	2	1	5	5 Obesity_Type_I
Female	21	2	2	3	1	2	1	3	1	2	1	1	4	4 Obesity_Type_I
Female	15	2	2	3	3	2	1	1	1	3	2	1	5	5 Obesity_Type_I
Male	30	2	2	3	3	2	1	1	1	2	0	2	1	1 Obesity_Type_I
Male	44	2	1	2	3	2	2	3	1	0	0	1	2	2 Obesity_Type_I
Male	23	1	2	2	3	3	2	1	1	1	1	3	1	1 Obesity_Type_I
Female	36	2	1	3	3	2	1	1	1	0	0	2	1	1 Obesity_Type_I

Gambar 3. 6 Dataset *Transformation*

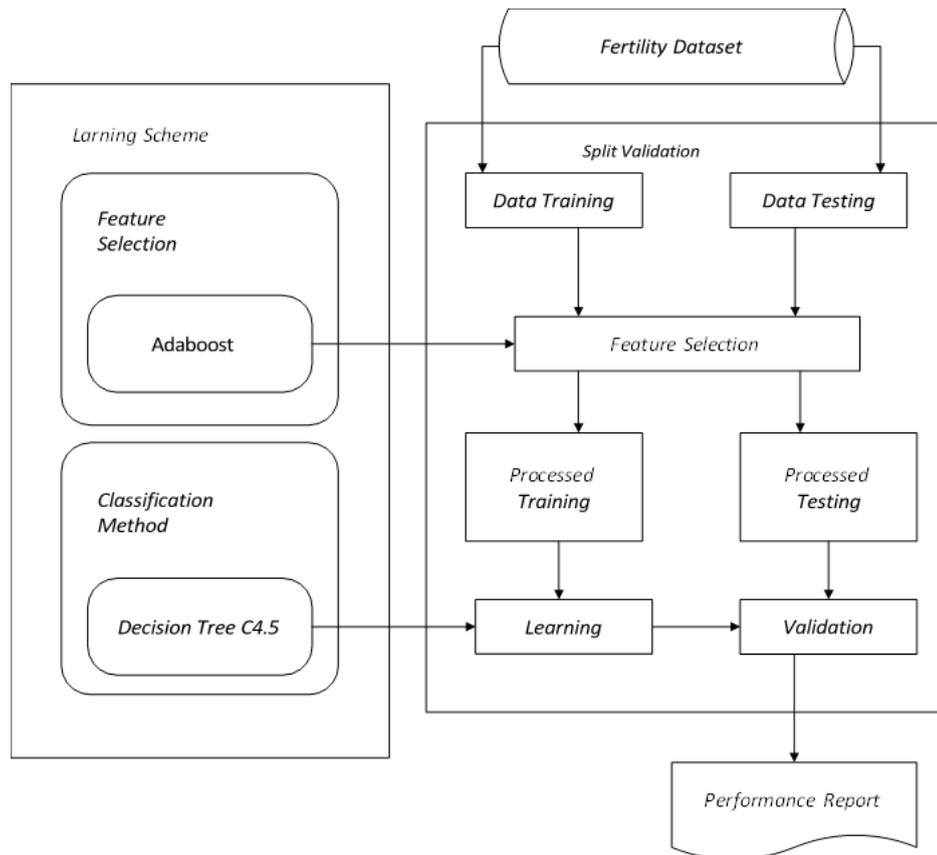
Berikut ini adalah Atribut yang digunakan untuk diolah. Penjelasan data dapat dilihat dibawah ini.

Tabel 3. 1 Atribut Dataset Level Obesitas

Atribut	Deskripsi	Kriteria
Jenis Kelamin	- Wanita - Pria	- Wanita - Pria
Usia	Usia	14 – 61
Riwayat Keluarga Memiliki Obesitas	Keluarga yang menderita obesitas	1 = Tidak 2 = Ya
FAVC	Atribut yang menunjukkan seringnya konsumsi makanan dengan highcalories	1 = tidak 2 = Ya
FCVC	Atribut yang menunjukkan frekuensi konsumsi sayuran	1 = Tidak pernah 2 = Kadang-kadang 3 = Selalu
NCP	Jumlah makanan pokok yang dikonsumsi setiap harinya	1 = Satu 2 = Dua 3 = Tiga 4 = lebih dari tiga
CAEC	Frekuensi konsumsi makanan di antara waktu makan	1 = Tidak. 2 = Kadang-kadang 3 = Sering 4 = Selalu
Smoke	Perokok atau tidak	1 = Tidak. 2 = Ya
CH2O	Jumlah air untuk diminum setiap hari	1 = Kurang dari satu liter 2 = Antara 1 dan 2 L 3 = Lebih dari 2 L

SCC	Pantau kalori yang dikonsumsi setiap hari	1 = Tidak.
		2 = Ya
FAF	Waktu Menggunakan Perangkat	0 = Saya tidak memiliki
		1 = 1 Or 2 hari
		2 = 2 atau 4 hari
		3 = 4 atau 5 hari
TUE	Saatnya menggunakan perangkat	0 = 0 - 2 jam
		1 = 3 - 5 jam
		2 = Lebih dari 5 jam
CALC	Frekuensi minum Alcohol	1 = Tidak.
		2 = Kadang-kadang
		3 = Sering
		4 = Selalu
Mtrans	Transportasi yang sering digunakan setiap hari	1 = Mobil
		2 = Sepeda motor
		3 = Sepeda
		4 = Transportasi Umum
		5 = Berjalan
Nobesity	Label	Obesitas Tipe I, Obesitas Tipe II Obesitas Tipe III.

4. Proses mining menggunakan metode *Boosting Decision Decision Tree* C4.5 diharapkan dengan menggunakan metode ini dapat menghasilkan tingkat akurasi yang lebih baik.



Gambar 3. 7 Alur Proses Boosting DT C.45

- a. Tahapan pertama yaitu pengumpulan *dataset fertility* yang diambil dari *Kaggle repository*.
- b. Berikutnya pembagian data menggunakan *split validation*, yaitu membagi data menjadi *data training* dan *data testing*.
- c. Pada tahap *learning schame* menggunakan *Ada Boost sebagai Boosting* untuk menghasilkan nilai akurasi yang lebih baik
- d. Selanjutnya proses klasifikasi menggunakan algoritma *Decision Tree C4.5* dan masuk ketahap validasi atau melihat kebenaran dan hasil dari proses sebelumnya.

### **3.2 Alat dan Bahan**

Berdasarkan Rancangan yang telah dibuat maka penelitian ini menggunakan perangkat keras Laptop MSI Modern 14 dengan Processor Intel Core i3 Generasi 11, 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/>)