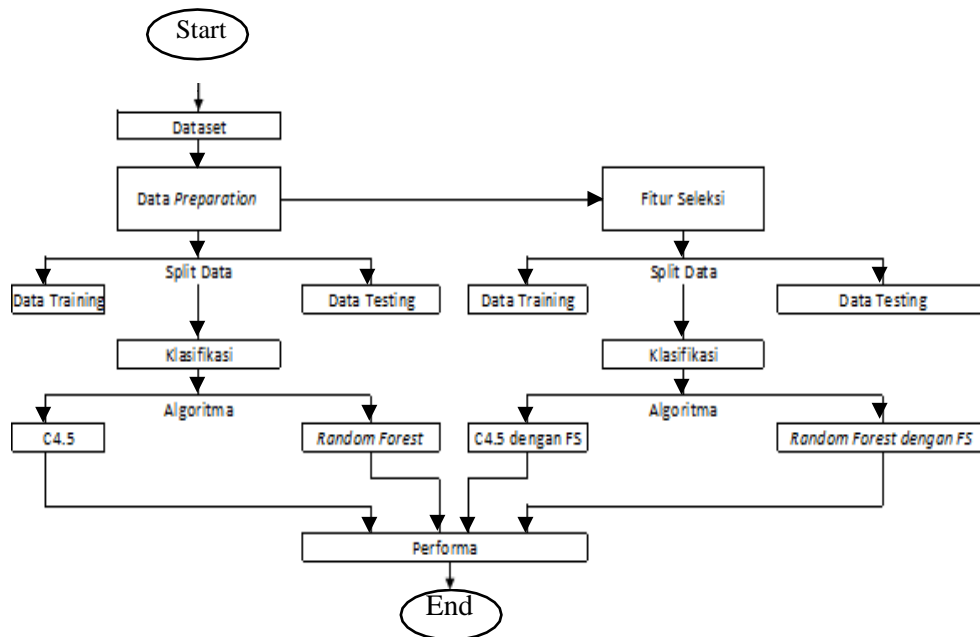


BAB III METODE PENELITIAN

3.1 Metode Penelitian

Pada bab ini akan disampaikan beberapa langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data yang diperoleh dari 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 Alur dalam Tahapan Penelitian :

a. Dataset.

Dataset adalah sekumpulan data yang disusun secara terstruktur. Biasanya, dataset dipresentasikan dalam bentuk tabel, alias baris dan kolom. Dataset terdiri atas dua yaitu dataset private dan dataset public. Dataset private

merupakan dataset yang diambil dari instansi atau organisasi perusahaan tempet kerja atau penelitian, contohnya bank, rumah sakit, sekolahan, industry dan lain-lain. Sedangkan dataset public adalah dataset yang diambil dari repository public yang disepakati oleh para peneliti data mining, contoh: uci repository (<https://archive.ics.uci.edu/>).

Data penelitian ini menggunakan dataset public yang diperoleh dari situs kaggle.com (<https://www.kaggle.com/datasets/mansoordaku/ckdisease>), terdiri atas 25 atribut (age, bp, sg, al, su,rbc, pc, pcc, ba, bgr, bu,sc, sod, pot, hemo, pcv, wbcc, rbcc, htn, dm, cad. Appet, pe, ane, class) dan 400 record.

Secara detail atribut data penelitian dapat dilihat pada tabel 3.1.

Tabel 3 1 Atribut Data Penelitian

NO	ATRIBUT	DESKRIPSI	TIPE	NILAI
1	age	Umur	Numerik	Umur dalam tahun
2	bp	Ukuran tekanan yang dihasilkan oleh darah	Numerik	mm/Hg
3	sg	Seriusnya suatu situasi atau kondisi	Numerik	1.005,1.010,1.015,1.020,1.025
4	al	Protein plasma darah	Numerik	0,1,2,3,4,5
5	su	Kadar gula darah	Numerik	0,1,2,3,4,5
6	rbc	Pengukuran sel darah merah	Nominal	Normal,abnormal
7	pc	Pengukuran sel nanah dalam urin	Nominal	Normal,abnormal
8	pcc	Pengukuran gumpalan sel nanah dalam urin	Nominal	Present,notpresent
9	ba	Pengukuran bakteri yang terdapat dalam tubuh	Nominal	Present,notpresent
10	bgr	Pengukuran kadar glukosa	Numerik	mgs/dl
11	bu	Pengukuran kadar urea dalam darah	Numerik	mgs/dl
12	sc	Pengukuran kreatin pada darah	Numerik	mgs/dl

13	sod	Pengukuran natrium pada darah	Numerik	mEq/L
14	pot	Pengukuran kalium pada darah	Numerik	mEq/L
15	hemo	Menghitung jumlah hemoglobin	Numerik	gms
16	pcv	Parameter pengukuran sel darah	Nominal	numeric
17	wbcc	Menghitung jumlah total sel darah merah	Nominal	Millions/cmm
18	rbcc	Menghitung jumlah total sel darah merah	Nominal	Millions/cmm
19	htn	Memiliki masalah hipertensi	Nominal	Yes,no
20	dm	Memiliki masalah diabetes	Nominal	Yes,no
21	cad	Coronary Artery Disease	Nominal	Yes,no
22	appet	Memiliki masalah nafsu makan	Nominal	Good,poor
23	pe	Pengukuran cairan pada betis	Nominal	Yes,no
24	ane	Pengukuran darah merah	Nominal	Yes,no
25	class	Target Class	Nominal	Ckd, notckd

b. Data Preparation,

Pada Tahap ini mengubah data mentah menjadi data yang dapat dianalisis, ini diperlukan untuk mendapatkan data yang berkualitas supaya menghasilkan model yang lebih efektif dan efisien.

c. Data Training dan Data Pengujian,

Data training digunakan untuk proses pelatihan algoritme C4.5 dengan algoritme *Random Forest* menggunakan 20% sebagai data uji dari total data yang ada di dataset. Data testing digunakan untuk menguji algoritme C4.5 dengan algoritme *Random Forest* menggunakan 80% data dari total data pada dataset.

- d. Penerapan Algoritma C4.5 dan Algoritme *Random Forest* pada Google Colab atau *Google Colaboratory* adalah sebuah *executable document* yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah di tulis melalui Google Drive. Beberapa manfaat Google Colab yaitu sesuai dengan namanya sendiri adalah *Collaborate*, dapat berkolaborasi dengan pengguna lainnya melalui berbagai coding secara online, Free GPU Google Colab memudahkan pengguna untuk menjalankan program computer dengan spek tinggi.[37]

Lakukan validasi terpisah untuk membagi dataset menjadi data training dan data testing. Penerapan algoritme C4.5 dengan algoritme *Random Forest* dapat di tool google colab kemudian melakukan penghitungan pada dataset yang tersedia. Selain perhitungan, google colab juga digunakan untuk melakukan tes untuk algoritme C4.5 dengan algoritme *Random Forest*, termasuk dalam seleksi fitur.

- e. Pengujian Akurasi.

Pada tahapan ini, dilakukan pengujian terhadap model yang dihasilkan untuk mendapatkan informasi model yang akurat. Evaluasi dilakukan dengan menggunakan metode *Confusion Matrix* dimana *Accuracy* , *precision*, *recall*, F1 Score, maupun Kurva ROC diperoleh. Proses pengujian metode yang diusulkan dilakukan dengan mengevaluasi perbandingan hasil *Accuracy*, *precision*, *recall*, F1 Score seluruh percobaan yang dilakukan menggunakan algoritme C4.5 dengan algoritme *Random Forest*, ditambah dengan penggunaan seleksi fitur dan memvalidasi model prediksi yang dianggap

paling optimal, dimana, semakin tinggi nilai *Accuracy* semakin baik pula metode yang digunakan.

3.2 Metode Pengumpulan Data

Pada tahap ini terdapat dua jenis data yang diperoleh untuk dijadikan sumber data yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan dari sumbernya langsung, sedangkan data sekunder yaitu data yang dikumpulkan dari peneliti sebelumnya, yang mana tingkat kebenaran data lebih tinggi.

Data yang digunakan pada penelitian ini menggunakan data sekunder. Data penelitian diambil dari hasil pemeriksaan data pasien penyakit ginjal kronis atau chronic kidney disease (CKD) yang diperoleh dari University of California Irvine (UCI) Machine Learning Data Repository, dengan jumlah 400 record yang terdiri dari 250 record (62,50%) pasien positif menderita penyakit ginjal kronis dan 150 record (37,50%) pasien negatif menderita penyakit tersebut. Dengan jumlah 24 atribut dan 1 atribut kelas. Secara detail data penyakit ginjal kronis dapat dilihat pada tabel 3.2.

Tabel 3 2 Data Penyakit Ginjal Kronis

id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	class	
1	40	102	1	0	normal	notprese	notprese	121	36	12					154	44	7800	5,2	yes	yes	no	good	no	no	ckd	
2	7	50	1,02	4	0	normal	notprese	notprese	18	0,8					11,3	38	6000		no	no	no	good	no	no	ckd	
3	2	62	80	1,01	2	3	normal	normal	notprese	notprese	423	53	1,8		9,6	31	7500		no	yes	no	poor	no	yes	ckd	
4	3	48	70	1,005	4	0	normal	abnorma	present	notprese	117	56	3,8	111	2,5	11,2	32	6700	3,9	yes	no	no	poor	yes	ckd	
5	4	51	80	1,01	2	0	normal	normal	notprese	notprese	106	26	1,4		11,6	35	7300	4,6	no	no	no	good	no	no	ckd	
6	5	60	90	1,015	3	0	normal	normal	notprese	notprese	74	25	1,1	142	3,2	12,2	39	7800	4,4	yes	yes	no	good	yes	no	ckd
7	6	68	70	1,01	0	0	normal	normal	notprese	notprese	100	54	24	104	4	12,4	36		no	no	no	good	no	no	ckd	
8	7	24		1,015	2	4	normal	abnorma	notprese	notprese	410	31	1,1		12,4	44	6900	5	no	yes	no	good	yes	no	ckd	
9	8	52	100	1,015	3	0	normal	abnorma	present	notprese	138	60	1,9		10,8	33	9600	4	yes	yes	no	poor	no	yes	ckd	
10	9	53	90	1,02	2	0	abnorma	abnorma	present	notprese	70	107	7,2	114	3,7	9,5	29	12100	3,7	yes	yes	no	poor	no	yes	ckd
11	10	50	60	1,01	2	4	abnorma	abnorma	present	notprese	490	55	4		9,4	28			yes	yes	no	good	no	yes	ckd	
12	11	63	70	1,01	3	0	abnorma	abnorma	present	notprese	380	60	2,7	131	4,2	10,8	32	4500	3,8	yes	yes	no	poor	yes	no	ckd
13	12	68	70	1,015	3	1	normal	normal	present	notprese	208	72	2,1	138	5,8	9,7	28	12200	3,4	yes	yes	yes	poor	yes	no	ckd
14	13	68	70								98	86	4,6	135	3,4	9,8			yes	yes	yes	poor	yes	no	ckd	
15	14	68	80	1,01	3	2	normal	abnorma	present	present	157	90	4,1	130	6,4	5,6	16	11000	2,6	yes	yes	poor	yes	no	ckd	
16	15	40	80	1,015	3	0	normal	normal	notprese	notprese	76	162	9,6	141	4,9	7,6	24	3800	2,8	yes	no	no	good	no	yes	ckd
17	16	47	70	1,015	2	0	normal	normal	notprese	notprese	99	46	2,2	138	4,1	12,6			no	no	no	good	no	no	ckd	
18	17	47	80								114	87	5,2	139	3,7	12,1			yes	no	no	poor	no	no	ckd	
19	18	60	100	1,025	0	3	normal	normal	notprese	notprese	263	27	1,3	135	4,3	12,7	37	11400	4,3	yes	yes	yes	good	no	no	ckd
20	19	62	60	1,015	1	0	abnorma	abnorma	present	notprese	100	31	1,6		10,9	30	5300	3,7	yes	no	yes	good	no	no	ckd	
21	20	61	80	1,015	2	0	abnorma	abnorma	notprese	notprese	173	148	3,9	135	5,2	7,7	24	9200	3,2	yes	yes	poor	yes	yes	ckd	
22	21	60	90								180	76	4,5		10,9	32	6200	3,6	yes	yes	yes	good	no	no	ckd	
23	22	48	80	1,025	4	0	normal	abnorma	notprese	notprese	95	163	7,7	136	3,8	9,8	32	6900	3,4	yes	no	no	good	no	yes	ckd
24	23	21	70	1,01	0	0	normal	normal	notprese	notprese									no	no	no	poor	no	yes	ckd	
25	24	42	100	1,015	4	0	normal	abnorma	notprese	present		50	1,4	129	4	11,1	39	8300	4,6	yes	no	no	poor	no	no	ckd
26	25	61	60	1,025	0	0	normal	normal	notprese	notprese	108	75	1,9	141	5,2	9,9	29	8400	3,7	yes	yes	no	good	no	yes	ckd
27	26	75	80	1,015	0	0	normal	normal	notprese	notprese	156	45	2,4	140	3,4	11,6	35	10300	4	yes	yes	no	poor	no	no	ckd
28	27	60	70	1,01	3	0	normal	abnorma	notprese	notprese	161	87	3,7	130	4	13,5	37	6600	4,1	yes	yes	no	poor	no	yes	ckd

3.3 Eksperimen dan Pengujian Data

Pada penelitian ini dilakukan proses eksperimen dan pengujian model menggunakan dataset sekunder Penyakit Ginjal Kronis diambil dari (<https://www.kaggle.com/datasets/mansoordaku/ckdisease>).

Dalam eksperimen penelitian ini menggunakan software dan hardware sebagai alat bantu dengan spesifikasi perangkat keras Laptop DELL dengan Processor Intel(R) Core(TM) i3-4030U CPU @ 1.90GHz 1.90 GHz, RAM 4,00 GB, sedangkan perangkat lunak yang digunakan Microsoft Excel dan Software tool google colablaboratory bahasa pemrograman python untuk pengolahan data.

3.4 Evaluasi dan Validasi Hasil

Hasil akhir dari penelitian ini adalah kegiatan validasi terhadap model yang digunakan, validasi ini dilakukan untuk menguji terhadap model prediksi yang dianggap paling optimal pada peningkatan nilai akurasi dalam pengklasifikasian data.