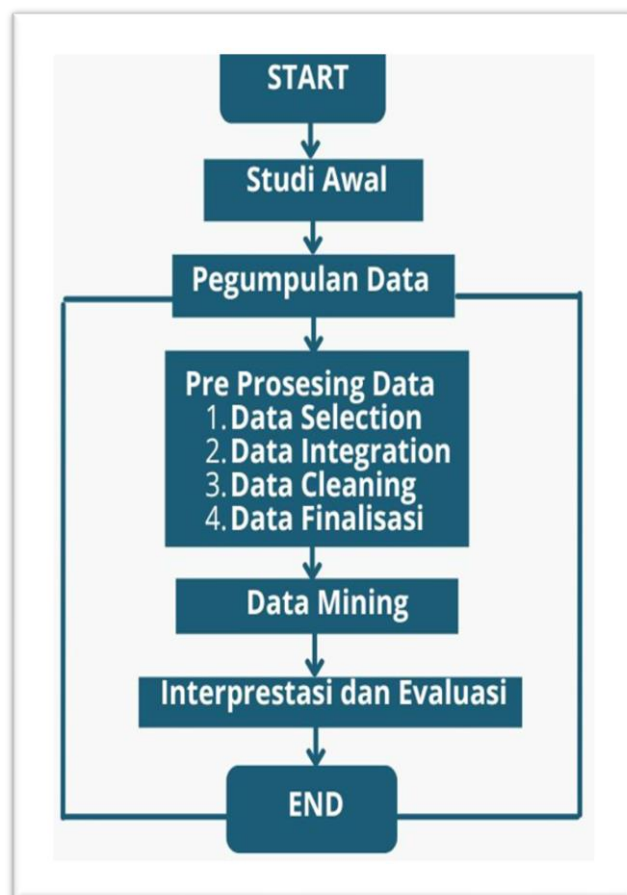


BAB III

METODOLOGI 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 ditunjukkan pada gambar 3.1.



Gambar 3.1 Alur Penelitian

Tahapan pada gambar 3.1. merupakan proses data mining pada penelitian ini yang dijelaskan sebagai berikut:

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/code/docxian/stroke-prediction) <https://www.kaggle.com/code/docxian/stroke-prediction>
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 dan 12 atribut untuk perhitungan data mining.

	A	B	C	D	E	F	G	H	I	J	K	
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		
22	70630	Female	71,0,0	Yes	Govt_job	Rural	193.94	22.4	smokes	1		
23	13861	Female	52,1,0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1		
24	68794	Female	79,0,0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1		
25	64778	Male	82,0,1	Yes	Private	Rural	208.3	32.5	Unknown	1		
26	4219	Male	71,0,0	Yes	Private	Urban	102.87	27.2	formerly smoked	1		
27	70822	Male	80,0,0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1		
28	38047	Female	65,0,0	Yes	Private	Rural	100.98	28.2	formerly smoked	1		
29	61843	Male	58,0,0	Yes	Private	Rural	189.84	N/A	Unknown	1		
30	54827	Male	69,0,1	Yes	Self-employed	Urban	195.23	28.3	smokes	1		
31	69160	Male	59,0,0	Yes	Private	Rural	211.78	N/A	formerly smoked	1		
32	43717	Male	57,1,0	Yes	Private	Urban	212.08	44.2	smokes	1		
33	33879	Male	42,0,0	Yes	Private	Rural	83.41	25.4	Unknown	1		
34	39373	Female	82,1,0	Yes	Self-employed	Urban	196.92	22.2	never smoked	1		
healthcare-dataset-stroke-data												

Gambar 3.2 Data Selection

Data yang tidak relevan atau mengandung nilai yang tidak sesuai akan dihapus agar tidak mengganggu analisis selanjutnya.

b) Data Integration

Proses integrasi dilakukan untuk mengubah skema struktur data dari bentuk mentah ke bentuk yang lebih terorganisir agar dapat diproses lebih lanjut pada tahap data preprocessing. Sebelum dilakukan integrasi, dataset masih dalam format awal dengan struktur yang belum optimal untuk analisis.

	A	B	C	D	E	F	G	H	I	J	K	
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		
22	70630	Female	71,0,0	Yes	Govt_job	Rural	193.94	22.4	smokes	1		
23	13861	Female	52,1,0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1		
24	68794	Female	79,0,0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1		
25	64778	Male	82,0,1	Yes	Private	Rural	208.3	32.5	Unknown	1		
26	4219	Male	71,0,0	Yes	Private	Urban	102.87	27.2	formerly smoked	1		
27	70822	Male	80,0,0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1		
28	38047	Female	65,0,0	Yes	Private	Rural	100.98	28.2	formerly smoked	1		
29	61843	Male	58,0,0	Yes	Private	Rural	189.84	N/A	Unknown	1		
30	54827	Male	69,0,1	Yes	Self-employed	Urban	195.23	28.3	smokes	1		
31	69160	Male	59,0,0	Yes	Private	Rural	211.78	N/A	formerly smoked	1		
32	43717	Male	57,1,0	Yes	Private	Urban	212.08	44.2	smokes	1		
33	33879	Male	42,0,0	Yes	Private	Rural	83.41	25.4	Unknown	1		
34	39373	Female	82,1,0	Yes	Self-employed	Urban	196.92	22.2	never smoked	1		
healthcare-dataset-stroke-data +												

Gambar 3.3 Data Sebelum Integrasi

Sebelum integrasi, jumlah data yang tersedia sebanyak 5.110 entri. Setelah proses integrasi dilakukan, jumlah data dapat berubah tergantung pada proses pembersihan dan penggabungan yang dilakukan. Dalam kasus ini, setelah integrasi jumlah data akhir yang tersimpan dan siap untuk tahap preprocessing adalah 5.110 entri, tetap sama karena tidak ada pengurangan atau penambahan data, hanya perubahan struktur dan format penyimpanan seperti pada gambar 3.4.

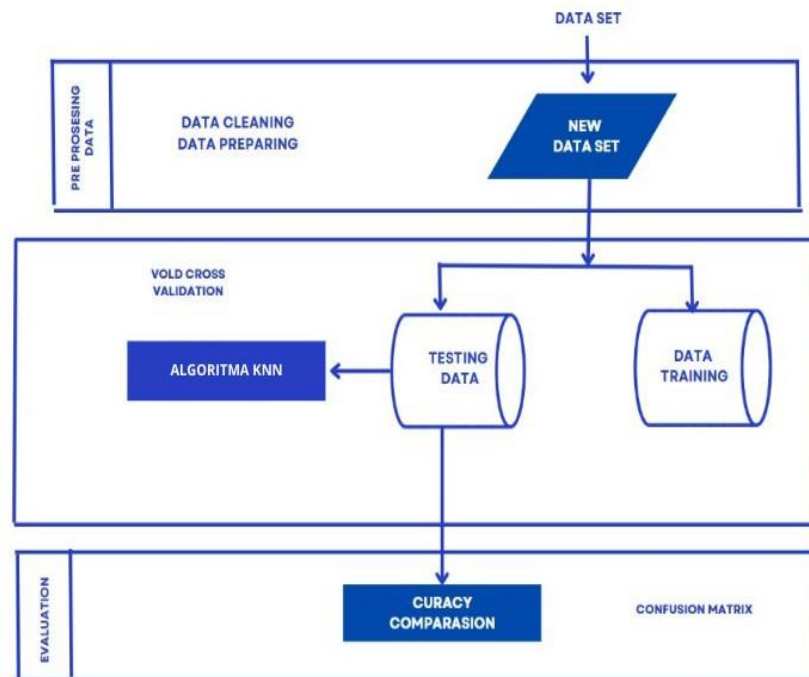
id	gender	age	hypertens	heart_dis	ever_mar	work_type	Residence	avg_glucose	bmi	smoking_status	stroke
9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoke	1
51676	Female	61	0	0	Yes	Self-empl	Rural	202.21	N/A	never smoked	1
31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
1665	Female	79	1	0	Yes	Self-empl	Rural	174.12	24	never smoked	1
56669	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoke	1
53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
10434	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
27419	Female	59	0	0	Yes	Private	Rural	76.15	N/A	Unknown	1
60491	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A	Unknown	1
5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
58202	Female	50	1	0	Yes	Self-empl	Rural	167.41	30.9	never smoked	1
56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
25226	Male	57	0	1	No	Govt_job	Urban	217.08	N/A	Unknown	1
70630	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
13861	Female	52	1	0	Yes	Self-empl	Urban	233.29	48.9	never smoked	1
68794	Female	79	0	0	Yes	Self-empl	Urban	228.7	26.6	never smoked	1
64778	Male	82	0	1	Yes	Private	Rural	208.3	32.5	Unknown	1
4219	Male	71	0	0	Yes	Private	Urban	102.87	27.2	formerly smoke	1
70822	Male	80	0	0	Yes	Self-empl	Rural	104.12	23.5	never smoked	1
38047	Female	65	0	0	Yes	Private	Rural	100.98	28.2	formerly smoke	1
61843	Male	58	0	0	Yes	Private	Rural	189.84	N/A	Unknown	1
54827	Male	69	0	1	Yes	Self-empl	Urban	195.23	28.3	smokes	1
69160	Male	59	0	0	Yes	Private	Rural	211.78	N/A	formerly smoke	1

Gambar 3.4 Data Sesudah Integration

Pada tahap ini, beberapa dataset yang memiliki atribut serupa atau berkaitan akan digabungkan, sementara atribut yang tidak relevan atau redundan dihapus untuk menjaga konsistensi data. Selain itu, tipe file yang sebelumnya dalam format .csv diekspor menjadi format Excel (.xlsx) untuk mempermudah proses analisis dan manipulasi data di tahap berikutnya.

4. Proses mining Algoritma KNN

Penelitian ini akan dilakukan dengan menerapkan dua metode yaitu Algoritma KNN dan *Optimasi Teknik Bagging* untuk penyakit Stroke. Alur tahapan penelitian yang dibuat sebagai kerangka kerja Algoritma KNN dapat dilihat seperti yang ditunjukkan pada gambar 3.5.



Gambar 3.5 Alur Penelitian KNN

- Tahapan pertama yaitu pengumpulan dataset Stroke yang diambil dari Kaggle repository.
- Berikutnya Preprosesing data yang bertujuan untuk membersihkan, merapikan, dan mempersiapkan data mentah sebelum diolah lebih lanjut.
- Selanjutnya proses klasifikasi menggunakan algoritma KNN
- Pengujian data yang digunakan dalam algoritma KNN adalah dataset prediksi stroke yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada gambar 3.6.

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
9046	1	67	0	1	1	0	1	229	37	3	1
51676	0	61	0	0	1	1	0	202	0	2	1
31112	1	80	0	1	1	0	0	106	33	2	1
60182	0	49	0	0	1	0	1	171	34	1	1
1665	0	79	1	0	1	1	0	174	24	2	1
56669	1	81	0	0	1	0	1	186	29	3	1
53882	1	74	1	1	1	0	0	70	27	2	1
10434	0	69	0	0	0	0	1	94	23	2	1
27419	0	59	0	0	1	0	0	76	0	0	1
60491	0	78	0	0	1	0	1	59	24	0	1
12109	0	81	1	0	1	0	0	80	30	2	1
12095	0	61	0	1	1	2	0	120	37	1	1
12175	0	54	0	0	1	0	1	105	27	1	1
8213	1	78	0	1	1	0	1	220	0	0	1
5317	0	79	0	1	1	0	1	214	28	2	1
58202	0	50	1	0	1	1	0	167	31	2	1
56112	1	64	0	1	1	0	1	192	38	1	1
34120	1	75	1	0	1	0	1	221	26	1	1
27458	0	60	0	0	0	0	1	89	38	2	1
25226	1	57	0	1	0	2	1	217	0	0	1
70630	0	71	0	0	1	2	0	194	22	1	1
13861	0	52	1	0	1	1	1	233	49	2	1

Gambar 3.6 Sampel Dataset

Tahap 1: Melakukan perhitungan Normalisasi Data guna untuk menyelaraskan data dengan interval angka 0 -1 untuk hasil perhitungan seperti dibawah ini.

Tabel 3.1 Normalisasi Data

Max	1	81	1	1	1	1	2	1	233,29	48,9	3	
Min	0	49	0	0	0	0	0	0	58,57	0	0	
id	gender	age	hypertension	heart_disease	ever_married	work_type	esidence_tpy	vg_glucose_lev	bmi	smoking_status	stroke	
9046	1	0,6	0	1	1	0	1	0,97367216	0,7484663	1	1	
51676	0	0,4	0	0	1	0,5	0	0,82211538	0	0,666666667	1	
31112	1	1	0	1	1	0	0	0,27100504	0,6646217	0,666666667	1	
60182	0	0	0	0	1	0	1	0,64480311	0,7034765	0,333333333	1	
1665	0	0,9	1	0	1	0,5	0	0,66134386	0,4907975	0,666666667	1	
56669	1	1	0	0	1	0	1	0,73054029	0,593047	1	1	
53882	1	0,8	1	1	1	0	0	0,06593407	0,5603272	0,666666667	1	
10434	0	0,6	0	0	0	0	1	0,20501374	0,4662577	0,666666667	1	
27419	0	0,3	0	0	1	0	0	0,10061813	0	0	1	
60491	0	0,9	0	0	1	0	1	0	0,4948875	0	1	
12109	0	1	1	0	1	0	0	0,12511447	0,607362	0,666666667	1	
12095	0	0,4	0	1	1	1	0	0,3542239	0,7525562	0,333333333	1	
12175	0	0,2	0	0	1	0	1	0,26293498	0,5582822	0,333333333	1	
8213	1	0,9	0	1	1	0	1	0,92301969	0	0	1	
5317	0	0,9	0	1	1	0	1	0,89010989	0,5766871	0,666666667	1	
58202	0	0	1	0	1	0,5	0	0,62293956	0,6319018	0,666666667	1	
56112	1	0,5	0	1	1	0	1	0,76144689	0,7668712	0,333333333	1	
34120	1	0,8	1	0	1	0	1	0,93131868	0,5276074	0,333333333	1	
27458	0	0,3	0	0	0	0	1	0,17542353	0,7730061	0,666666667	1	
25226	1	0,3	0	1	0	1	1	0,90722299	0	0	1	
70630	0	0,7	0	0	1	1	0	0,77478251	0,4580777	0,333333333	1	
13861	0	0,1	1	0	1	0,5	1	1	1	0,666666667	1	

Tahap 2: Melakukan Normalisasi pada data uji seperti tahap 1 untuk hasil perhitungan seperti dibawah ini:

Tabel 3.2 Normalisasi Data Uji

id	gender	age	hypertension	heart_disease	ver_married	rk_type	residence_type	glucose_le	bmi	moking_statu	stroke
64778	1	1	0	1	0	0	0	1	1	0	1
4219	1	1	0	0	0	0	1	0	0,8369231	1	1
70822	1	1	0	0	0	1	0	0	0,7230769	0,666666667	1
38047	0	1	0	0	0	0	0	0	0,8676923	1	1
61843	1	1	0	0	0	0	0	1	0	0	1
54827	1	1	0	1	0	1	1	1	0,8707692	0,333333333	1

Tahap 3: Menghitung Euclidean Distance pada masing-masing data uji untuk hasil perhitungan seperti dibawah ini.

Tabel 3.3 Euclidean Distance

Euclidean Distance					
Data uji 1	Data uji 2	Data uji 3	Data uji 4	Data uji 5	Data uji 6
1,804264	1,76445	2,279272	2,270167	2,184681	1,668808
2,262015	2,258186	1,947328	1,783396	1,756446	2,29415
1,445453	1,790526	1,750096	1,796321	1,788121	2,112058
2,30764	1,963574	2,343158	1,971594	2,153961	2,020512
2,252182	2,213416	1,925459	1,715938	1,991734	2,536026
2,058694	1,252096	1,902581	1,899768	1,833358	2,131521
1,886289	2,058593	2,018832	2,06344	2,094515	2,340435
2,121781	1,193419	1,799297	1,206199	1,786216	2,029425
2,298161	2,275958	2,108325	1,798797	1,732549	2,563525
2,294325	1,768075	2,122772	1,774196	1,98476	2,417633
2,316038	2,043371	2,005637	1,481397	2,07542	2,783461
1,994992	2,440388	1,899528	1,993439	2,30055	1,851801

2,358263	1,814988	2,214675	1,822424	2,110448	2,121469
1,736293	2,128143	2,403099	2,571701	1,737187	1,921626
1,907779	1,985949	2,401333	1,997998	2,187413	2,01907
2,431171	2,389151	2,132896	1,934745	2,252613	2,341861
1,582591	1,812421	2,220666	2,305783	1,996461	1,498051
2,091444	1,846656	2,235663	2,337223	1,85346	2,18649
2,14629	1,253302	1,858634	1,257583	1,97446	1,932032
1,889737	2,248188	2,07379	2,671787	1,889121	0,96576
2,133816	2,293236	1,683686	1,819388	1,849712	2,157236
2,552593	2,274661	2,467856	2,28033	2,558382	2,097959

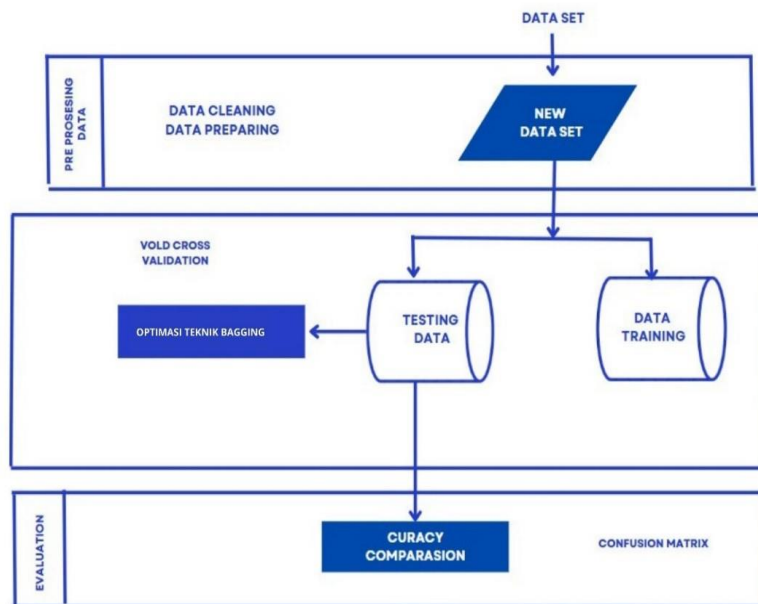
Tahap 4: Menentukan nilai k dengan mengambil nilai terkecil dari distance sesuai k untuk hasil perhitungan seperti table 3.4.

Tabel 3.4 Nilai K

Nilai <i>k</i>					
Uji ke 1 10	Uji ke 2 10	Uji ke 3 10	Uji ke 4 10	Uji ke 5 10	Uji ke 6 10
1	1				1
		1	1	1	
1	1	1	1	1	
	1				1
		1	1		
1	1	1		1	
1		1			
	1	1	1	1	1
			1	1	
	1		1		
		1	1		
1		1			1
	1		1		
1				1	1
1					1
1	1				1
1	1			1	
	1	1	1	1	1
1				1	1
		1	1	1	
					1

5. Proses mining *Optimasi Teknik Bagging*

Alur tahapan penelitian dibuat sebagai kerangka kerja dengan *Optimasi Teknik Bagging* dapat dilihat seperti yang ditunjukkan pada gambar 3.6.



Gambar 3.6. Alur Penelitian Menggunakan *Optimasi Teknik Bagging*

- Tahapan pertama yaitu pengumpulan dataset Stroke yang diambil dari Kaggle.
- Berikutnya Preprosesing data yang bertujuan untuk membersihkan, merapikan, dan mempersiapkan data mentah sebelum diolah lebih lanjut.
- Selanjutnya proses klasifikasi menggunakan *Optimasi Teknik Bagging*
- Pengujian data prediksi stoke yang digunakan dalam *Optimasi Teknik Bagging*.

3.1 Studi Literatur

Studi literatur adalah serangkaian kegiatan yang berkenaan dengan metode pengumpulan data pustaka, membaca dan mencatat, serta mengelolah bahan penelitian. Studi Literatur adalah merupakan penelitian yang dilakukan oleh peneliti dengan mengumpulkan sejumlah buku buku, majalah yang berkaitan dengan masalah dan tujuan penelitian [12]

Teknik ini dilakukan dengan tujuan untuk mengungkapkan berbagai teori-teori yang relevan dengan permasalahan yang sedang dihadapi/diteliti sebagai bahan rujukan dalam pembahasan hasil penelitian. Pengertian Lain tentang Studi literatur adalah mencari referensi teori yang relefan dengan kasus atau permasalahan yang ditemukan.

Secara Umum Studi Literatur adalah cara untuk menyelesaikan persoalan dengan menelusuri sumber-sumber tulisan yang pernah dibuat sebelumnya. Dengan kata lain, istilah Studi Literatur ini juga sangat familier dengan sebutan studi pustaka. Dalam sebuah penelitian yang hendak dijalankan, tentu saja seorang peneliti harus memiliki wawasan yang luas terkait objek yang akan diteliti. Jika tidak, maka dapat dipastikan dalam persentasi yang besar bahwa penelitian tersebut akan gagal.

3.2 Jadwal Penelitian

Penelitian ini dilakukan di kampus Institut Informatika dan Bisnis Darmajaya Bandar Lampung dengan perkiraan periode penelitian dari bulan Juni

sampai dengan Desember 2024. Pada Tabel 3.6 merupakan jadwal dan tahapan penelitian tugas akhir yang dilakukan.

Tabel 3.5. *Timeline* penelitian tugas akhir

Jenis Kegiatan	Bulan																											
	Juni				Juli				Agustus				September				Oktober				November				Desember			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Persiapan Penelitian																												
Perencanaan																												
Pengumpulan Data																												
Seminar Proposal																												
Pengolahan Data																												
Pembuatan Laporan Hasil																												

3.3 Alat dan bahan penelitian

Dalam penelitian ini, untuk melakukan penambangan data, dibutuhkan perangkat keras dan lunak untuk menunjang kelancaran dalam proses penambangan. Perangkat tersebut diantaranya:

1. Perangkat keras yang digunakan untuk melakukan penambangan data adalah sebagai berikut:
 - a. Processor core i3
 - b. RAM 4 Gb
 - c. SSD 256
2. Perangkat lunak yang digunakan untuk melakukan penambangan data adalah sebagai berikut:
 - a. Sistem operasi windows 10
 - b. Software Rapid Miner
 - c. Bahan yang digunakan diambildari

<https://www.kaggle.com/code/docxian/stroke-prediction>

