

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Pada penelitian ini hal penting yang akan penulis lakukan yaitu mempersiapkan bahan penelitian, dalam hal ini penulis akan menggunakan *dataset* berupa data kanker serviks yang di dapat dari *website UCI Machine Learning Repository*. Sehingga data tersebut dapat penulis klasifikasikan dengan menerapkan algoritma KNN dan C4.5.

3.2 Peralatan Penelitian

Penelitian ini akan dilakukan dengan menggunakan peralatan yang terdiri dari perangkat keras dan perangkat lunak.

1. Perangkat keras yang digunakan:
 - a. Laptop *Intel(R) Celeron(R) CPU N2830 @ 2.16GHz 2.16 GHz*
 - b. RAM 4GB
 - c. HDD 500GB
 - d. SSD 128GB
2. Perangkat lunak yang digunakan:
 - a. *64-bit windows 10 pro operating system*
 - b. *Browser*
 - c. *Microsoft Word 2013*
 - d. *Microsoft Excel 2013*
 - e. *RapidMiner*
 - f. *Google Colabatory*

3.3 Lokasi dan Waktu Penelitian

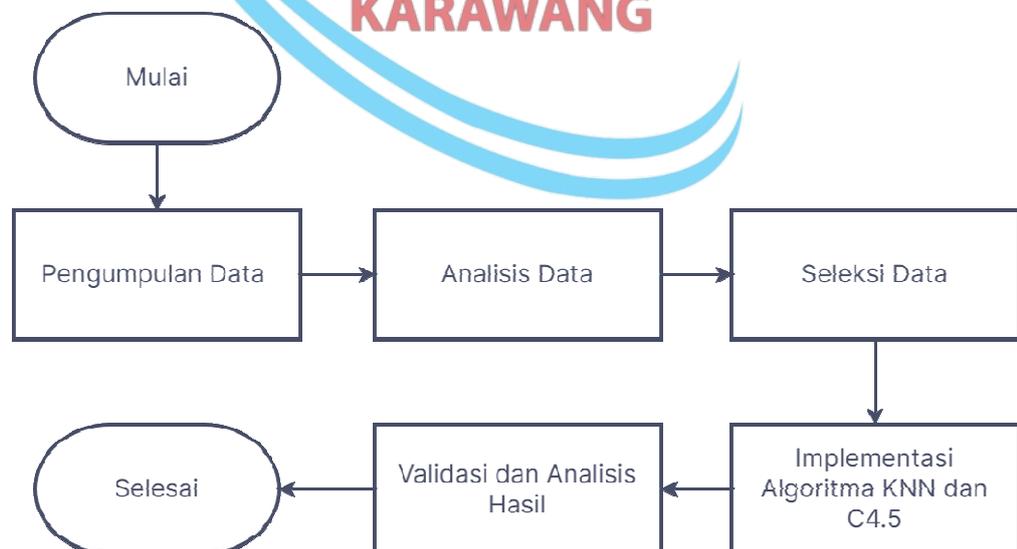
Lokasi penelitian ini akan dilakukan dirumah, karena untuk mematuhi protokol kesehatan dimasa pandemi Covid-19, sejak bulan Desember 2021. Adapun waktu penelitian terdapat pada tabel dibawah ini:

Tabel 3.1 Waktu Penelitian

No.	Kegiatan	Bulan							
		Des 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	Mei 2022	Jun 2022	Jul 2022
1	Analisis Kebutuhan	■	■						
2	Pengumpulan Data	■	■						
3	Analisis Data			■	■				
4	Seleksi Data				■				
5	Pengolahan Data					■	■	■	
6	Validasi dan Analisis Hasil								■

3.4 Prosedur Penelitian

Dalam penelitian ini terdapat tahapan-tahapan yang harus dipahami dan dilakukan. Terdapat pada gambar 3.1 Alur penelitian.



Gambar 3.1 Alur Penelitian

3.4.1 Pengumpulan Data

Pengumpulan data adalah tahapan yang bertujuan untuk memperoleh informasi maupun data yang berkaitan dengan penelitian. Bahan atau data yang

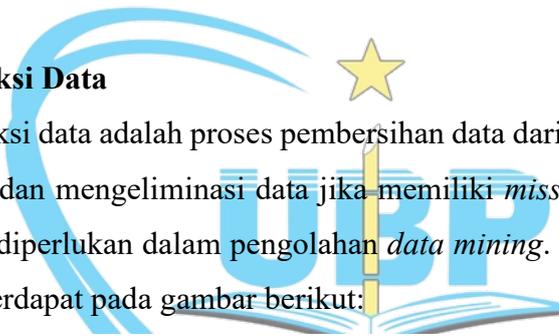
akan digunakan dalam penelitian ini yaitu *dataset* yang diperoleh dari *website UCI Machine Learning Repository* oleh DR. Sobar, Prof. Rizanda Machmud, dan Adi Wijaya, PhD candidate. Yang berisi data resiko kebiasaan penyakit kanker serviks dengan jumlah total atribut 19 dari 72 data, sehingga diharapkan dapat mendukung untuk proses klasifikasi penyakit kanker serviks.

3.4.2 Analisis Data

Analisis data bertujuan untuk memahami data agar dapat diseleksi dan dikelola. Sehingga memudahkan dalam menentukan kesalahan atau *error* dalam mengakurasi algoritma KNN dan C4.5 terhadap klasifikasi penyakit kanker serviks.

3.4.3 Seleksi Data

Seleksi data adalah proses pembersihan data dari atribut-atribut yang tidak diperlukan dan mengeliminasi data jika memiliki *missing value/duplicate value* yang tidak diperlukan dalam pengolahan *data mining*. Data sebelum dan setelah di seleksi terdapat pada gambar berikut:



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
10	10	10	10	4	7	0	10	1	0	7	5	10	4	0	0	0	0	0
11	10	10	10	10	10	7	0	0	10	10	0	0	10	10	10	10	10	10
12	10	10	0	0	10	0	10	1	0	0	0	0	0	0	0	0	0	0
13	10	10	10	10	10	7	0	1	0	0	0	0	10	10	10	10	10	10
14	0	10	0	0	10	0	0	1	0	0	0	0	10	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	10	10	10	10	10	0	0	0	10	10	10	10	10	10	10	10	10	10
17	10	10	10	10	10	0	0	0	10	10	10	10	10	10	10	10	10	10
18	10	0	10	0	10	0	0	0	10	10	0	0	10	10	10	10	10	10
19	0	10	10	10	10	10	0	0	10	10	10	10	10	10	10	10	10	10
20	10	10	10	10	0	10	0	0	10	10	10	10	10	10	10	10	10	10

Gambar 3.2 Data Sebelum di Seleksi

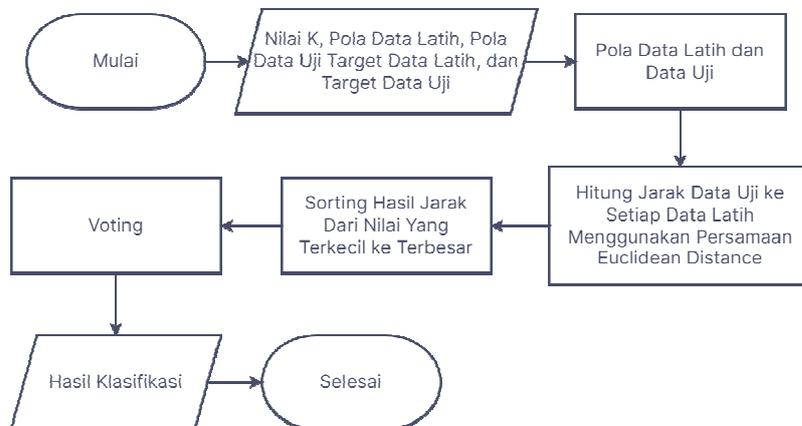
id	duk_sosial_instrumental	pengetahuan_pemberdayaan	kemampuan_pemberdayaan	keinginan_pemberdayaan	dukungan_sosial_instrumental
0	12	12	11	2	positif
1	0	8	6	2	positif
2	11	3	3	15	positif
3	4	2	1	8	positif
4	12	5	4	7	positif
..	..	-	-
15	12	12	11	2	negatif
16	11	11	11	11	negatif
17	11	12	11	11	negatif
18	12	11	11	11	negatif
19	11	11	11	11	negatif

Gambar 3.3 Data Setelah di Seleksi

Pada gambar 3.3 terdapat hasil seleksi berdasarkan data yang memiliki *missing value* atau data yang tidak lengkap memiliki nilai *default* dengan menggunakan nilai median. Sehingga data yang sebelumnya sebanyak 72 data dan 19 atribut menjadi 63 data dan 5 atribut yang diantaranya adalah dukungan sosial instrumental, pengetahuan pemberdayaan, kemampuan pemberdayaan dan keinginan pemberdayaan lalu untuk kanker serviks dijadikan untuk atribut kelas jika data bernilai 1 maka data dinyatakan positif jika 0 maka dinyatakan negatif.

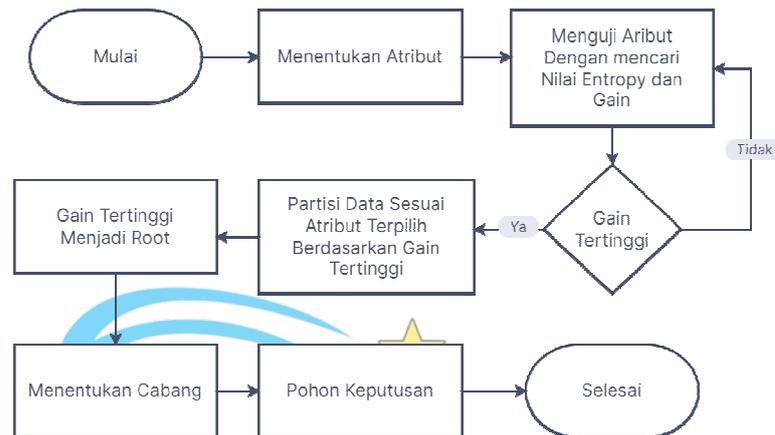
3.4.4 Penerapan Algoritma KNN dan C4.5

Pada proses ini yaitu menerapkan algoritma KNN, dengan penghitungan menggunakan *platform RapidMiner* dan program *Python*. Selain menerapkan pada *platform*, pada tahap ini juga akan dilakukan penghitungan secara manual dengan menggunakan rumus algoritma KNN dan C4.5.



Gambar 3.4 Flowchart KNN

Gambar 3.4 Merupakan cara kerja algoritma KNN, yang diawali dengan menentukan nilai k , menentukan pola data latih/data uji, dan menentukan target data latih/data uji. Kemudian proses pola data uji dan data latih. Lalu hitung jarak data uji ke data latih dengan *Euclidean Distance*. Selanjutnya urutkan dari jarak yang tekecil hingga terbesar. Kemudian dilakukan voting dan kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi.



Gambar 3.5 Flowchart C4.5

Gambar 3.5 Merupakan cara kerja algoritma C4.5, yang diawali dengan menentukan atribut yang akan digunakan. Kemudian mencari *gain* tertinggi berdasarkan hasil penghitungan *entropy* dari masing masing atribut agar dapat dilakukan pengujian. Apabila ditemukan *gain* tertinggi, maka *gain* tersebut akan menjadi *root* awal. Selanjutnya mencari *gain* tertinggi dari hasil setiap partisi untuk menentukan cabang. Kemudian pembuatan pohon keputusan untuk mengetahui hasil dari klasifikasi.

3.4.5 Validasi dan Analisis Hasil

Pada tahap validasi ini diambil dari hasil akurasi algoritma KNN dan C4.5. kemudian algoritma yang mempunyai nilai akurasi yang tinggi merupakan algoritma terbaik pada proses klasifikasi ini. untuk mendapatkan nilai akurasi dalam klasifikasi kanker serviks, digunakan persamaan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} * 100\% \quad (7)$$

3.4.6 Metode Analisis

Pada penelitian ini metode algoritma yang digunakan adalah *K-Nearest Neighbor* (KNN) dan C4.5. Jenis penelitian ini adalah melakukan pengujian dengan tujuan untuk menciptakan pengetahuan baru dari teori-teori yang sudah ada. Penerapan pada penelitian ini adalah dimana *dataset* yang digunakan untuk pengujian bersumber dari *website UCI Machine Learning Repository* yang berisi data resiko kebiasaan penyakit kanker serviks, kemudian dilakukan pengujian dengan metode algoritma KNN dan C4.5 agar dapat memperbandingkan hasil akurasi terbaik dari kedua algoritma tersebut dalam mengklasifikasi data resiko kebiasaan penyakit kanker serviks.

