

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek Penelitian ini Dapodik SD dan SMP di Karawang, dengan mengambil data Dapodik SD dan SMP pada Tahun Ajaran. 2020/2021. Data yang diambil merupakan Data Dapodik SD dan SMP selama 1 Tahun Ajaran

Penelitian ini dilaksanakan di Dinas Pendidikan Pemuda dan Olahraga Kabupaten Karawang di Jalan. Surotokunto No.72, Adiarsa Timur, Kec. Karawang Timur, Kabupaten Karawang, Jawa Barat 41311. Adapun waktu pelaksanaan penelitian ini dimulai dari bulan Desember 2021 sampai dengan Juli 2022. Perincian waktu penelitian terdapat pada Tabel berikut.

Tabel 3.1. Waktu Penelitian

Kegiatan	Des-21				Jan-22				Feb-22			
	1	2	3	4	1	2	3	4	1	2	3	4
Studi Literatur												
Pengumpulan Data												
Proposal												
Kegiatan	Apr-22				Mei-22				Jun-22			
	1	2	3	4	1	2	3	4	1	2	3	4
Preprocessing												
Perhitungan Manual												
Perhitungan Program Python												
Evaluasi												

3.2 Peralatan Penelitian

1. *Hardware* yang digunakan dalam penelitian ini menggunakan laptop dengan spesifikasi :

a. Processor : 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz
2.50GHz

b. *Memory Size* : 8GB RAM

2. *Software*

a. Windows 11

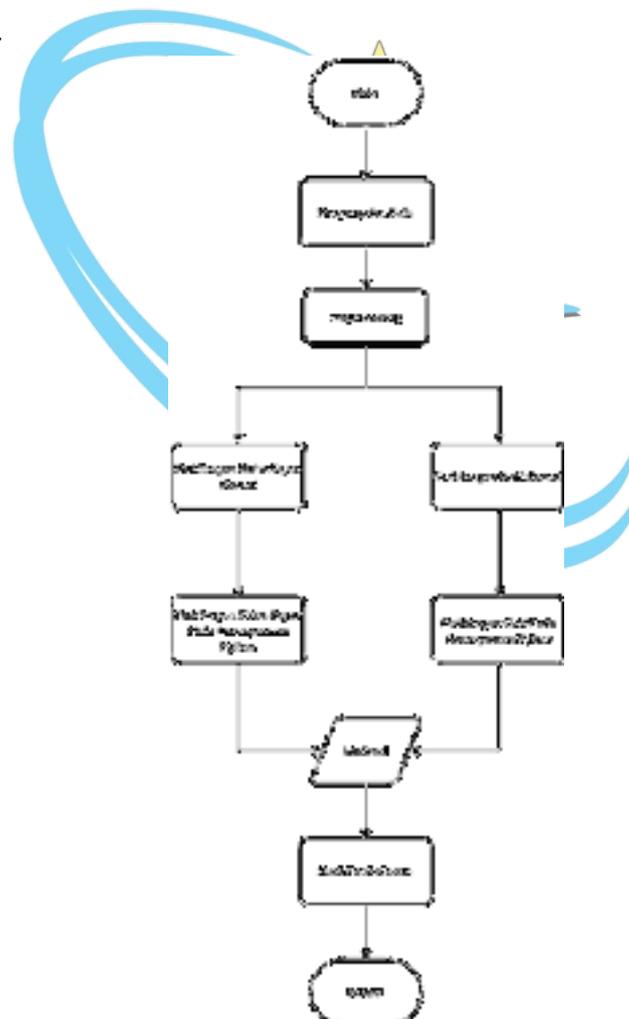
b. Microsoft Word 2021

c. Microsoft Excel 2021

d. Python

3.3 Prosedur Penelitian

Tahapan penelitian Tugas Akhir yang akan dilakukan pertama adalah mulai merancang topik yang akan digunakan. Lalu, kedua adalah Pengumpulan data mengambil data yang diperlukan saja. Ketiga adalah Preprocessing data menghapus beberapa atribut yang tidak digunakan. keempat adalah melakukan perhitungan *Naïve Bayes* dan *C.45* manual . Lalu, kelima melakukan perhitungan *C.45* program python. keenam melakukan perhitungan program python dengan masing – masing algoritma yang berbeda, jika sudah lanjut ke evaluasi menghitung *Confusion Matrix* dari masing – masing perhitungan lalu terlihat hasil akurasi nya. keenam adalah selesai.



Gambar 3.1. Alur Prosedur Penelitian

3.4 Pengumpulan Data

Penelitian ini dimulai dengan melakukan pengumpulan data Dapodik SD, SMP tahun ajaran 2020 – 2021. Peneliti mengumpulkan data, dengan mendatangi Dinas Pendidikan Pemuda dan Olahraga bagian program dan pelaporan. Data yang diperoleh merupakan campuran sekolah SD dan SMP Kabupaten Karawang, Data terkumpul sebanyak 100 sekolah. Variabel data terdiri dari No, NPSN, Nama Sekolah, Ruang Kelas, Meja Siswa, Kursi Siswa, Peserta Didik.

3.5 Preprocessing Data

Pada Preprocessing Data terdapat perubahan Data Numerik menjadi Kategorikal sekaligus reduksi pada variabel No, NPSN, Nama Sekolah. Hal ini dilakukan karena variabel-variabel tersebut tidak sebagai faktor penentu dalam penghitungan Sarana Prasarana ini. Sedangkan variabel Ruang Kelas, Meja Siswa, Kursi Siswa, Peserta Didik. Setelah tahap ini selesai maka akan menghasilkan data set yang dapat digunakan pada perhitungan manual dan penerepan algoritma *Naive Bayes* dengan *C.45* pada aplikasi Python.

3.6 Perhitungan Manual

Pada proses perhitungan manual penulis menggunakan aplikasi Microsoft Excel 2021 dan mengimplementasikan algoritma Naive Bayes dalam melakukan klasifikasi. Data akan dikelompokkan menjadi 3 yaitu Banyak, Sedang, Sedikit.

1. Perhitungan Manual *Naive Bayes*

Langkah - langkah dalam algoritma *Naive Bayes* sebagai berikut.

- a. Mempersiapkan data set atau data latih yang akan dilakukan perhitungan probabilitas dengan algoritma *naive bayes*.
- b. Kemudian lakukan perhitungan probabilitas kelas
- c. Setelah itu hitung nilai probabilitas dari seluruh atribut yang ada dalam data dengan rumus *Naive Bayes*.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X

$P(X|H)$ = Probabilitas X berdasarkan kondisi H

$P(H)$ = Probabilitas hipotesis H

$P(X)$ = Probabilitas X

- d. Melakukan perhitungan prediksi terhadap data testing menggunakan model *Naïve Bayes*.

2. Perhitungan Manual C.45

- a. Menentukan atribut sebagai akar dan menghitung nilai informasi *gain* atribut. Untuk memilih atribut sebagai akar, perlu menghitung *gain* menggunakan rumus.

$$\text{Gain}(S, A) = \text{Entropy}(s) - \sum_{i=1}^s \frac{|S_i|}{|S|} * \text{Entropy}(s) \quad [2]$$

- b. Sebelum mendapatkan nilai *gain* maka di cari terlebih dahulu nilai *entropy*, untuk mencari nilai *entropy* menggunakan rumus.

$$\text{Entropy}(S) = \sum_{i=0}^n - p_i * \log_2 p_i \quad [3]$$

- c. Kemudian menghitung tiap atribut *entropy* yaitu jumlah ruang kelas, meja siswa, kursi siswa, peserta didik, perhitungan dilakukan sampai tidak ada atribut didalam *record* yang dipisahkan, tidak terdapat *record* cabang yang kosong.
- d. Jika sudah mendapatkan nilai *entropy* kemudian dilakukan perhitungan nilai *gain* tiap atribut, *gain* (total ruang kelas), *gain* (total meja siswa), *gain* (total kursi siswa), *gain* (total peserta didik).

3.7 Perhitungan Program Python

Setelah melakukan perhitungan manual peneliti akan membandingkan hasilnya dengan perhitungan menggunakan Program Python dengan algoritma *Naive Bayes* dan C.45, sehingga dapat dilihat perbandingannya dengan perhitungan manual.

3.8 Evaluasi

Proses evaluasi pada penelitian ini dengan menggunakan metode *Confusion Matrix*. Metode ini merupakan sebuah metode evaluasi cara kerjanya mengukur kinerja dalam bentuk akurasi, presisi serta *recall*. Pada tahapan inilah hasil dari proses klasifikasi akan dilihat tingkat kebenarannya.

Setelah perhitungan klasifikasi selesai dilakukan selanjutnya akan

dievaluasi *Confusion Matrix* untuk mengetahui performa akurasi, presisi dan *recall*, dengan rumus sebagai berikut.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (1)$$

$$\text{Presisi} = \frac{TP}{FP + TP} * 100\% \quad (2)$$

$$\text{Recall} = \frac{TP}{FN + TP} * 100\% \quad (3)$$

Keterangan :

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative.

