

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek pada penelitian tugas akhir ini adalah data yang diambil dari data siswa kelas X jurusan IPA dan IPS tahun ajaran 2019 sebanyak 321 data siswa yang digunakan. Lokasi ini mengambil data nya dari SMA Negeri 2 Karawang.

Tabel 3.1 Waktu Penelitian

Kegiatan	Des 2021	Jan 2022	Feb 2022	Mar 2022	April 2022	Mei 2022	Juni 2022	Juli 2022
Pengumpulan Data	■							
Seleksi Data		■						
Perhitungan Manual excel dan Python		■	■	■	■	■		
Evaluasi						■	■	■

3.2 Peralatan Penelitian

1. *Hardware* yang digunakan dalam penelitian ini menggunakan laptop dengan spesifikasi *Processor* AMD A9-9425 *RADEON* R5, 5 **KARAWANG** COMPUTE CORES 2C+3G 3.10 GHz, *Memory Size* 4.00 GB:
2. *Software*
 - a. Windows 10
 - b. Microsoft Word 2010
 - c. Microsoft Excell 2010
 - d. Aplikasi Google Colab

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 akan diperlukan saja. Ketiga Seleksi Data adalah menghapus data atau mengambil data yang akan digunakan saja. Keempat adalah melakukan perhitungan *Microsoft Excell* dan Algoritma *Naïve Bayes* manual. Lalu, Kelima melakukan perhitungan *Algoritma Naïve Bayes* Pemrograman *Python*. Keenam adalah selesai.



Gambar 3.1 Alur Prosedur Penelitian

3.3.1 Pengumpulan Data

Tahapan pertama yang dilakukan dalam melaksanakan penelitian ini adalah dengan melakukan pengumpulan data melalui salah satu sekolah dan mendapatkan data-data siswa tentang penerapan data *mining* untuk memprediksi penjurusan siswa. *Dataset* merupakan sebuah kumpulan data yang diambil dari beberapa informasi yang dikelola menjadi sebuah informasi untuk melakukan teknik pada ilmu data *mining*.

Table 3.2 Variabel Data

No	Variabel
1	Nama
2	Nilai Mtk Semester 1
3	Nilai Fisika Semester 1
4	Nilai Mtk Semester 2
5	Nilai Fisika Semester 2
6	Keterangan

3.3.2 Seleksi Data

Pada tahapan seleksi data ini merupakan proses pemilihan data set yang akan digunakan dan tidak digunakan dalam penelitian tersebut. Seleksi variabel dilakukan untuk menghapus yang digunakan untuk proses pengklasifikasian. Proses seleksi variabel dapat dilihat di tabel:

Table 3.3 Seleksi Variabel

Varibel	Digunakan	Keterangan
Nama	✓	Digunakan
Nilai Mtk Semester 1	✓	Digunakan
Nilai Fisika Semester 1	✓	Digunakan
Nilai Mtk Semester 2	✓	Digunakan
Nilai Fisika Semester 2	✓	Digunakan

3.3.3 Perhitungan Excel dan Perhitungan Rapidminer dan Python

Pada tahapan ini, peneliti menggunakan aplikasi *Microsoft Excel*, *Rapidminer*, dan *Python* untuk menerapkan algoritma Naïve Bayes. Disini data akan dibagi menjadi 2, yaitu jurusan IPA dan IPS. Setelah perhitungan menggunakan *Microsoft Excel* selesai selanjutnya akan melakukan perhitungan *Rapidminer* dan *Python*, selanjutnya peneliti akan membandingkan hasil perhitungan menggunakan *Microsoft Excel* dengan hasil perhitungan menggunakan *Rapidminer* dan *Python* kemudian akan mendapatkan hasil keakurasiannya.

3.3.4 Evaluasi

Setelah melakukan implementasi Naïve Bayes, maka akan dihasilkan dan evaluasi *Confusion Matrix* untuk bisa mengetahui hasil akurasi, presisi, dan *recall*. *Confusion Matrix* adalah bentuk dari hasil proses klasifikasi yang akan dilihat kebenarannya. *Confusion Matrix* untuk mengetahui performa akurasi, presisi, *recall* berikut ini adalah rumus *Confusion Matrix* sebagai berikut :

1. Akurasi

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \text{"Hasil"}$$

Akurasi adalah perbandingan prediksi yang betul (*Positif dan Negative*) dengan keseluruhan data.

2. Presisi

$$\text{Presisi} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \text{"Hasil"}$$

Presisi adalah perbandingan prediksi yang betul dibandingkan dengan keseluruhan hasilnya yang diprediksinya yaitu *positive*

3. Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \text{"Hasil"}$$

Recall adalah perbandingan dari prediksi betul positif dengan dibandingkan seluruh data yang betul *positif*

Keterangan Rumus *Confusion Matrix*:

1. TP (*True Positive*) yaitu jumlah data yang kelas aktual dan prediksi merupakan kelas yang positif.
2. TN (*True Negative*) yaitu banyaknya data yang kelas aktual merupakan kelas negative sedangkan kelas prediksi merupakan kelas *negative*.
3. FP (*False Negative*) yaitu banyaknya suatu data kelasnya aktual merupakan kelas negative sedangkan kelas prediksi merupakan kelas positif.

4. FN (*False Negative*) yaitu keseluruhan data yang kelas aktualnya merupakan kelas positif dan sedangkan kelas prediksi merupakan kelas *negative*.

