

BAB III

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Pada bab ini dijelaskan tahap-tahap yang dilakukan dalam melaksanakan penelitian. Tahapan penelitian yang dilakukan adalah sebagai berikut:

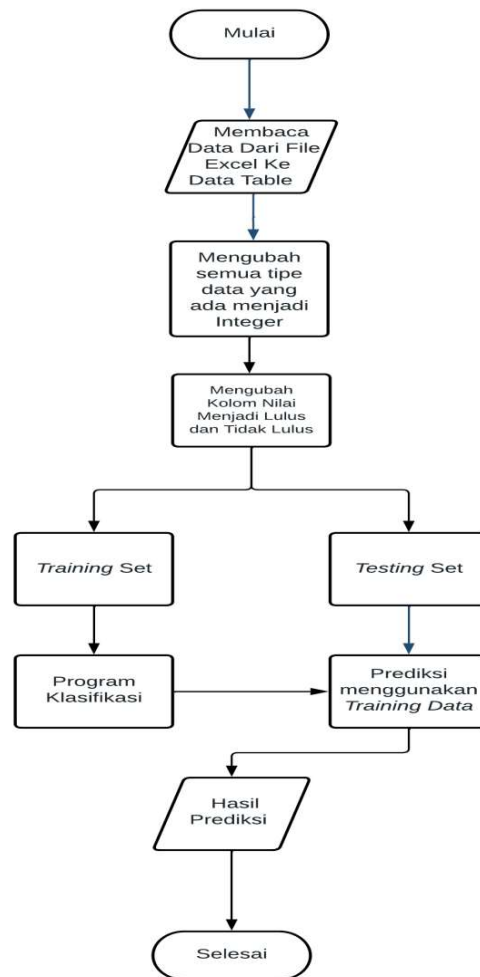
1. Melakukan studi literatur yaitu mengumpulkan referensi seperti artikel, buku, dan laporan penelitian yang berhubungan dengan permasalahan yang dibahas.
2. Melakukan pencarian *dataset* untuk dibuat penelitian performa siswa.
3. Melakukan transformasi data dari nilai angka menjadi sistem *binary classification* (*Passed* dan *Failed*).
4. Membuat program klasifikasi dengan empat algoritma *machine learning* (DT, KNN, SVM, dan NB).
5. Menarik kesimpulan dari hasil prediksi dan evaluasi performa dari setiap algoritma *machine learning* yang digunakan.

3.2 Instrumen Penelitian

Perangkat pendukung yang digunakan pada penelitian ini untuk membuat sistem klasifikasi adalah Jupyter Notebook. Jupyter Notebook merupakan sebuah aplikasi yang biasa digunakan untuk membuat program AI, *machine learning*, dan analisa data. Jupyter Notebook dapat bekerja dengan baik untuk membantu dalam mencari makna dan memberikan informasi penting yang terdapat pada data yang digunakan.

3.3 Perancangan Penelitian

Secara garis besar, perancangan penelitian untuk membuat sistem klasifikasi dibagi menjadi tiga, yaitu pembacaan data, mengubah nilai angka menjadi sistem *binary classification*, dan pembuatan sistem klasifikasi dengan *machine learning*. Diagram alir perancangan penelitian dapat dilihat pada Gambar 3.1 di bawah ini.



Gambar 3.1 Diagram Alir Penelitian

Berdasarkan Gambar 3.1 di atas, bisa dilihat tahapan penelitian yang dilakukan dari awal mulai hingga didapatkan nilai akurasi dari percobaan setiap algoritma. Tahapan penelitian seperti pada gambar di atas ini harus dilakukan dengan baik dan secara berurutan sehingga bisa didapatkan hasil prediksi setiap algoritma yang diharapkan. Perancangan model *machine learning* pada penelitian ini menggunakan tahapan sebagai berikut.

1. Melakukan pembacaan data penelitian.
2. Melakukan Transformasi data dari tipe *string* ke *integer*.
3. Melakukan pembagian *training* dan *testing* data.
4. Membuat sistem *tuning* parameter untuk mencari parameter terbaik dari masing-masing model *machine learning* (DT, KNN, SVM, dan NB).

5. Setelah didapatkan hasil parameter terbaik, selanjutnya parameter terbaik tersebut digunakan untuk membuat model *machine learning* yang dilatih dengan *training data* untuk setiap algoritma yang digunakan.
6. Hasil dari setiap model di atas kemudian digunakan untuk membuat prediksi dari *testing data*. Hasil dari model ini juga menghasilkan *confusion matrix*, *accuracy*, *precision*, *recall*, *F1score*, dan *MCC*.

3.4 Pengambilan *Dataset*

Dataset merupakan sekumpulan data yang berasal dari informasi-informasi di masa lalu dan disimpan dalam memori pada kondisi tidak terhubung. *Dataset* dapat ditambahkan, dihapus, atau memperbarui baris dalam memori sehingga siap untuk dikelola menjadi suatu informasi yang baru. *Dataset* terdiri dari dua jenis, yakni *private dataset* dan *public dataset*. *Private dataset* merupakan *dataset* yang diambil dari suatu organisasi untuk dijadikan sebagai objek penelitian, misalnya data rumah sakit, sekolah, bisnis dan sebagainya. Sedangkan *public dataset* merupakan *dataset* yang diambil dari *public repository* yang telah disepakati oleh pakar peneliti *data mining*.

Dataset yang digunakan dalam penelitian ini adalah *private dataset*. Data percobaan dikumpulkan dari siswa SMKN 4 Kota Tangerang. Hasilnya didapatkan satu *dataset* dengan data target yang berisi data nilai siswa SMKN 4 Kota Tangerang pada mata pelajaran Matematika, PKn, dan Bahasa Inggris.

3.5 *Dataset*

Dataset ini berisi data siswa yang berasal dari SMKN 4 Kota Tangerang. *Dataset* ini memiliki 32 variabel yang dapat digunakan untuk penelitian. Variabel yang digunakan pada penelitian ini merupakan variabel yang bisa digunakan untuk membuat prediksi performa siswa. Variabel dari faktor ekonomi, faktor keluarga, jenis kelamin, dan umur bisa digunakan untuk membuat prediksi performa siswa. Variabel nilai siswa pada mata pelajaran semester sebelumnya juga bisa digunakan untuk membuat prediksi performa siswa. Variabel yang digunakan ini dapat dilihat pada Tabel 3.1 di bawah ini.

Tabel 3.1 Variabel Penelitian

No	Variabel	Penjelasan
1	<i>Sex</i>	Jenis kelamin siswa (laki-laki atau perempuan)
2	<i>Age</i>	Umur siswa (numerik)
3	<i>Medu</i>	Tingkat edukasi ibu (1 s.d. 5)
4	<i>Fedu</i>	Tingkat edukasi ayah (1 s.d. 5)
5	<i>Studytime</i>	Waktu belajar siswa dalam satu minggu (numerik)
6	<i>Famrel</i>	Hubungan dalam keluarga (1-5)
7	<i>Freetime</i>	Waktu luang (numerik)
8	<i>Health</i>	Tingkat kesehatan siswa (numerik)
9	<i>Famsize</i>	Jumlah anggota keluarga (lebih kecil dari 3 atau lebih dari tiga)
10	<i>Pstatus</i>	Status tinggal dengan orang tua (ya atau tidak)
11	<i>Mjob</i>	Pekerjaan Ibu (numerik)
12	<i>Fjob</i>	Pekerjaan Ayah (numerik)
13	<i>reason</i>	Alasan memilih sekolah ini (nominal: dekat dengan rumah, reputasi sekolah, pembelajaran yang baik atau lainnya)
14	<i>schoolsup</i>	Mengikuti pembelajaran tambahan seperti bimbingan di luar sekolah (ya atau tidak)
15	<i>Traveltime</i>	Waktu perjalanan dari rumah ke sekolah (numerik)
16	<i>Famsud</i>	Biaya dukungan dari keluarga (Nominal)
17	<i>Activities</i>	Mengikuti ekstrakurikuler (ya atau tidak)
18	<i>Higher</i>	Apakah ingin melanjutkan pendidikan ke jenjang yang lebih tinggi (ya atau tidak)
19	<i>Internet</i>	Akses internet di rumah (ya atau tidak)
20	<i>Romantic</i>	Memiliki hubungan dengan lawan jenis (ya atau tidak)
21	<i>Goout</i>	Apakah sering bermain dengan teman (ya atau tidak)
22	<i>Sosmed</i>	Waktu bermain sosial media (numerik)
23	<i>Onlinegame</i>	Waktu bermain game online (numerik)
24	<i>Fisikas2</i>	Nilai Fisika semester 2
26	<i>Kimias2</i>	Nilai Kimia semester 2
27	<i>PKNs2</i>	Nilai PKn semester 2
28	<i>Binggriss2</i>	Nilai Bahasa Inggris semester 2
29	<i>Matematikas2</i>	Nilai matematika semester 2
30	<i>PKNs3 (Target data)</i>	Nilai PKn semester 3
31	<i>Binggriss3 (Target data)</i>	Nilai Bahasa Inggris semester 3
32	<i>Matematikas3 (Target data)</i>	Nilai Matematika semester 3

Berdasarkan Tabel 3.1 di atas, digunakan 32 variabel untuk membuat sistem prediksi performa siswa. Tabel 3.1 tersebut juga menjelaskan apa isi dari setiap variabel yang digunakan lengkap dengan rentang data yang ada pada setiap variabel. Variabel yang memiliki penjelasan s.d. menandakan bahwa itu adalah rentang data dari variabel tersebut, yang menyatakan jika tertulis 1 s.d. 5 maka jika nilainya 1 itu menandakan sangat buruk dan jika nilainya adalah 5 maka itu

menandakan sangat baik. Jumlah data yang berhasil dikumpulkan pada penelitian ini ada sebanyak 231 data.

3.6 Predikat Nilai Siswa

Predikat nilai siswa merupakan sebuah nilai kategori yang menyatakan nilai prestasi atau nilai akademis siswa berdasarkan rentang skor tertentu. Penelitian ini menggunakan pembagian predikat siswa seperti pada Tabel 3.2 di bawah ini.

Tabel 3.2 Predikat Nilai Siswa

Predikat Nilai	Keterangan	Rentang Skor
A	Lulus	80 s.d.100
B	Tidak Lulus	0 s.d. 79

Berdasarkan Tabel 3.2 di atas, penelitian ini membuat sistem prediksi performa siswa dengan *machine learning* untuk memprediksi tingkat kelulusan siswa dengan predikat lulus untuk nilai diatas 80, sehingga pada hasil prediksi dengan *machine learning* akan menghasilkan 2 keluaran. Keluaran yang pertama yaitu adalah lulus (diprediksikan mendapat nilai 80 atau lebih), dan yang kedua adalah tidak lulus (diprediksikan mendapat nilai antara 0 sampai 79). Batas nilai pembagian dari nilai 80 dipilih karena nilai 80 merupakan standar yang ingin dicapai oleh guru pada lembaga pendidikan di SMKN 4 Kota Tangerang.

3.7 Training Data dan Testing Data

Tujuan dari penelitian ini adalah untuk mencari model dengan performa prediksi terbaik. Penelitian ini menggunakan satu jenis pembagian data, yaitu sebesar 60% untuk *training* data dan 40% untuk *testing* data. Pembagian data untuk mata pelajaran Matematika dapat dilihat pada Tabel 3.3 di bawah ini.

Tabel 3.3 *Binary Classification Training and Testing Data* (Matematika)

<i>Binary Classification</i> (Matematika)	
Nilai	Total Data
Lulus	73
Tidak Lulus	158
Total Data	231

Berdasarkan Tabel 3.3 di atas, dapat dilihat bahwa total data yang berhasil dikumpulkan ada sebanyak 231 data. Berdasarkan data yang berhasil dikumpulkan tersebut, sebanyak 73 siswa dinyatakan lulus, sedangkan sisanya yang tidak lulus ada sebanyak 158 data. Hasil dari data tersebut kemudian dibagi untuk *training* data dan *testing* data. Hasil pembagian untuk *training* data dengan pembagian sebesar 60% dapat dilihat pada Tabel 3.4 di bawah ini.

Tabel 3.4 Matematika *Binary Classification Training Data* (60%)

<i>Binary Classification Training Data</i> (Matematika)	
Nilai	Total Data
Lulus	41
Tidak Lulus	97
Total Data	138

Berdasarkan Tabel 3.4 di atas, dapat dilihat bahwa pada *training* data dengan pembagian sebesar 60% terdapat 138 data. Data untuk siswa yang lulus ada sebanyak 41 data, sedangkan untuk siswa tidak lulus ada sebanyak 97 data. Hasil pembagian data untuk *testing* data dengan pembagian sebesar 40% dapat dilihat pada Tabel 3.5 di bawah ini.

Tabel 3.5 Matematika *Binary Classification Testing Data* (40%)

<i>Binary Classification Testing Data</i> (Matematika)	
Nilai	Total Data
Lulus	32
Tidak Lulus	61
Total Data	93

Berdasarkan Tabel 3.5 di atas, dapat dilihat bahwa pada *testing data* sebesar 40% terdapat total 93 data. Terdapat juga 32 data untuk siswa yang lulus dan ada 61 data untuk siswa yang tidak lulus. Hasil pembagian pada mata pelajaran Bahasa Inggris dapat dilihat pada Tabel 3.6 di bawah ini.

Tabel 3.6 Bahasa Inggris *Binary Classification Training and Testing Data*.

<i>Binary Classification</i> (Bahasa Inggris)	
Nilai	Total Data
Lulus	48
Tidak Lulus	183
Total Data	231

Berdasarkan Tabel 3.6 di atas, dapat dilihat bahwa total data yang berhasil dikumpulkan untuk mata pelajaran bahasa inggris ada sebanyak 231 data. Dari data yang berhasil dikumpulkan tersebut, sebanyak 48 siswa lulus, sedangkan sisanya adalah siswa yang tidak lulus ada sebanyak 183 data. Hasil dari data tersebut kemudian akan dibagi untuk *training* data dan *testing* data. Hasil pembagian untuk *training* data dengan pembagian 60% ini dapat dilihat pada Tabel 3.7 di bawah ini.

Tabel 3.7 Bahasa Inggris *Binary Classification Training Data* (60%).

<i>Binary Classification Training Data</i> (Bahasa Inggris)	
Nilai	Total Data
Lulus	27
Tidak Lulus	111
Total Data	138

Berdasarkan Tabel 3.7 di atas, dapat dilihat bahwa pada *training* data dengan pembagian sebesar 60% terdapat 138 data. Data untuk siswa yang lulus dengan ada sebanyak 27 data, sedangkan untuk siswa yang tidak lulus ada sebanyak 111 data. Hasil pembagian data untuk *testing* data dengan pembagian sebesar 40% dapat dilihat pada Tabel 3.8 di bawah ini.

Tabel 3.8 Bahasa Inggris *Binary Classification Testing Data* (40%).

<i>Binary Classification Testing Data</i> (Bahasa Inggris)	
Nilai	Total Data
Lulus	21
Tidak Lulus	72
Total Data	93

Berdasarkan Tabel 3.8 di atas, dapat dilihat bahwa pada *testing data* sebesar 40% terdapat total 93 data. Terdapat juga 21 data untuk siswa yang lulus dan ada 72 data untuk siswa yang tidak lulus. Hasil pembagian pada mata pelajaran PKn dapat dilihat pada Tabel 3.9 di bawah ini.

Tabel 3.9 PKn *Binary Classification Training and Testing Data*.

<i>Binary Classification (PKn)</i>	
Nilai	Total Data
Lulus	166
Tidak Lulus	65
Total Data	231

Berdasarkan Tabel 3.9 di atas, dapat dilihat bahwa total data yang berhasil dikumpulkan untuk mata pelajaran PKn ada sebanyak 231 data. Dari data yang berhasil dikumpulkan tersebut, sebanyak 166 siswa lulus, sedangkan sisanya adalah siswa yang tidak lulus ada sebanyak 65 data. Hasil dari data tersebut kemudian dibagi untuk *training* data dan *testing* data. Hasil pembagian untuk *training* data dengan pembagian 60% ini dapat dilihat pada Tabel 3.10 di bawah ini.

Tabel 3.10 PKn *Binary Classification Training Data (60%)*.

<i>Binary Classification Training Data (PKn)</i>	
Nilai	Total Data
Lulus	100
Tidak Lulus	38
Total Data	138

Berdasarkan Tabel 3.10 di atas, dapat dilihat bahwa pada *training* data dengan pembagian sebesar 60% terdapat 138 data. Data untuk siswa yang lulus ada sebanyak 100 data, sedangkan untuk siswa yang tidak lulus ada sebanyak 38 data. Hasil pembagian data untuk *testing* data dengan pembagian sebesar 40% dapat dilihat pada Tabel 3.11 di bawah ini.

Tabel 3.11 PKn *Binary Classification Testing Data (40%)*.

<i>Binary Classification Testing Data (PKn)</i>	
Nilai	Total Data
Lulus	66
Tidak Lulus	27
Total Data	93

Berdasarkan Tabel 3.11 di atas, dapat dilihat bahwa pada *testing* data sebesar 40% terdapat total 93 data. Terdapat juga 66 data untuk siswa yang lulus dan ada 27 data untuk siswa yang tidak lulus.