

BAB 3

METODE PENELITIAN

3.1 Instrumen Penelitian

Bahan dan peralatan yang dibutuhkan dalam penelitian ini antara lain :

3.1.1 Bahan

Bahan yang digunakan dalam penelitian ini yaitu data siswa kelas SMA Negeri 2 Pemalang tahun ajaran 2015/2016.

3.1.2 Peralatan

Peralatan yang digunakan dalam penelitian ini berupa perangkat keras dan perangkat lunak. Berikut adalah perangkat keras dan perangkat lunak yang diperlukan :

1. Kebutuhan hardware :
 - a. Prosesor Intel® Core™ i3-3120M CPU @ 2.50GHz
 - b. Layar monitor 14 inch
 - c. Ram 4,00 GB
 - d. *Harddisk* 500 GB
2. Kebutuhan perangkat lunak :
 - a. Sistem :
 - i) *Windows Edition* : *Microsoft Windows 7 Ultimate*
 - ii) *System Type* : *64-bit Operating System*
 - b. *Data Mining* :
 - i) Matlab (*Matrix Laboratory*) version R2012a (7.14.0.739)

3.2 Metode Pengumpulan Data

Data dan informasi yang sesuai dengan permasalahan yang dihadapi diperlukan dalam penyusunan laporan penelitian. Untuk mendapatkan solusi dari permasalahan yang dihadapi, kemudian data tersebut diolah. Untuk mendapatkan

informasi yang akurat dan berkualitas, maka diperlukan cara dan metode yang sesuai dengan kebutuhan penelitian. Dalam penelitian ini digunakan data primer yaitu data siswa kelas X SMA Negeri 2 Pemalang tahun ajaran 2015/2016 yang diperoleh melalui *softcopy* secara langsung (observasi) dari sumber data yaitu SMA Negeri 2 Pemalang yang beralamat di jalan Jendral Sudirman nomor 14 Pemalang. Data yang diperoleh sebanyak 389 *record* terdiri dari 16 variabel atau atribut.

Berikut ini adalah sampel data siswa kelas X SMA Negeri 2 Pemalang :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	no_da	n	nama	jns_k	tgl_lhr	un_bin	un_bing	un_mtk	un_ipa	rtotal_bhs	rtotal_ips	rtotal_ipa	mnt_bk	mnt_1	alasan	jurusan
2	001	0005576845	AGUNG TRILAKSONO	L	03/04/2000	92	70	87.5	75	81	85.3	86	IPA	IPA	sesuai dengan prestas	IPA
3	005	0006057765	JULIETIKA PUTRI MELINIA	P	01/07/2000	92	82	45	50	83.4	76.5	76.9	IPS	IPA	Karna ilmu IPA lebih d	IPS
4	006	0006532103	Ayu Septiana Amaliyah	P	10/09/2000	86	50	60	55	80.1	80.5	82.3	IPA	IPA	karena peminatan MIF	IPA
5	007	0006532099	syahfara ashari putri	P	16/07/2000	96	76	77.5	82.5	83.8	87.8	87.5	IPA	IPA	karena menurut saya	IPA
6	008	0006357487	Divya Madini	P	04/06/2000	90	86	50	65	86.5	80.2	78.9	BHS	BHS	karena menurut saya,	BHS
7	009	0005594858	Jantra Wisesa Gati	L	16/03/2000	74	80	52.5	57.5	82	79.8	78.1	IPA	IPA	karena saya sangat be	IPA
8	010	0006593806	VIKA OKTAVIA	P	01/10/2000	84	76	42.5	60	85.2	80.4	80.7	IPA	IPA	Saya senang mempelaj	IPA
9	012	9997718227	Rofiatul Adawiyah	P	29/11/2000	82	70	52.5	57.5	81	82.2	81.6	IPS	IPS	mendalami imu alam	IPS
10	013	0005594264	PEPARING GUSTI MASHARATI	L	12/11/2000	80	56	72.5	62.5	81.3	84.9	83.9	IPA	IPA	saya merasa lebih mu	IPA
11	015	0006011368	DANANG WAHYU UTOMO	L	26/10/2000	82	72	52.5	67.5	80.2	84.2	81	IPA	IPA	untuk mempermudah	IPA
12	017	0005576329	SOLEH TEGUH MARGONO	L	31/03/2000	90	88	65	70	81.1	78.2	77.8	IPA	IPA		IPA
13	018	9996796353	TAUFIK HIDAYAT	L	02/09/1999	78	52	65	87.5	77.5	78.2	77.8	IPA	IPA		IPA
14	019	0006411252	Elsa Rahma Prameswari	P	08/08/2000	82	68	57.5	62.5	81.5	79.7	77.2	BHS	IPA	karena saya minat	IPA
15	021	0005594866	ANTY SETIANING TYAS	P	19/04/2000	84	68	72.5	75	82.7	85.8	83.8	IPA	IPA	Lebih suka ke eksak di	IPA
16	022	9997199895	Maya Kartikaningtias	P	10/12/1999	92	70	72.5	55	83.7	81.4	83.4	IPA	IPA	karena saya suka pela	IPA
17	023	0007059052	DIVA RISKY ANANDA	P	01/02/2000	80	76	47.5	60	81.8	81.1	80.3	IPA	IPS	cita cita saya ingin me	IPA
18	025	0006530884	Dias Indah Melisawati	P	04/06/2000	84	56	65	67.5	81.1	86.5	86.4	IPA	IPA	sebenarnya saya meny	IPA
19	026	9996779195	Ayu Lestari	P	15/03/1999	84	64	35	47.5	81.6	79.3	77.6	IPS	IPS		IPS
20	027	0006530867	RULIS IRAWATI	P	16/05/2000	94	78	67.5	70	81.7	83	81.9	IPA	IPA	karena saya insyaAlla	IPA
21	028		PUJI LESTARI	P	21/03/2000	88	72	42.5	65	81	81	80.2	IPA	IPA	ingin lebih mantap dal	IPA
22	029	0005975346	Embun Fajrina Purnamasari	P	16/05/2000	90	80	57.5	77.5	83.7	82.9	83	IPS	IPS	tidak terlalu minat	IPS
23	030	9987798715	SAIUS RIROTUN	P	15/05/1999	84	52	65	57.5	80.2	81.1	80.5	IPA	IPA	karena suka pelajaran	IPA

Gambar 3.1 Sampel Data Siswa kelas X SMA Negeri 2 Pemalang

Dari gambar 3.1 terlihat bahwa data siswa kelas X SMA Negeri 2 Pemalang terdiri dari 16 variabel atau atribut.

Berikut ini adalah penjelasan nama atribut dari data tersebut :

Tabel 3.1 Atribut Data Awal

No.	Nama Atribut	Keterangan
1.	no_daftar	Nomor pendaftaran siswa
2.	Nisn	Nomor induk siswa nasional

No.	Nama Atribut	Keterangan
3.	Nama	Nama Siswa
4.	jns_k	Jenis kelamin siswa
5.	tgl_lhr	Tanggal lahir siswa
6.	un_bin	Nilai Ujian Nasional SMP mata pelajaran Bahasa Indonesia
7.	un_bing	Nilai Ujian Nasional SMP mata pelajaran Bahasa Inggris
8.	un_mtk	Nilai Ujian Nasional SMP mata pelajaran Matematika
9.	un_ipa	Nilai Ujian Nasional SMP mata pelajaran IPA
10.	rtotal_bhs	Nilai rata-rata raport SMP mata pelajaran Bahasa Indonesia dan Bahasa Inggris dari semester 1 sampai dengan semester 5
11.	rtotal_ips	Nilai rata-rata raport SMP mata pelajaran Matematika dan IPS dari semester 1 sampai dengan semester 5
12.	rtotal_ipa	Nilai rata-rata raport SMP mata pelajaran Matematika dan IPA dari semester 1 sampai dengan semester 5
13.	mnt_bk	Jurusan yang direkomendasikan oleh guru BK
14.	mnt_1	Jurusan yang diminati oleh siswa
15.	Alas an	Alasan siswa memilih jurusan yang diminati
16.	Jurusan	Jurusan siswa

3.3 Pengolahan Data Awal

Tidak semua data dan variabel atau atribut yang diperoleh dari data siswa baru kelas X SMA Negeri 2 Pematang tahun ajaran 2015/2016 akan di gunakan atau diolah, karena penelitian yang akan dilakukan kali ini memiliki batasan-batasan data yang digunakan.

Guna mendapatkan data yang berkualitas diperlukan beberapa tahap pengolahan data awal sehingga data siap untuk digunakan.

Berikut ini adalah teknik-teknik yang dilakukan :

1. Pembersihan Data

Data yang tidak berkualitas akan menghasilkan *data mining* yang tidak berkualitas pula. Keputusan yang berkualitas harus didasarkan pada data yang berkualitas (kesalahan atau ketidak benaran dapat terjadi apabila terdapat data ganda atau hilang).

Pekerjaan-pekerjaan pembersihan data:

- a. Melengkapi nilai yang hilang dari suatu atau beberapa data (*missing value*).
- b. Mengidentifikasi atau menghilangkan *outliers* dan memperhalus *data noise*.
- c. Memperbaiki ketidak konsistenan data.
- d. Memecahkan redundansi yang disebabkan oleh integrasi data.

2. *Data integration and transformation*

Penggabungan beberapa sumber data dan mentransformasikan atau mengubah data kedalam format yang sesuai untuk diproses penggalian data. Tugas yang dilakukan pada tahap ini yaitu :

- a. Mengintegrasikan berbagai database atau file-file
- b. Transformasi data (*Data transformation*)
- c. Normalisasi dan *aggregation*

3. *Data reduction*

Mereduksi *dataset* dengan mengurangi jumlah atribut dan *record* agar lebih sedikit namun bersifat informatif.

- a. Mendapatkan representasi dalam volume data yang sudah berkurang tetapi menghasilkan hasil analitis yang sama atau serupa.
- b. Diskritisasi data, yaitu bagian dari reduksi data untuk data numerik.

Beberapa variabel akan dihapus karena tidak mempengaruhi hasil prediksi penjurusan siswa. Atribut yang digunakan sebanyak 8 atribut, 1 atribut sebagai ID, dan 1 atribut sebagai label atau target.

Tabel 3.2 Atribut yang Digunakan

No.	Nama Atribut	Keterangan
1.	Nama	Nama Siswa
2.	un_mtk	Nilai Ujian Nasional SMP mata pelajaran Matematika
3.	un_ipa	Nilai Ujian Nasional SMP mata pelajaran IPA
4.	rtotal_bhs	Nilai rata-rata raport SMP mata pelajaran Bahasa Indonesia dan Bahasa Inggris dari semester 1 sampai dengan semester 5
5.	rtotal_ips	Nilai rata-rata raport SMP mata pelajaran Matematika dan IPS dari semester 1 sampai dengan semester 5
6.	rtotal_ipa	Nilai rata-rata raport SMP mata pelajaran Matematika dan IPA dari semester 1 sampai dengan semester 5
7.	mnt_1	Jurusan yang diminati oleh siswa
8.	Jurusan	Jurusan siswa

3.4 Metode Penelitian

Dalam penelitian ini, metode *data mining* yang diusulkan berupa model standarisasi *data mining* CRISP-DM yang terdiri dari tahap-tahap [12][15]:

3.4.1 Pemahaman Bisnis (*Bussiness Understanding*)

Penentuan tujuan proyek dan kebutuhan secara detil dalam lingkup bisnis atau unit penelitian secara keseluruhan, dan juga menerjemahkan tujuan dan batasan dari data yang diambil dari Sekolah Menengah Atas (SMA) kelas X menjadi formula dari permasalahan *data mining* mulai dari menyiapkan strategi awal hingga metode yang dibutuhkan untuk mencapai tujuan.

3.4.2 Pemahaman Data (*Data Understanding*)

Dalam penelitian ini digunakan data primer yaitu data siswa kelas X SMA Negeri 2 Pemalang tahun ajaran 2015/2016 yang diperoleh melalui *softcopy* secara langsung (observasi). Data yang digunakan adalah data siswa baru atau siswa kelas X tahun ajaran 2015/2016 SMA Negeri 2 Pemalang.

3.4.3 Pengolahan Data (*Data Preparation*)

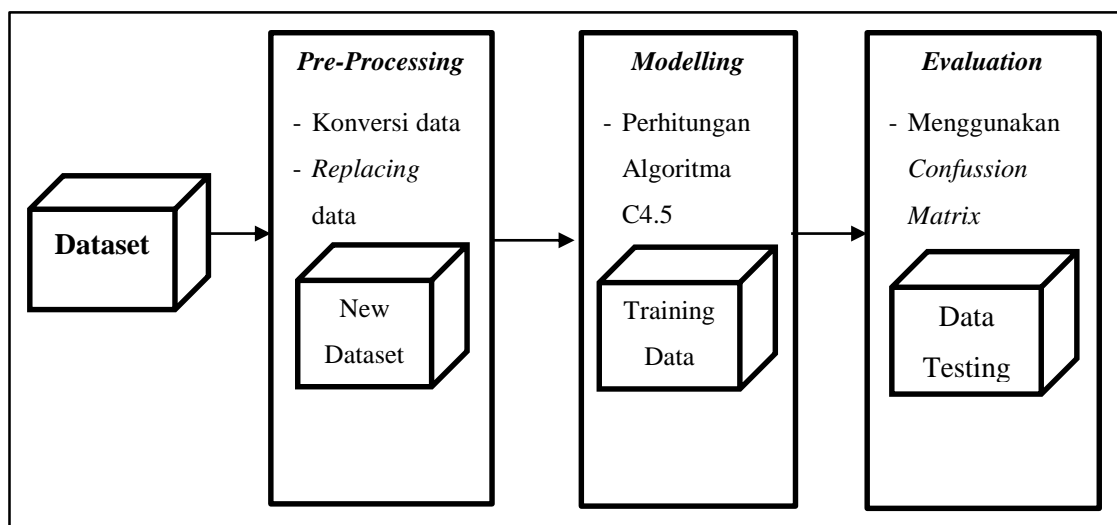
Berikut adalah tahapan pengolahan data :

1. Proses pertama yang dilakukan adalah menentukan data yang akan diolah. Dari data yang telah diperoleh, tidak semua data akan diolah karena tidak semua masuk dalam kriteria penelitian. Penelitian yang akan dilakukan memiliki batasan-batasan data yang akan digunakan. Data awal yang diperoleh terdiri dari 389 *record* data.
2. Proses kedua yang dilakukan adalah menentukan *variable* atau atribut yang akan digunakan dari proses pertama. Terdapat 16 atribut/*variabel* pada data awal, atribut/*variabel* yang akan digunakan sebanyak 8 atribut/*variabel* adalah nama, *n_mtk*, *un_ipa*, *rtotal_bhs*, *rtotal_ips*, *rtotal_ipa*, *mnt_1*, jurusan.
3. Proses ketiga yang dilakukan adalah penanganan data *missing value*. *Missing value* adalah data yang tidak lengkap dikarenakan atribut tidak tercatat maupun atribut memang tidak dimiliki dsb. Penanganan *missing value* dilakukan dengan penghapusan *record* yang kosong. Jumlah data awal adalah 389 data, terdapat data yang *missing value* sehingga menjadi 382 data *record* yang dapat digunakan.
4. Proses keempat yang dilakukan adalah melakukan konversi data. Data yang telah dipilih kemudian dikonversi guna mempermudah proses penambahan data pada sebagian atribut. Data akan diproses menggunakan alat bantu *data mining*. Konversi dilakukan pada atribut *un_mtk*, *un_ipa*, *rtotal_bhs*, *rtotal_ips*, *rtotal_ipa*, dan *mnt_1*.

3.4.4 Pemodelan (*Modelling*)

Decision tree C4.5 dalam penelitian ini merupakan metode klasifikasi yang digunakan, sedangkan untuk evaluasi serta pengukuran tingkat akurasi menggunakan kerangka kerja *confussion matrix*.

Berikut adalah gambar pemodelan *data mining* :



Gambar 3.2 Model Penelitian yang Diusulkan

3.4.5 Evaluasi (*Evaluation*)

Tahap dimana dilakukan validasi dan pengukuran tingkat akurasi hasil yang dicapai oleh model yang telah ditetapkan. Tahap ini dilakukan menggunakan *framework Confussion Matrix* dengan *tools* Matlab.

3.4.6 Penyebaran (*Deployment*)

Output dari penelitian ini akan menunjukkan hasil klasifikasi penjurusan siswa SMA Negeri 2 Pematang yang sudah diproses menggunakan *software* matlab. Hasil tersebut akan disajikan dalam bentuk sebuah sistem pendukung keputusan yang dapat digunakan oleh pihak sekolah sebagai dasar pertimbangan untuk melakukan penjurusan siswa dengan variabel atau atribut yang telah ditentukan.