

BAB III

METODE PENELITIAN

3.1 Instrumen Penelitian

Pada penelitian ini bahan dan peralatan yang diperlukan sebagai berikut:

3.1.1 Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah data siswa kelas X SMA 1 Kajen tahun ajaran 2015/2016, yang beralamat di Jalan Mandurorejo Kajen, Kabupaten Pekalongan.

3.1.2 Peralatan

Peralatan yang digunakan dalam melakukan penelitian berupa kebutuhan perangkat keras dan perangkat lunak. Berikut adalah kebutuhan sistem :

Kebutuhan perangkat keras :

- Prosesor AMD A6-4455 APU with Radeon(tm) HD Graphics 2.10 GHz
- Layar monitor 14 inch
- Ram 4,00 GB
- Harddisk 500 GB

Kebutuhan perangkat lunak :

1. Sistem :
 - a. Windows Edition : Microsoft Windows 7 Ultimate
 - b. System Type : 64-bit Operating System
2. Data Mining
 - a. Matlab version R2012a (7.14.0.739)

Tools yang digunakan untuk mengolah data dalam klasifikasi penjurusan siswa SMA 1 Kajen dengan menggunakan metode data mining.

b. RapidMiner 5.3.013

Tools yang digunakan untuk melihat tingkat keakurasian data yang diteliti dengan menggunakan algoritma *Naive Bayes*.

3.2 Prosedur Pengambilan atau Pengumpulan Data

Dalam penyusunan suatu laporan penelitian diperlukan data dan informasi yang sesuai dengan permasalahan yang dihadapi. Kemudian data tersebut diolah untuk mendapatkan solusi permasalahan yang ada. Untuk mendapatkan data dan informasi yang akurat dan berkualitas maka diperlukan metode dan cara sesuai dengan kebutuhan penelitian. Penelitian ini menggunakan data primer yaitu data siswa kelas X tahun ajaran 2015/2016 yang diperoleh melalui *softcopy* secara langsung dari sumber data yaitu SMA 1 Kajen, data diperoleh sebanyak 332 record dan terdiri dari 15 variabel atau atribut. Berikut contoh data siswa SMA 1 Kajen.

No	D	E	F	G	O	W	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	BE	BR
pendaftaran	NAMA	Jns kelamin	Asal SMP	RAPORT IPA	RAPORT MTK	RAPORT IPS	RT KKM				NILAI UN		RT un	Min at	dj uru	RT2 RAPORT	nilai	REKOMENDASI		
5				rt	rt	rt	IPA	MAT	rt	IPS	IPA	MAT	Mat/ipa	Mat/ipa	skan	Mat/ipa	itas	IQ	Rekom 1	
6	C.3043	ARYA ARINAL HAQ	L	SMPN 1 KAJEN	82,00	81,17	80,83	80	80	80	80	77,5	72,5	75,0	A	A	81,6	C	105	SMA (IPA)
7	A.027	Aulia Putri Faradisa	P	SMP 1 KAJEN	87,50	88,50	87,17	80	80	80	80	87,5	87,5	87,5	A	A	88,0	B	108	SMA (IPA)
8	B.022	AYU IDHA MUFIDHA	P	SMP 1 KAJEN	90,17	89,83	88,83	80	80	80	80	95,0	97,5	96,3	A	A	90,0	B	105	SMA (IPA)
9	A.078	Cindemas Pujanengrat	P	SMP 1 KAJEN	88,17	90,50	81,83	80	80	80	80	85,0	97,5	91,3	A	A	89,3	B	106	SMA (IPA)
10	B.2074	DENY GEOVANY	L	SMP 1 WONOPRING	81,50	80,17	79,17	74	75	74,5	75	65,0	75,0	70,0	A	A	80,8	C	104	SMA (IPA)
11	A.074	Dewi Faradiba	P	SMP 1 KAJEN	88,17	91,83	84,17	80	80	80	80	82,5	97,5	90,0	A	A	90,0	B	106	SMA (IPA)
12	A.108	DHANY AJIPERWATA	L	SMP 1 KARANGANYU	89,67	94,67	88,17	72	71	71,5	73	80,0	90,0	85,0	A	A	92,2	B	105	SMA (IPA)
13	A.079	Diah Fara Dipa	P	SMP 1 KAJEN	87,33	90,67	87,50	80	80	80	80	80,0	92,5	86,3	A	A	89,0	B	106	SMA (IPA)
14	C.3020	DIAN TRI IMAWAN	L	SMP PONDOK SELAJI	89,83	88,67	0,00	76	75	75,5	0	80,0	85,0	82,5	A	A	89,3	B	106	SMA (IPA)
15	A.1027	EVA NOVITASARI	P	SMP 1 BOJONG	81,83	79,67	83,17	72	73	72,5	74	85,0	75,0	80,0	A	A	80,8	C	105	SMA (IPA)
16	B.2079	FABIOLA CAROLINE N.H	P	SMP KANISIUS TIRTO	92,67	94,67	93,33	62	62	62	69	77,5	87,5	82,5	A	A	93,7	A	*	.
17	B.135	FAUZYYAH DESY ANANDA	P	SMP 1 BOJONG	90,50	86,33	90,00	72	73	72,5	74	92,5	92,5	92,5	A	A	88,4	B	105	SMA (IPA)
18	C.3027	GHOZI KORI AQILLA	L	SMPN 1 KAJEN	83,33	88,50	83,83	80	80	80	80	87,5	77,5	82,5	A	A	85,9	C	107	SMA (IPA)
19	A.191	HAPPY RINDU ROMANTIKA	P	SMP 1 SRAGI	89,83	88,00	93,67	75	75	75	75	82,5	85,0	83,8	A	A	88,9	B	102	SMA (IPA)
20	A.1099	HASNA KHOLIVIYA	P	MTS AL FATAH TALANG	88,17	88,17	94,83	75	74	74,5	75	75,0	70,0	72,5	A	A	88,2	B	102	SMA (IPA)
21	B.2008	HERSI MITA RIZKI	P	SMP 1 KAJEN	83,33	86,83	83,00	80	80	80	80	95,0	70,0	82,5	A	A	85,1	C	111	SMA (IPA)
22	B.2060	HILDA RESTU WIYANSA	P	MTS SALAFIYAH PAJANG	86,67	87,00	91,17	70	70	70	70	80,0	77,5	78,8	A	A	86,8	B	104	SMA (IPA)
23	A.1059	ILHAM AMIN NURDIN	L	SMP NU KAJEN	81,17	81,17	83,17	70	70	70	70	75,0	80,0	77,5	A	A	81,2	B	104	SMA (IPA)
24	A.1094	IMAM KAMALUDDIN	L	SMP 1 KAJEN	83,00	85,50	82,00	80	80	80	80	75,0	70,0	72,5	A	A	84,3	C	108	SMA (IPA)
25	B.043	INTAN YOLANDA SARI	P	SMP 1 KAJEN	91,83	89,50	89,83	80	80	80	80	82,5	90,0	86,3	A	A	90,7	B	102	SMA (IPA)
26	B.044	ITA VERA	P	SMP 1 KAJEN	92,17	95,00	90,00	80	80	80	80	82,5	97,5	90,0	A	A	93,6	B	107	SMA (IPA)
27	C.3089	KALAM HANAN	L	SMPN 1 KAJEN	83,17	85,33	81,00	80	80	80	80	65,0	82,5	73,8	A	A	84,3	C	101	SMA (IPS)

Gambar 3.1 : Contoh Data Siswa

Dari gambar 3.1 terlihat data siswa SMA 1 kajen terdiri dari 15 variabel. Berikut adalah keterangan nama atribut dan penjelasan data siswa di atas.

Tabel 3.1 : Atribut data awal

No	Nama atribut	Penjelasan
1	No pendaftaran	No pendaftaran siswa
2	Nama	Nama siswa
3	Jenis Kelamin	Jenis Kelamin Siswa
4	Nama SMP	Nama SMP asal siswa
5	Rata-rata Raport IPA	Nilai rata-rata raport SMP mata pelajaran IPA selama 5 semester
6	Rata-rata Raport Matematika	Nilai rata-rata raport SMP mata pelajaran Matematika selama 5 semester
7	Rata-rata Raport IPS	Nilai rata-rata raport SMP mata pelajaran IPS selama 5 semester
8	Rata-rata Raport KKM	Rata-rata Raport KKM
9	Nilai UN IPA	Nilai Ujian Nasional SMP mata pelajaran IPA
10	Nilai UN Matematika	Nilai Ujian Nasional SMP mata pelajaran Matematika
11	Minat	Jurusan yang diminati siswa
12	Jurusan	Jurusan siswa
13	Nilai Kualitas	Kualitas nilai berdasarkan KKM SMP asal
14	IQ	Nilai IQ siswa
15	Rekomendasi Jurusan	Jurusan rekomendasi guru BK

3.3 Pengolahan Data Awal

Dari data siswa baru kelas X SMA 1 Kajen tahun ajaran 2015/2016 yang diperoleh, tidak semua data dan atribut/variabel akan digunakan atau diolah karena penelitian yang akan dilakukan kali ini, memiliki batasan-batasan data yang digunakan.

Untuk mendapatkan data yang berkualitas diperlukan beberapa tahap pengolahan data awal sehingga data siap untuk digunakan. Berikut adalah teknik yang dilakukan untuk pengolahan data awal :

1. *Data Cleaning*

Data yang tidak berkualitas akan menghasilkan data mining yang tidak berkualitas. Keputusan yang berkualitas harus didasarkan pada data yang berkualitas. Data yang hilang akan menyebabkan ketidakbenaran atau menyesatkan. Ekstraksi data, pembersihan, dan transformasi data merupakan tugas utama dalam *data warehouse*.

Tugas-tugas data cleaning:

- a) Mengisi nilai data yang hilang dan tidak lengkap (*missing value*).
- b) Mengidentifikasi atau menghilangkan *outliers* dan memperhalus *data noise*.
- c) Memperbaiki ketidak konsistenan data.
- d) Memecahkan redudansi yang disebabkan oleh integrasi data.

2. *Data Integration and Transformation*

Penggabungan beberapa sumber data dan mentransformasikan data kedalam format yang sesuai untuk diproses pengalihan data.

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

3. *Data Reduction*

Mengurangi jumlah atribut dan record data agar lebih sedikit sehingga bersifat informatif.

- a) Mendapatkan representasi dalam *volume* data yang sudah berkurang namun tetap menghasilkan analisis yang sama.
- b) Diskritisasi data merupakan bagian dari reduksi data dan bagian penting untuk data numerik.

Beberapa atribut akan dihilangkan karena tidak mempengaruhi hasil klasifikasi penjurusan siswa. Atribut yang lain akan tetap diikuti, ada sebanyak 9 atribut, 1 atribut sebagai ID dan 1 atribut sebagai label.

Tabel 3.2 : Atribut yang digunakan

No	Nama atribut	Penjelasan
1	Nama	Nama siswa
2	Rata-rata Raport IPA	Nilai rata-rata raport SMP mata pelajaran IPA selama 5 semester
3	Rata-rata Raport Matematika	Nilai rata-rata raport SMP mata pelajaran Matematika selama 5 semester
4	Nilai UN IPA	Nilai Ujian Nasional SMP mata pelajaran IPA
5	Nilai UN Matematika	Nilai Ujian Nasional SMP mata pelajaran Matematika
6	Minat	Jurusan yang diminati siswa
7	Jurusan	Jurusan siswa
8	Nilai Kualitas	Kualitas nilai berdasarkan KKM SMP asal
9	IQ	Nilai IQ siswa

3.4 Metode Penelitian

Metode yang digunakan pada penelitian ini adalah model standarisasi data mining, yaitu CRISP-DM (*Cross Industry Standart Process for Data Mining*). Berikut adalah langkah-langkah metode CRISP-DM :

3.4.1 Pemahaman Bisnis (*Bussiness Understanding*)

Penelitian ini mengutamakan penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan, serta menerjemahkan tujuan dan batasan dari data yang diperoleh dari SMA 1 Kajen sehingga menjadi formula dari permasalahan data mining mulai dari menyiapkan strategi awal hingga metode yang dibutuhkan untuk menapai tujuan.

3.4.2 Pemahaman Data (*Data Understanding*)

Pada penelitian ini data yang digunakan adalah data primer. Data diperoleh melalui *softcopy* secara langsung dari SMA 1 Kajen. Data yang didapatkan adalah data siswa kelas X SMA 1 Kajen tahun ajaran 2015/2016.

3.4.3 Pengolahan Data (*Data Preparation*)

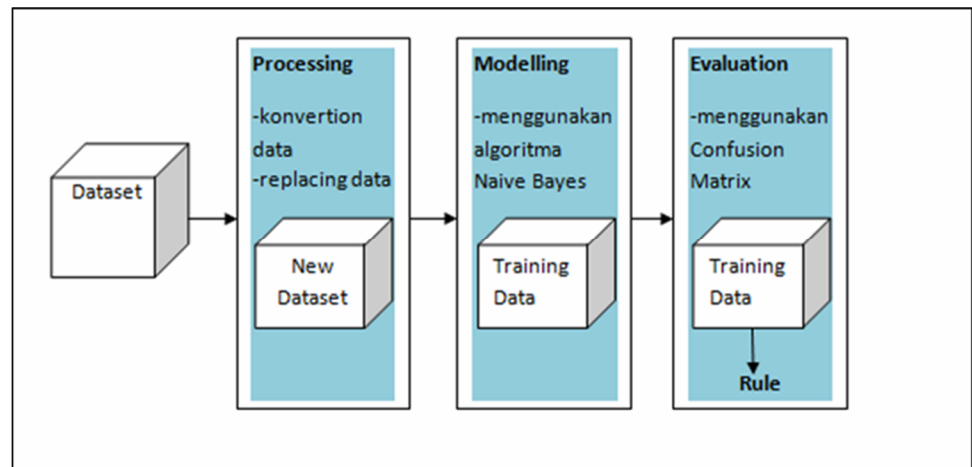
- a. Tahap Pertama yang dilakukan adalah penentuan data yang akan diolah. Dari data yang telah diperoleh, tidak semua data akan diolah karena tidak semua masuk dalam kriteria penelitian, penelitian yang akan dilakukan memiliki batasan-batasan data yang akan digunakan. Data awal terdiri dari 332 *record* data.
- b. Tahap Kedua yang dilakukan adalah menentukan atribut yang akan digunakan dari tahap pertama. Terdapat 15 variabel pada data awal, variabel yang akan digunakan sebanyak 9 variabel

adalah nama siswa, rata-rata nilai raport ipa, rata-rata nilai raport matematika, nilai ujian nasional ipa, nilai ujian nasional matematika, nilai kualitas, tes IQ, minat jurusan dan jurusan.

- c. Tahap Ketiga yang dilakukan adalah penanganan data *missing value*. *Missing value* adalah data yang tidak lengkap dikarenakan atribut tidak tercatat maupun atribut memang tidak dimiliki. Penanganan *missing value* dilakukan dengan penghapusan *record* yang kosong. Jumlah data awal adalah 332 data, terdapat data yang *missing value* sehingga menjadi 304 data *record* yang dapat digunakan.
- d. Tahap Keempat atau yang terakhir adalah melakukan konversi data. Guna memudahkan dalam analisa proses data mining maka data dengan atribut yang telah ditentukan atau dipilih akan dilakukan konversi. Konversi dilakukan pada atribut rata-rata nilai raport ipa, rata-rata nilai raport matematika, nilai ujian nasional ipa, nilai ujian nasional matematika, nilai kualitas, nilai IQ dan minat jurusan.

3.4.4 Pemodelan (*Modeling*)

Dalam penelitian ini metode yang digunakan adalah algoritma *Naive Bayes Classification*. Untuk pengukuran tingkat akurasi klasifikasi penjurusan siswa ini menggunakan *Tools* RapidMiner dan Matlab. Berikut adalah gambaran pemodelan penelitian :



Gambar 3.2 : Model Penelitian

3.4.5 Validasi dan Evaluasi (*Evaluation*)

Pada tahap ini dilakukan validasi dan pengukuran tingkat keakuratan dari hasil yang dicapai oleh model dengan menggunakan *tools* Matlab yaitu *Confusion Matrix* untuk menguji tingkat keakurasian model yang digunakan.

3.4.6 Penyebaran (*Deployment*)

Hasil dari penelitian ini berupa analisis yang mengarah ke *Decision Support System* (DSS) yang diharapkan dapat digunakan oleh pihak sekolah SMA 1 Kajen khususnya guru BK sebagai bahan pertimbangan untuk proses penjurusan siswa serta menentukan langkah penjurusan siswa dengan atribut yang telah dijabarkan.