

BAB III

METODE PENELITIAN

3.1 Desain penelitian

Model yang digunakan dalam penelitian ini adalah model CRISP-DM, dengan langkah-langkah sebagai berikut:

3.1.1 Pemahaman bisnis (*Business understanding*)

Pendidikan sebagai modal utama bagi setiap orang dinilai sangat penting keberadaannya. Dengan berpendidikan, setiap individu akan mempunyai nilai lebih di mata orang yang memandangnya. Dunia universitas sudah bukan lagi barang mewah dan sudah seharusnya setiap anak dapat mengenyam pendidikan hingga level sarjana. Untuk dapat masuk ke perguruan tinggi, biasanya sekolah menengah atas melakukan usaha-usaha, seperti mendaftarkan calon lulusannya ikut seleksi secara bersama, mengikutsertakan tes yang di selenggarakan oleh perguruan tinggi agar lulusannya dapat lolos.

Kualitas sebuah sekolah menengah atas dapat juga dilihat dari rasio kelulusan siswanya lolos seleksi perguruan tinggi. Jika setiap tahunnya siswa yang lolos seleksi perguruan tinggi meningkat, maka akan naik pula kualitas sekolah tersebut yang akan disertai tingkat kepercayaan masyarakat terhadap sekolah tersebut juga akan naik. Namun jika terjadi penurunan siswa yang lolos seleksi dan terus terjadi, maka akan menjadi masalah tersendiri. Untuk itu diperlukan pemantauan dan evaluasi secara berkala terhadap kecenderungan tingkat kelulusan siswa yang lolos seleksi perguruan tinggi untuk dapat memberikan pertimbangan kebijakan yang akan diambil oleh sekolah tersebut.

3.1.2 Pemahaman data (*Data understanding*)

Data yang digunakan merupakan data siswadari tahun 2015-2017 dan data nilai dari semester 1 – 4. Data siswa tersebut di dapat dari pangkalan data siswa EMIS Madrasah yang dikelola oleh operator madrasah, sedangkan data nilai didapat dari LEGGER madrasah yang dimiliki bagian kurikulum madrasah. Selain itu juga data yang digunakan adalah data siswa yang lolos seleksi SBMPTN.

Gambar 4.1 Contoh tabel data siswa (*sebagian data terpotong*)

Nama Siswa	Tempat Lahir	Tanggal Lahir (dd/mm/yyyy)	Jenis Kelamin	Tingkat / Kelas	Jumlah Saudara	Jenis Sekolah	Status Sekolah	Kabupaten/Kota Lokasi Sekolah	Alamat
Achmad Haikal Vaicena	Terbanggi Besar	22/08/2000	L	10	2	1	2	LAMPUNG TENGAH	jl.lintas timur km 79,Perum Kopkar Dwi Karya
Afif Efendi	Pujokerto	14/07/1999	L	10	2	2	2	metro utara	pujokerto
Amelya Nur Afifah	Siraman	07/12/1999	P	10	1	1	1	KOTA METRO	DUSUN II RT/RW 005/003
Anisa Oktaviana	Sidodadi	12/10/1999	P	10	2	1	1	lampung timur	sidodadi
Arif Ahadi Fajar	Metro	01/01/1998	L	10	2	1	1	KOTA METRO	Iring Mulyo metro timur
Cindy Puspita Sari	Sri Pendowo	28/07/2000	P	10	3	1	1	LAMPUNG TIMUR	SRIPENDOW
Danang Prayogo	Srimenanti	15/01/2000	L	10	1	1	1	LAMPUNG TIMUR	SRIMENANTI
Dara Mutiara Mylan Kidr	Sekampung	12/04/2000	P	10	1	1	1	Lampung Timur	BATANGHARI
Darul Efendi	Cimarias	20/09/1999	L	10	2	2	1	LAMPUNG TIMUR	cimarias, bangun rejo, lampung tengah
Della Susiani	way kanan	20/05/2000	P	10	2	2	1	WAY KANAN	way kanan
Dian Fitriani	Tulang Bawang	26/02/2000	P	10	3	1	1	tulang bawang barat	indraloka jaya,way kenanga,tulang bawang barat
Dinda Dwi Meliani	Tanjung Intan	08/03/2000	P	10	2	1	1	lampumh timur	tegal yoso
Ega Sita Larasaty	Liwa	05/07/2000	P	10	4	2	2	KOTA METRO	38 polos, banjarejo lamtim
Elvis Armando Bastian P	Kibang	14/02/2000	L	10	2	1	1	Lampung Timur	Kibang. kecamatan Metro Kibang, kabupaten Lampu
Fadhil Nur Hidayat	Lampung Tengah	10/10/2000	L	10	3	2	1	LAMPUNG TIMUR	desa jatidatar lampung tengah
Gusty Welliansyah	Yosodadi	20/08/2000	L	10	3	1	1	METRO TIMUR	banjarejo 38 polos
Halik Alfizanur Rohma	Metro	18/10/2000	P	10	2	1	1	LAMPUNG TIMUR	srisawahan
Intan Mustika Ayu	Sukadana Timur	20/06/2001	P	10	2	1	2	LAMPUNG TIMUR/SUKADA	sukadana timur
Luci Novita Sari	Banding	07/11/2000	P	10	3	1	2	LAMPUNG TIMUR/sukaDA	SUKADANA TENGAH,SUKADANA,LAMPUNG TIMUR
Masdar Khilmy Alamri	Purwosari	06/03/2000	L	10	3	1	2	metro barat	28 purwosari metro utara kota metro
Mery Qoirunisa	Bumi Nabung Baru	06/01/2000	P	10	2	1	1	tulang bawang	bratasena, tulang bawang
Muhammad Nuri Yasin	Catur Tunggal	08/11/1999	L	10	2	2	1	LAMPUNG TIMUR	38h banjarejo lampung timur

Gambar 4.2 Contoh tabel daftar nilai siswa (*sebagian data terpotong*)

NAMA SISWA	JURUSAN	SEMESTER 1										SEMESTER 2									
		MTK	IND	ING	FIS	KIM	BIO	SOS	EKO	GEO	AVG	MTK	IND	ING	FIS	KIM	BIO	SOS	EKO	GEO	AVG
Annisa Dwi Apriliana	IPA	86	85	81	86	86	86				85	81	83	83	82	83	83				83
Ayu Rantika	IPA	86	85	81	82	86	82				84	88	85	84	85	85	85				85
Dike Herawati	IPA	85	85	87	84	84	84				85	82	83	76	82	85	85				82
Eliyana Wulan Sari	IPA	83	81	82	83	83	83				83	80	84	82	82	83	83				82
Febri Widya Astuti	IPA	81	81	87	81	85	80				83	85	83	85	85	86	85				85
Firda Arifah	IPA	85	79	81	79	86	80				82	86	85	87	87	88	88				87
Iqbal Maulana Assidiq	IPA	81	85	86	82	76	82				82	82	84	84	83	85	85				84
Khamdani Yahya	IPA	80	81	85	88	85	79				83	82	85	83	83	84	84				84
M. Wahyu Hadi Santoso	IPA	79	79	81	83	81	83				81	82	83	83	83	84	84				83
Nabiila Tuzzakyyah	IPA	84	84	81	83	82	80				82	81	83	83	83	84	84				83
Nabila Citra Pertiwi	IPA	86	85	81	84	83	84				84	84	83	85	85	86	85				85
Nurul Amanah Agustina	IPA	82	81	80	83	81	84				82	83	84	84	84	86	85				84
Olvi Lorensky Yeztiani	IPA	85	80	82	83	83	84				83	80	82	82	82	84	84				82
Putri Maharani	IPA	80	81	79	81	83	80				81	83	84	87	86	89	89				86
Riko Ahmad Gusfafa	IPA	84	85	89	84	83	84				85	85	84	84	83	85	85				84
Riky Al Rauf Gusfara	IPA	80	80	83	81	81	83				81	84	84	85	85	85	85				85
Riska Aprilia	IPA	83	80	81	85	83	82				82	86	85	85	85	86	86				86
Sari Ayu Wulandari	IPA	87	85	82	84	82	84				84	86	85	86	85	86	86				86
Sri Wahyu Ningsing	IPA	83	80	79	83	87	80				82	85	85	85	85	86	85				85
Stefanny Noviantari	IPA	80	85	81	83	86	76				82	87	83	87	85	87	86				86
Aisyah Azzahra	IPS	79	80	85				83	81	83	82	84	83	86				86	86	86	85
Annisa Dini Sholeha	IPS	83	84	80				83	79	83	82	83	83	87				87	88	87	86
Asri Anggi Anda Yani	IPS	80	78	85				84	85	83	83	86	84	86				85	86	86	86
Febri Adi Laksono	IPS	85	79	81				84	80	83	82	84	82	85				85	85	85	84
Melda Noer Fadhillah	IPS	83	80	86				83	79	83	82	87	83	86				85	86	86	86
Mella Adistria	IPS	81	80	78				83	79	83	81	83	84	85				85	86	86	85

3.1.3 Persiapan data (*Data preparation*)

Dalam tahap ini, data siswa dan data nilai digabung menjadi satu untuk dijadikan *dataset*. Dataset didapat dari pembersihan data dari *missing value* atau data yang tidak lengkap serta menghapus atribut yang dinilai tidak perlu untuk diikuti sertakan . Data siswa yang diambil hanya data siswa yang lolos seleksi serta sebagian siswa yang tidak lolos seleksi.

Gambar 4.3 Contoh dataset setelah dilakukan *cleaning* dan penggabungan
(*sebagian data terpotong*)

JK	TEMPAT LAHIR	ALAMAT	PENDIDIKAN ORTU	PEKERJAAN ORTU	PENGHASILAN ORTU	Jumlah	Jenis SMP	Status SMP	MTK 4	IND 0 4	ENG 4	URSHM MTK	URSHM INDO	URSHM ENG	URSHM IPA	Label
P	Metro	Metro	SMP	Petani	1000000-2000000	4	MTs	Swasta	81	80	82	52	58	62	51	Lulus
L	Metro	Tulang Bawang Ba	SMA	Petani	2000000-3000000	0	SMP	Negeri	80	80	81	51	58	61	61	Lulus
P	Lampung Tengah	Lampung Tengah	SMP	Petani	1000000-2000000	2	SMP	Negeri	78	80	83	52	57	51	52	Lulus
P	Lampung Timur	Lampung Timur	Sarjana	Guru	2000000-3000000	1	SMP	Negeri	85	88	82	53	63	53	51	Lulus
P	Tulang Bawang	Tulang Bawang	Sarjana	Pegawai Swa	2000000-3000000	2	SMP	Negeri	82	85	84	51	61	55	54	Lulus
P	OKU Timur	OKU Timur	SMP	Wiraswasta	2000000-3000000	3	MTs	Swasta	81	82	85	51	60	52	62	Lulus
P	Lampung Timur	Lampung Timur	SMP	Wiraswasta	2000000-3000000	1	SMP	Negeri	80	82	82	52	61	51	61	Lulus
P	Lampung Tengah	Lampung Tengah	SMP	Petani	1000000-2000000	3	SMP	Negeri	80	83	83	56	62	50	60	Lulus
P	Lampung Tengah	Lampung Tengah	SMP	Wiraswasta	2000000-3000000	2	SMP	Negeri	80	80	82	54	59	53	53	Lulus
L	Lampung Tengah	Metro	SMP	Wiraswasta	2000000-3000000	2	MTs	Swasta	89	87	90	53	61	53	53	Tidak Lulus
P	Jakarta	Metro	SMP	Lainnya	1000000-2000000	2	MTs	Negeri	90	87	91	52	65	53	54	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMP	Petani	2000000-3000000	2	MTs	Negeri	89	89	90	55	64	63	56	Tidak Lulus
P	Metro	Metro	SMP	Pedagang	2000000-3000000	2	SMP	Negeri	87	82	87	61	58	61	53	Tidak Lulus
L	Metro	metro	SMA	PNS	2000000-3000000	3	SMP	Negeri	87	82	87	64	65	53	50	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMA	Pegawai Swa	2000000-3000000	3	SMP	Negeri	88	83	88	53	61	57	51	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMA	Wiraswasta	1000000-2000000	0	SMP	Negeri	88	83	88	57	60	56	54	Tidak Lulus
P	Lampung Tengah	Lampung Timur	SMA	Petani	3000000-4000000	2	SMP	Negeri	82	77	82	51	57	62	53	Tidak Lulus
P	Metro	metro	SMA	Petani	2000000-3000000	2	MTs	Swasta	89	88	90	50	61	64	60	Tidak Lulus
P	Lampung Tengah	Lampung Tengah	SMA	PNS	2000000-3000000	5	SMP	Negeri	90	89	91	60	64	56	61	Tidak Lulus
P	Lampung Timur	Lampung Timur	Sarjana	guru	3000000-4000000	1	SMP	Negeri	89	87	90	61	60	53	54	Tidak Lulus
P	Lampung Utara	Lampung Utara	SMP	Petani	1000000-2000000	4	MTs	Swasta	87	88	88	54	57	56	50	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMP	Petani	1000000-2000000	2	SMP	Negeri	90	89	90	52	55	62	51	Tidak Lulus
L	Lampung Timur	Lampung Timur	SMA	Petani	1000000-2000000	3	MTs	Swasta	86	88	86	50	56	61	60	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMP	Petani	1000000-2000000	3	SMP	Negeri	90	89	91	66	61	54	54	Tidak Lulus
L	Lampung Tengah	Lampung Tengah	SMA	Petani	2000000-3000000	1	SMP	Negeri	87	88	88	53	60	53	56	Tidak Lulus
P	Lampung Timur	Lampung Timur	SMP	Petani	1000000-2000000	2	SMP	Negeri	90	81	90	50	58	57	50	Tidak Lulus
P	Lampung Tengah	Lampung Tengah	Sarjana	Guru	3000000-4000000	3	SMP	Negeri	88	87	89	58	57	63	53	Tidak Lulus
L	Lampung Tengah	metro	SMP	Petani	1000000-2000000	4	SMP	Negeri	89	84	89	52	61	56	60	Tidak Lulus
L	Palembang	Palembang	SMP	Wiraswasta	2000000-3000000	3	MTs	Swasta	89	87	90	53	64	56	51	Tidak Lulus

Adapun atribut yang akan digunakan, yakni

1. Jenis kelamin

Atribut jenis kelamin dipilih karena atribut ini akan memberikan informasi tentang perbedaan antara keduanya terhadap lolos nya seleksi masuk perguruan tinggi.

Variabel jenis kelamin hanya terdiri dari 2, yaitu laki-laki dan perempuan.

2. Alamat

variabel alamat terdiri dari 12 kabupaten yang ada dilampung, sedangkan jika terdapat siswa yang beralamat di luar lampung, maka di kategorikan sebagai luar lampung.

3. Pendidikan orang tua

Atribut pendidikan orang tua dipilih karena berpengaruh terhadap motivasi untuk melanjutkan pendidikan ke jenjang yang lebih tinggi. Variabel pendidikan orang tua terdiri dari 4 tipe, yakni SMP, SMA, Diploma dan Sarjana.

4. Pekerjaan orang tua

Variabel pekerjaan orang tua terdiri dari 10 jenis, yakni petani, guru, PNS, wiraswasta, pegawai swasta, pedagang, buruh, dosen, sopir dan lain-lainnya.

5. Penghasilan orang tua

Penghasilan orang tua dibedakan menjadi 3 kategori sebagai berikut:

- a. Rp 1000.000 – Rp 2000.000
- b. Rp 2000.000 – Rp 3000.000
- c. Rp 3000.000 – Rp 4000.000

6. Jumlah saudara

Variabel jumlah saudara terdiri dari 6 kategori, jumlah 0 atau tidak mempunyai saudara hingga yang memiliki saudara berjumlah 5 orang.

7. Jenis SMP

Variabel jenis SMP merupakan jenis SMP asal siswa. Hanya 2 jenis, SMP dan MTs.

8. Status SMP

Variabel status SMP berisi tentang status kepemilikan dari sekolah tersebut. Dalam hal ini, hanya terdapat 2 status, yaitu sekolah dengan status negeri dan swasta.

9. Nilai mata pelajaran Ujian Nasional

Nilai Mapel UN yang di ambil pada kasus ini terdiri dari 3 mapel, yakni Matematika, Bahasa Indonesia dan Bahasa Inggris. Masing-masing mapel diambil nilainya dari semester 1 hingga semester 4.

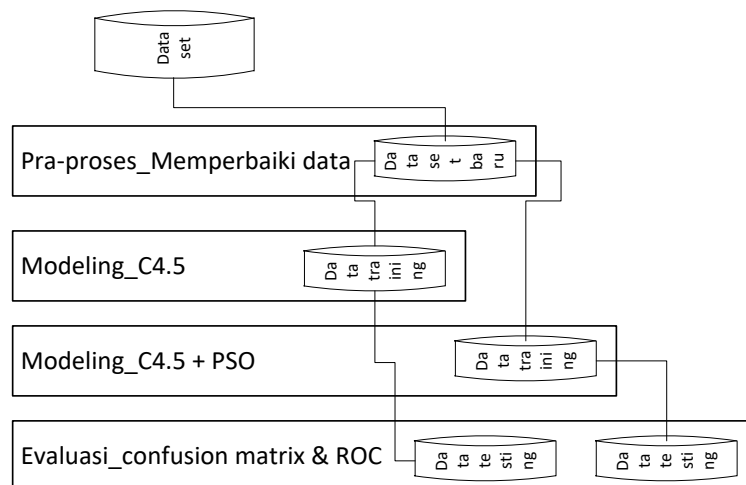
10. Nilai rata-rata UNSMP

Variabel nilai rata-rata mata pelajaran UN SMP merupakan nilai 3 mata pelajaran yang di ujikan secara nasional pada jenjang SMP. Adapaun pelajaran yang diambil sampelnya yakni Matematika, Bahasa Indonesia dan Bahasa Inggris.

3.1.4 Pemodelan (*modelling*)

Pada tahapan pemodelan ini, terdapat dua metode yang akan digunakan yaitu algoritma C4.5 yang selanjutnya dikombinasikan dengan algoritma Particle Swarm Optimization. Kombinasi dilakukan dengan mencari bobot dari setiap atribut. Untuk membandingkan dan pengujiannya, akan menggunakan framework RapidMiner sehingga akan terlihat peningkatan akurasi.

Gambar 4.4 Pemodelan penelitian



3.1.5 Validasi dan evaluasi

Dalam tahapan validasi dan evaluasi, akan menggunakan framework RapidMiner. Dengan framework tersebut akan dilakukan pengujian keakurasian dengan beberapa teknik yaitu *confusion matrix* dan kurva ROC. Cara perhitungan confusion matrix menggunakan persamaan dibawah ini:

$$Accuracy = \frac{a+d}{a+c+d+b} \text{ atau } \frac{TP+TN}{TP+FP+TN+FN}$$

3.1.6 Penyebaran (*Deployment*)

Hasil penelitian ini adalah analisa yang mengarah ke penunjang keputusan yang dapat digunakan oleh pihak sekolah untuk memberikan pertimbangan dalam penentuan langkah terkait upaya meloloskan siswa masuk ke perguruan tinggi melalui jalur seleksi, sehingga dapat meningkatkan kualitas dari sekolah tersebut.

3.2 Alat dan bahan penelitian

Dalam penelitian ini, untuk melakukan penambangan data, dibutuhkan perangkat keras dan lunak untuk menunjang kelancaran dalam proses penambangan. Perangkat tersebut diantaranya:

3.2.1 Kebutuhan Perangkat keras

Perangkat keras yang digunakan untuk melakukan penambangan data adalah sebagai berikut:

1. Processor core i3
2. RAM minimal 4 Gb
3. Minimal harddisk kosong 100 Gb

3.2.2 Perangkat lunak

Perangkat lunak yang digunakan untuk melakukan penambangan data adalah sebagai berikut:

1. Sistem operasi minimal windows 7
2. Software RapidMiner