

BAB III

METODE PENELITIAN

Metodologi atau tahapan penelitian diperlukan sebagai kerangka dan panduan proses penelitian, sehingga rangkaian proses penelitian dapat dilakukan secara terarah, teratur dan sistematis.

3.1 Pengumpulan Data

Pengumpulan data awal dimulai dengan pengumpulan data dalam bentuk data rekaman calon mahasiswa baru yang diperoleh dari UPT Pangkalan Data dan Pengembangan Sistem Informasi UNISNU Jepara yang berupa data calon mahasiswa baru tahun 2019-2020 yang digunakan sebagai dasar untuk memprediksi heregistrasi calon mahasiswa baru.

3.1.1 Observasi

Observasi merupakan sebuah cara pengumpulan data dengan pengamatan secara langsung dengan tujuan untuk memperoleh data atau informasi yang berkaitan dengan heregistrasi calon mahasiswa baru di UNISNU Jepara. Pada tahapan observasi, peneliti memperoleh data awal sebanyak 3.969 *record* dengan 18 (delapan belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 16 (enam belas) atribut reguler yaitu tahun pendaftaran, gelombang pendaftaran, jalur pendaftaran, program kelas, status awal pendaftaran, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jenis sekolah asal, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi yang berasal dari tahun 2019-2020 yang nantinya akan dilakukan *pre-processing* data dan digunakan sebagai *data training* dan *data testing*.

3.1.2 Wawancara

Wawancara dilakukan dengan cara mengajukan pertanyaan kepada bagian UPT Pangkalan Data dan Pengembangan Sistem Informasi yang memegang data

mahasiswa dan Biro III UNISNU yang memegang bidang penerimaan mahasiswa baru. Pada tahap wawancara ini juga mempunyai tujuan agar data atau informasi yang diperoleh dapat dipertanggungjawabkan.

3.1.3 Studi Pustaka

Studi Pustaka adalah metode dimana perolehan data-datanya melalui buku-buku, artikel website dan jurnal yang berkaitan dengan dengan penelitian ini. Data yang diperoleh dari buku-buku, artikel *website* dan jurnal mengenai algoritma data mining naive bayes dan prediksi heregistrasi calon mahasiswa baru.

3.2 Pengolahan Data Awal

Pada tahap ini meliputi proses pengolahan data calon mahasiswa baru tahun 2019-2020. Pada proses ini dilakukan seleksi terhadap data dan atribut yang akan digunakan untuk pemodelan karena tidak semua data dapat digunakan dan semua atribut akan digunakan. Berikut adalah beberapa teknik pengolahan data yang dilakukan :

1. Menyusun dan Mengkategorikan Data

Pada tahapan ini atribut yang dikonversi terdapat pada atribut tahun pendaftaran, usia, penghasilan orangtua, dan nilai UN. Sedangkan atribut nama, gelombang pendaftaran, jalur pendaftaran, program kelas, status awal pendaftaran, jenis kelamin, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, jenis sekolah asal, jurusan sekolah asal, informasi pendaftaran, dan status heregistrasi nilainya masih tetap seperti pada data awal. Berikut adalah kriteria atribut yang dikonversi untuk mengkategorikan data :

Tabel 3.1 Kriteria Atribut

No.	Atribut	Keterangan	Nilai Atribut
1	Tahun Pendaftaran	2019	I
		2020	II

2.	Usia	16-20 Tahun	I
		21-25 Tahun	II
		26-30 Tahun	III
		31-35 Tahun	IV
		36-40 Tahun	V
		41-45 Tahun	VI
		46-50 Tahun	VII
		51-55 Tahun	VIII
3.	Pendapatan Orangtua	< 1.000.000	I
		1.000.000-2.000.000	II
		2.000.000-3.000.000	III
		3.000.000-4.000.000	IV
		4.000.000-5.000.000	V
		> 5.000.000	VI
4.	Nilai UN	21-25	I
		26-30	II
		31-35	III
		36-40	IV
		41-45	V
		46-50	VI
		51-55	VII
		56-60	VIII
		61-65	IX
		66-70	X
		71-75	XI
		76-80	XII
		81-85	XIII
		86-90	XIV
		91-95	XV
		96-100	XVI
5.	Status Heregistrasi	Ya	
		Tidak	

2. Seleksi Atribut

Tahapan ini dilakukan untuk menyeleksi atribut yang akan digunakan dari semua atribut pada data awal [25]. Pada penelitian ini atribut yang akan digunakan yaitu 14 (empat belas) atribut dengan rincian 1 (satu) atribut ID yaitu nama, 12 (dua belas) atribut reguler yaitu tahun

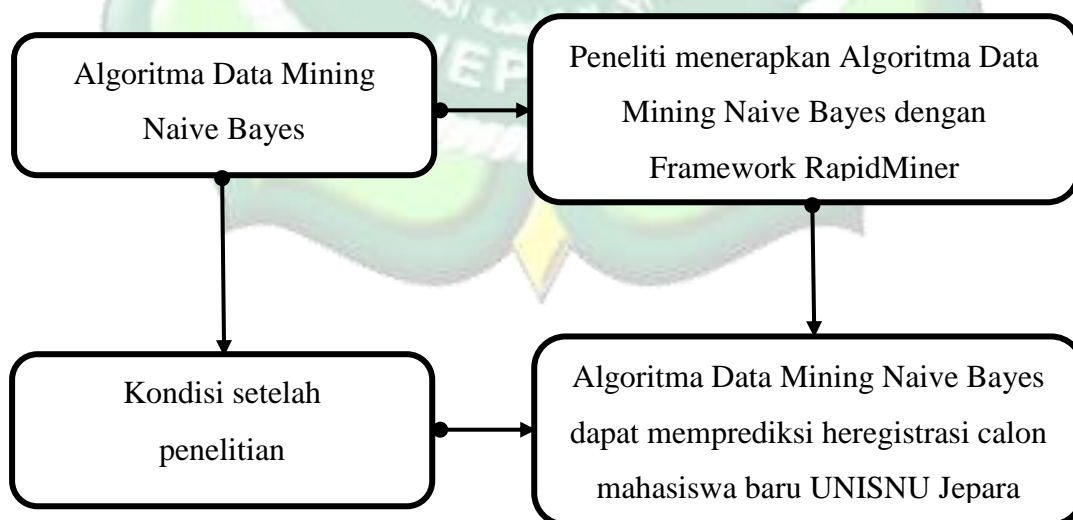
pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orangtua, jurusan sekolah asal, nilai UN, informasi pendaftaran, dan 1 (satu) atribut kelas yaitu status heregistrasi.

3. Pembersihan Data

Pada sebuah data terdapat data tidak bersih jika mengandung kotoran yang berupa nilai kosong (*missing value*). Semakin banyak missing value pada suatu data semakin tinggi pula tingkat kekotoran data tersebut. Maka diperlukan pembersihan data guna memberikan hasil yang baik dan mencegah terjadinya *error*. Pembersihan data dilakukan dengan cara mengisi dan membuang nilai-nilai yang kosong [26].

3.3 Metode Yang Diusulkan

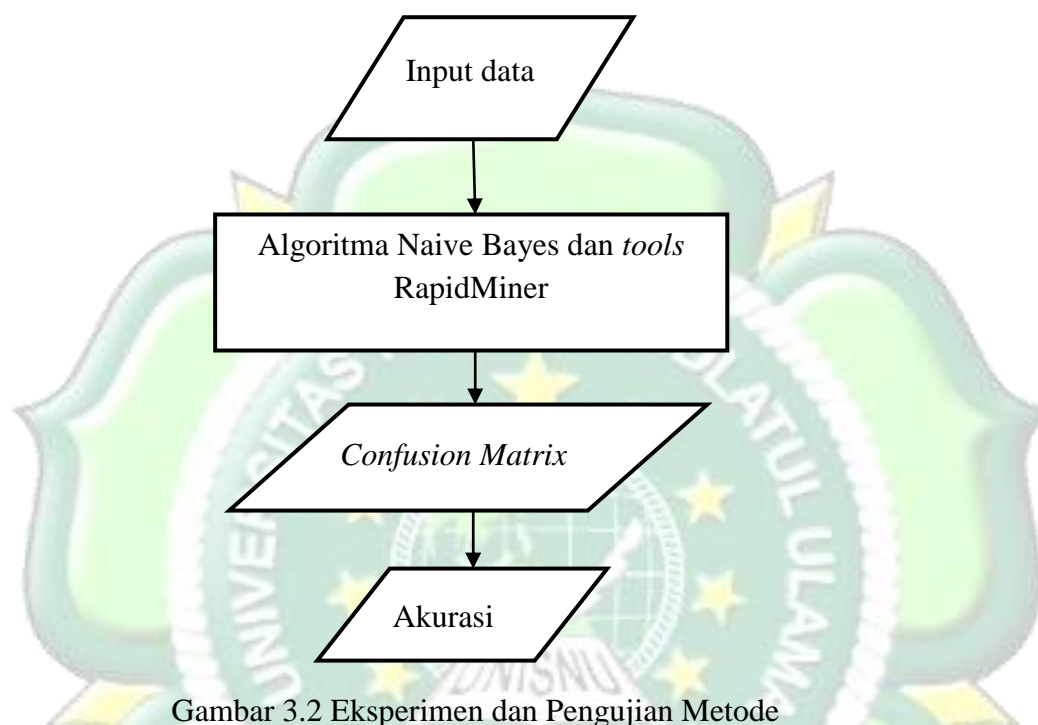
Pada tahap ini dijelaskan metode yang akan digunakan untuk prediksi heregistrasi calon mahasiswa baru. Proses secara bertahap dimulai dari pengumpulan data lalu pengolahan data/*pre-processing* data yaitu *integrasi*, *selection* dan *cleaning*. Selanjutnya dilakukan penerapan untuk prediksi heregistrasi calon mahasiswa baru menggunakan metode Klasifikasi dengan algoritma Naive Bayes yang diharapkan menghasilkan nilai akurasi yang tinggi.



Gambar 3.1 Metode yang diusulkan

3.4 Experimen dan Pengujian Metode

Pada tahap ini dilakukan penerapan model untuk menyelesaikan prediksi heregistrasi calon mahasiswa baru dengan metode Klasifikasi algoritma Naive Bayes menggunakan *tools* RapidMiner. Untuk mengetahui tingkat akurasi maka akan diuji dengan menggunakan metode *Confusion Matrix*.



Gambar 3.2 Eksperimen dan Pengujian Metode

3.5 Evaluasi dan Validasi Hasil

Pada tahap terakhir model yang diusulkan pada penelitian ini akan diuji dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi. *Confusion Matrix* akan menggambarkan tingkat akurasi mulai dari label *True Positive* (TP) yaitu prediksi positif yang benar, label *False Positive* (FP) yang menampilkan prediksi positif yang salah, label *False Negative* (FN) yang menampilkan prediksi negatif yang benar, dan label *True Negative* (TN) yang menampilkan prediksi negatif yang salah.

Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif) dibandingkan dengan seluruh *data testing*. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.