

BAB III

METODE PENELITIAN

1.1 Rancangan Penelitian

Metode penelitian pada penelitian ini adalah penelitian eksperimen dengan tahapan penelitian seperti contoh pada Gambar 3.1 :

1. Pengumpulan data

Penelitian ini diawali dengan melakukan pengumpulan data. Data yang diperoleh adalah data beasiswa di Universitas Peradaban.

2. Pengolahan data awal

Setelah data terkumpul, langkah selanjutnya adalah mengolah data yang ada. Menyeleksi atribut data yang akan digunakan dan data yang tidak relevan dihilangkan.

3. Model/ metode yang diusulkan

Peneliti mengusulkan model yang akan digunakan. Model yang digunakan tersebut berupa metode dalam teknik prediksi *data mining* yaitu klasifikasi dengan algoritma C4.5 yang dioptimisasi untuk pemilihan atribut yang digunakan.

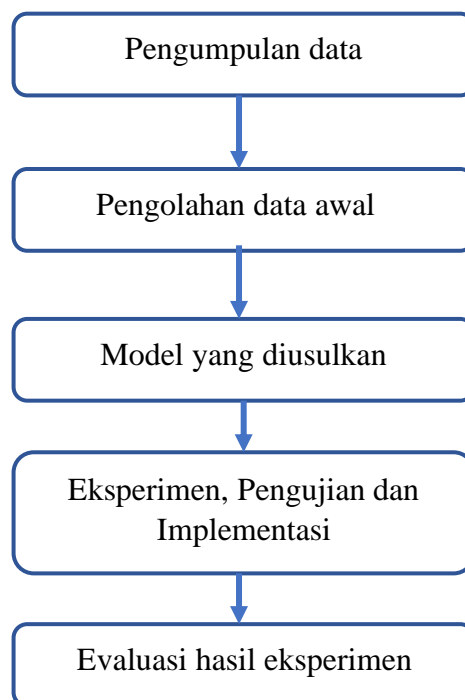
4. Eksperimen, pengujian model dan implementasi

Pengujian model, dari algoritma yang sudah ditentukan maka *dataset* yang ada diolah. Sehingga menghasilkan model yang diinginkan. Model yang dihasilkan selanjutnya diimplementasi dalam program.

5. Evaluasi dan validasi hasil

Setelah dilakukan eksperimen terhadap semua *dataset* dengan model yang diusulkan maka akan menghasilkan nilai-nilai akurasi dan performa. Kemudian hasil tersebut dianalisa dan dievaluasi. Dari hasil evaluasi dapat ditarik kesimpulan dari penelitian dan eksperimen ini

Pada penelitian ini akan mengusulkan sebuah metode klasifikasi data untuk prediksi penerima beasiswa. Hasil *dataset* yang baru akan diolah menggunakan algoritma C4.5 dan akan dibangun pohon keputusan untuk menghasilkan sebuah *rule*.



Gambar 3. 1 Metode Penelitian

1.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah sumber data mahasiswa yang terdaftar dalam seleksi penerima beasiswa yang diperoleh dari bagian kemahasiswaan Universitas Peradaban. Data diperoleh sebanyak 105 data (*record*), dengan atribut nama mahasiswa, status, nama mahasiswa sesuai penulisan pada rekening, nama bank, jenis kelamin, nim, jurusan, S1/D4/D3, semester, IPK, kedisiplinan, keaktifan dan hasil *dataset* tersebut bisa dilihat pada lampiran 1.2 tabel *dataset* beasiswa pada halaman 53.

1.3 Pengolahan Data Awal

Pada tahap pengolahan data langkah awal yaitu menetapkan atribut sebagai variabel keputusan dan *dataset* yang masih berisi data informasi pendukung, hal ini dilakukan pengolahan dengan membersihkan data informasi pendukung tersebut. Transformasi data pada masing-masing *dataset* dilakukan ke dalam bentuk yang dibutuhkan dengan menggunakan metode klasifikasi. Berikut merupakan salah satu contoh *dataset* beasiswa yang telah ditransformasikan dengan menghilangkan informasi pendukung seperti nama, jenis kelamin, jurusan semester dan lain-lain yang akan diwakili dengan atribut nim ditunjukkan pada Tabel 3.2. Sehingga

informasi yang dibutuhkan untuk selanjutnya dilakukan proses klasifikasi adalah *record* dan atribut dari IPK, kedisiplinan dan keaktifan.

Tabel 3. 1 Tabel *Dataset* yang Telah Ditransformasi

NIM	IPK	kedisiplinan	Keaktifan	Hasil
412140038	3.67	V	V	YA
42115027	3.86	K	V	YA
42116043	3.32	V	V	YA
42415017	3.42	V	V	YA
42415009	3.41	V	V	YA
40112034	3.32	X	X	TDK
40112038	3.56	X	X	TDK
40112035	3.47	X	X	TDK

Keterangan pada kolom kedisiplinan :

V : disiplin

K : kurang disiplin

X : tidak disiplin

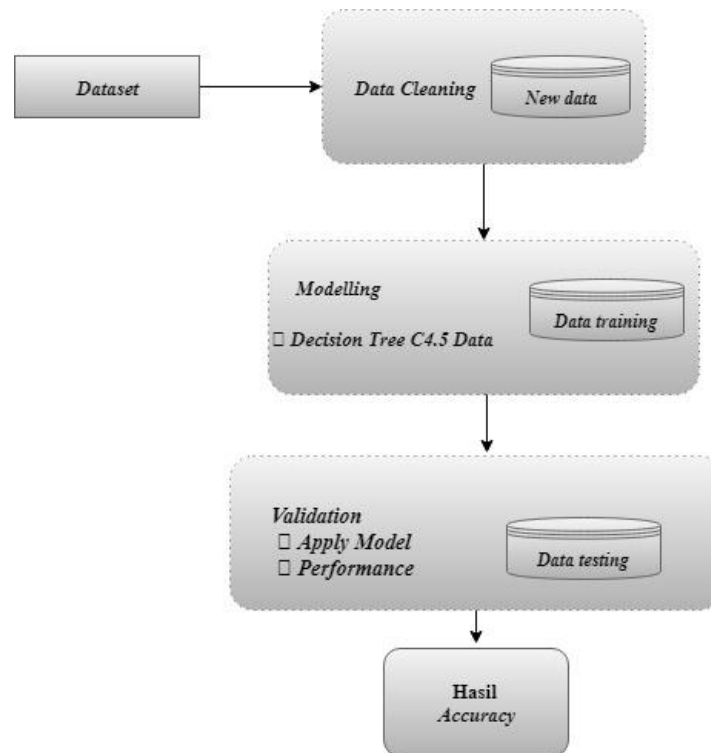
Keterangan pada kolom keaktifan :

V : aktif

X : tidak aktif

1.4 Metode yang Diusulkan

Metode ini akan digambarkan secara skematik dan disertai dengan formula perhitungan. Model akan dibentuk dari data yang sudah diolah, dan hasil pengolahan model akan diukur dengan model yang ada saat ini. Metode yang diusulkan ditunjukkan pada Gambar 3.2.



Gambar 3. 2 Metode yang Diusulkan

Pada penelitian ini dalam pengujian *dataset* pertama dilakukan adalah melakukan proses *preprocessing* yaitu *data cleaning*, setelah dilakukan *data cleaning* menghasilkan *dataset* baru. Setelah menghasilkan *dataset* baru akan dilakukan proses pemodelan menggunakan algoritma C4.5. dalam proses perhitungan algoritma C4.5 *dataset* akan dibagi menjadi dibagi jadi dua bagian yaitu *data training* dan *data testing*.

Pada pengujian validasi menggunakan *split validation* jumlah (k) yang dengan melakukan pemodelan dengan metode *decision tree* C4.5, selanjutnya memasukan *apply model* dan *performa*. *Validation* merupakan proses untuk mengevaluasi keakuratan prediksi dari model. Pada proses *split validation* menghasilkan *confusion matrix* dan akurasi dari model.

1.5 Eksperimen dan Pengujian Model

Tahapan eksperimen pada penelitian ini adalah:

1. Menyiapkan *dataset* beasiswa untuk eksperimen.
2. Melakukan pengolahan data awal untuk menghasilkan *dataset* baru.

3. Hasil *dataset* baru diolah menggunakan algoritma C4.5 untuk menghasilkan suatu model atau *rule*.
4. Pada perhitungan Algoritma C4.5 *dataset* akan dibagi jadi dua bagian yaitu data *training* dan data *testing*.
5. Melakukan evaluasi menggunakan *confusion matrix*.

Pada penelitian ini dalam melakukan eksperimen menggunakan komputer untuk melakukan proses perhitungan terhadap model yang diusulkan yang ditunjukkan pada Tabel 3.3. Penggunaan komputer dengan spesifikasi komputer sebagai berikut:

Tabel 3. 2 Spesifikasi Komputer yang Digunakan

<i>Processor</i>	Intel Core i5 7200U
Memori	4 GB
<i>Hardisk</i>	1 TB
Sistem Operasi	Windows 10

1.6 Evaluasi dan Validasi

Model yang terbentuk akan diuji dengan menggunakan *confusion matrix* untuk mengetahui tingkat akurasi. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi *negative* yang benar, dan prediksi *negative* 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.

Panduan untuk mengklasifikasikan AUC (*Area Under Curve*) [27].

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*