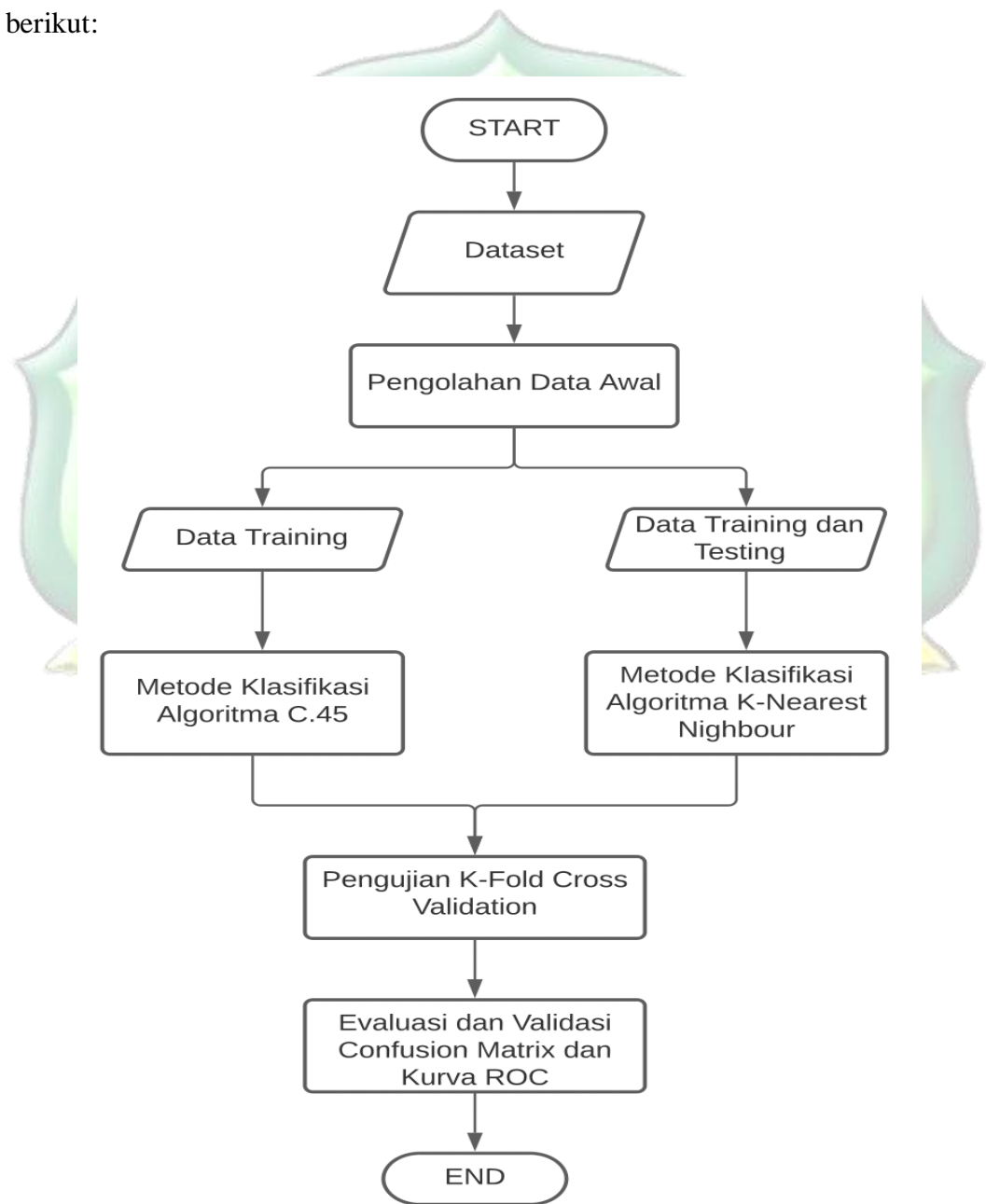


## BAB III

### METODE PENELITIAN

#### 3.1 Desain Penelitian

Pada tahapan penelitian ini peneliti menggunakan metode analisis kuantitatif. Adapun langkah-langkah dalam tahapan penelitian yang diilustrasikan sebagai berikut:



Gambar 3. 1 Metode Penelitian

Pada tahapan penelitian membutuhkan dataset yang akan digunakan sebagai data *training* dan data *testing*. Langkah selanjutnya yaitu pengolahan data awal untuk dilakukan pengujian *k-fold cross validation* pada data testing dan data training dengan menggunakan komparasi melalui dua metode yaitu Algoritma *K-Nearest Neighbour* (K-NN) dan Algoritma *C4.5*. Setelah pengujian, dilanjutkan dengan klasifikasi pada data testing dan data training yang belum diketahui kategorinya. Serta evaluasi dan validasi hasil menggunakan *confusion matrix* dan *curva ROC* (*Receiver Operating Characteristic*) sebagai *performance*.

### 3.1.1 Pengumpulan Data

Pada tahapan pengumpulan data, data yang didapatkan yaitu data mahasiswa fakultas sains dan teknologi pada tahun / ajaran 2016 – 2018 yang digunakan sebagai data training dan data testing. Adapun teknik dari pengumpulan data pada penelitian ini adalah sebagai berikut :

#### 1. Observasi

Observasi merupakan metode yang dilakukan untuk mengetahui objek yang akan diteliti secara langsung. Observasi dilakukan guna untuk mendapatkan data atau informasi terkait dengan data mahasiswa selama masa studi yang terdapat di Universitas Islam Nahdlatul Ulama (UNISNU) Jepara. Pada saat observasi peneliti mendapatkan data sebanyak 1.367 data. Data tersebut digunakan sebagai data *training* dan data *testing*, dari data mahasiswa selama masa studi yang diambil sebanyak 1.280, karena terdapat 87 data yang kosong dan belum diketahui. Dari data tersebut terdapat beberapa atribut, yaitu : NIM (Nomor Induk Mahasiswa), nama, prodi, kelas, jenis kelamin, angkatan, domisili, status menikah, pekerjaan orang tua, status mahasiswa, jumlah SKS, IPS 1, IPS 2, IPS 3, IPS 4, dan keterangan.

#### 2. Wawancara

Wawancara dilakukan dengan cara mengajukan pertanyaan kepada bagian UPT Pusat Data IT UNISNU yang memegang bidang masalah data mahasiswa yaitu Bapak Ahmad Zulfa, SE. Pada tahap wawancara ini juga mempunyai tujuan agar data atau informasi yang diperoleh dapat dipertanggungjawabkan.

### 3. Studi Pustaka

Studi pustaka merupakan teknik pengumpulan data dengan tinjauan pustaka ke perpustakaan dan pengumpulan bukubuku, bahan-bahan tertulis serta referensi-referensi yang relevan dengan penelitian yang sedang dilakukan. Studi kepustakaan juga menjadi bagian penting dalam kegiatan penelitian karena dapat memberikan informasi lebih mendalam yang berhubungan dengan masalah yang diangkat oleh peneliti [23].

#### 3.1.2 Pengolahan Data Awal

Pengolahan data awal merupakan tahap yang penting agar data yang akan digunakan mempunyai kualitas yang baik. Hal ini dilakukan karena tidak semua data dan atribut dapat digunakan. Berikut adalah beberapa teknik pengolahan data yang dilakukan :

##### 1. Menyusun dan Mengkategorikan Data

Menyusun dan mengkategorikan data diperlukan dengan mengkonversi atribut yang berupa nominal juga disebut kategorial menjadi atribut numeric memiliki nilai yang berupa kuantitas yang diukur dan dinyatakan dalam nilai *integer* atau *riil* [10]. Data yang diubah terdapat pada atribut status menikah, pekerjaan orang tua, status mahasiswa, jumlah SKS, IPS 1, IPS 2, IPS 3, IPS 4, dan keterangan. Berikut merupakan kriteria variable yang dikonversi :

Tabel 3. 1 Kriteria Variabel

No.	Atribut	Keterangan	Nilai Variabel
1.	Status Menikah	Menikah	1
		Belum Menikah	2
2.	Pekerjaan Orang Tua	TNI / POLRI	1
		PNS	2
		Pegawai Swasta	3
		Wiraswasta	4
		Sopir	5
		Petani / Nelayan	6

No.	Atribut	Keterangan	Nilai Variabel
		Tukang Kayu / Ukir / Tukang Batu	7
		Pensiun	8
		Pedagang / PKL	9
		Tidak Bekerja	10
		Lain-lain	11
3.	Status Mahasiwa	Lulus	1
		Aktif	2
		Non Aktif	3
		Keluar	4
		Drop Out	5
4.	Indeks Prestasi Semester (1)	3.51-4.00	1
		3.01-3.50	2
		2.51-3.00	3
		2.01-2.50	4
		1.51-2.00	5
		1.01-1.50	6
		0.51-1.00	7
		0.01-0.50	8
5.	Indeks Prestasi Semester (2)	3.51-4.00	1
		3.01-3.50	2
		2.51-3.00	3
		2.01-2.50	4
		1.51-2.00	5
		1.01-1.50	6
		0.51-1.00	7
		0.01-0.50	8
6.	Indeks Prestasi Semester (3)	3.51-4.00	1
		3.01-3.50	2
		2.51-3.00	3
		2.01-2.50	4
		1.51-2.00	5
		1.01-1.50	6
		0.51-1.00	7
		0.01-0.50	8
7.	Indeks Prestasi Semester (4)	3.51-4.00	1
		3.01-3.50	2
		2.51-3.00	3
		2.01-2.50	4

No.	Atribut	Keterangan	Nilai Variabel
		1.51-2.00	5
		1.01-1.50	6
		0.51-1.00	7
		0.01-0.50	8
8.	Jumlah SKS	122 - 144 sks	1
		97 - 121 sks	2
		73 - 96 sks	3
		49 - 72 sks	4
		25- 48 sks	5
		1 - 24 sks	6
9.	Keterangan	Tepat	
		Tidak Tepat	

## 2. Seleksi Atribut

Seleksi atribut dilakukan untuk mengambil variabel tertentu dari data awal yang diperoleh [9]. Pada penelitian ini variabel independen yang digunakan yaitu atribut NIM, nama, status menikah, pekerjaan orang tua, status mahasiswa, jumlah SKS, IPS 1, IPS 2, IPS 3, dan IPS 4. Sedangkan variabel dependen yang digunakan yaitu keterangan ketepatan studi mahasiswa dengan *output* tepat atau tidak tepat.

## 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 [10].

## 4. Normalisasi Data

Nilai-nilai atribut data yang berbeda perlu dilakukan normalisasi atau standarisasi agar proses data mining berjalan dengan baik [24]. Terdapat metode normalisasi data yaitu : *normalisasi min-max*, *normalisasi dengan penskalaan decimal* dan *normalisasi z-score*. Pada penelitian ini menggunakan *normalisasi min-max* berikut merupakan rumus *normalisasi min-max* [10] :



$$V' = \frac{v - \text{min}A}{\text{max}A - \text{min}A} \cdot (\text{new max}A - \text{new min}A) + \text{new min}A$$

Dimana :

$V'$  = Nilai dari data baru hasil dari normalisasi

$V$  = Nilai dari data sebelum dinormalisasi

New max A = Batas nilai maksimum terbaru

New min A = Batas nilai minimum terbaru

Max A = Nilai maksimum pada kolom

Min A = Nilai minimum pada kolom

### 3.1.3 Lokasi Penelitian

Penelitian ini dilaksanakan di Universitas Islam Nahdlatul Ulama (UNISNU) yang berada di Jl. Taman Siswa Pekeng Tahunan, Kabupaten Jepara, Provinsi Jawa Tengah. Peneliti memilih tempat tersebut dikarenakan belum adanya metode dalam menentukan ketepatan studi mahasiswa dengan mudah, cepat dan akurat yang dikarenakan jumlah mahasiswa yang masuk dengan yang lulus tidak seimbang tiap tahunnya.

### 3.1.4 Pengujian Model

Data akan diproses menggunakan dua metode yaitu *algoritma C4.5* dan *algoritma K-Nearest Neighbour* yang akan menghasilkan model kemudian akan dilakukan pengujian. Pengujian model dilakukan untuk memperkirakan estimasi dan akurasi dengan menggunakan *k-fold cross validation*. Percobaan yang akan dilakukan sebanyak  $k$ , Nilai  $k$  yang digunakan yaitu 10 atau disebut *10-fold cross validation* yang kemudian diambil nilai terbaik dan diaplikasikan ke dalam metode yang diusulkan.

### 3.1.5 Metode yang Diusulkan

Adapun metode yang diusulkan pada penelitian ini adalah metode klasifikasi data mining dengan *Algoritma C4.5* dan *K-Nearest Neighbor*. Data diolah menggunakan *tool* RapidMiner studio versi 9.8 dan akan menghasilkan suatu model. Berikut adalah tahapan pengolahan data menggunakan algoritma *K-Nearest Neighbor* pada RapidMiner:

1. Menentukan nilai  $k$  yang menghasilkan *performance* bernilai tinggi.
2. Menghitung *euclidean distance* pada masing-masing objek terhadap data *training* dan data *testing*.
3. Mengurutkan hasil perhitungan *euclidean distance* dan mengkategorikan berdasarkan tetangga terdekat terhadap nilai  $k$  yang ditetapkan.
4. Menentukan klasifikasi berdasarkan nilai  $k$ .

Tahapan pengolahan data menggunakan algoritma *C4.5* pada RapidMiner :

1. Menghitung entropy total dan gain yang paling besar.
2. Menentukan pohon keputusan yang akan menghasilkan sebuah rule yang akan diimplementasikan pada program.

### 3.1.6 Evaluasi dan Validasi Hasil

Data yang sudah diolah menggunakan algoritma *C4.5* dan *K-Nearest Neighbor* melalui *tool* RapidMiner versi 9.8 dan dilakukan pengujian model menggunakan 10-fold cross validation, selanjutnya dilakukan evaluasi dan validasi hasil menggunakan *confusion matrix* dan curva ROC (*Receiver Operating Characteristic*) sebagai *performance* yang akan menghasilkan nilai *Accuracy*. Penelitian ini diharapkan mampu memberikan hasil dengan nilai akurasi tinggi sehingga dapat meminimalisir kesalahan sasaran.