

## **BAB 3**

### **METODE PENELITIAN**

#### **3.1 Metode Pengumpulan Data**

##### **3.1.1 Sumber Data**

Jenis sumber data yang didapatkan peneliti adalah data primer dan data sekunder.

1. Data primer

Didapatkan peneliti secara langsung melalui teknik wawancara yang dilakukan terhadap Ketua Program Studi Teknik Informatika-S1 Universitas Dian Nuswantoro

2. Data sekunder

Peneliti dapatkan dari PSI Universitas Dian Nuswantoro yang berisi data mahasiswa, KHS, serta transkrip nilai mahasiswa Teknik Informatika-S1. Data mahasiswa yang didapatkan sebanyak 1903 diantaranya 555 mahasiswa angkatan 2011, 752 mahasiswa angkatan 2012, 596 mahasiswa angkatan 2013 dan terdiri dari atribut NIM, IPK, Kd\_dosenwali. Didapatkan 12520 *record* pada data KHS mahasiswa Teknik Informatika-S1 yang terdiri dari atribut Tahun Ajaran, NIM, dan IPS. Didapatkan 76952 *record* pada data transkrip nilai mahasiswa yang terdiri dari atribut NIM, KodeMK, dan Nilai. Sedangkan untuk data yang dapat mendukung penelitian ini penulis dapatkan dari buku, jurnal, internet dan sebagainya. Berikut ini pada Tabel 3.1, Tabel 3.2, dan Tabel 3.3 merupakan contoh data yang didapatkan dari PSI Universitas Dian Nuswantoro.

**Tabel 3.1 Contoh Data Mahasiswa**

nim	ipk	kd_dosenwali
A11.2011.05830	3,21	0686.11.1995.067
A11.2011.05832	2,75	0686.11.1995.067
A11.2011.05833	3,08	0686.11.1995.067
A11.2011.05835	3,37	0686.11.1995.067
A11.2011.05837	3,29	0686.11.1995.067
A11.2011.05838	2,85	0686.11.1995.067
A11.2011.05839	3,61	0686.11.1995.067
A11.2011.05840	2,9	0686.11.1995.067
A11.2011.05841	3,39	0686.11.1995.067

**Tabel 3.2 Contoh Data KHS Mahasiswa**

ta	nim	ips
20111	A11.2011.05830	3:07
20112	A11.2011.05830	3:01
20121	A11.2011.05830	3
20122	A11.2011.05830	3:05
20131	A11.2011.05830	2:09
20132	A11.2011.05830	3:18
20141	A11.2011.05830	3:05
20111	A11.2011.05832	2:05
20112	A11.2011.05832	2:04

**Tabel 3.3 Contoh Data Transkrip Nilai**

nim	kodeMK	nilai
A11.2011.05830	A11.44212	B
A11.2011.05830	A11.54101	A
A11.2011.05830	A11.54102	A
A11.2011.05830	A11.54103	A
A11.2011.05830	A11.54104	B
A11.2011.05830	A11.54105	B
A11.2011.05830	A11.54106	A
A11.2011.05830	A11.54201	B
A11.2011.05830	A11.54202	B

### **3.1.2 Pengumpulan Data**

Metode dalam pengumpulan data pada penelitian ini yaitu :

1. Wawancara

Wawancara dilakukan terhadap Ketua Program Studi Teknik Informatika-S1 Universitas Dian Nuswantoro untuk mengetahui secara dalam mengenai konsep dari peminatan tersebut.

2. Studi Pustaka

Studi pustaka dilakukan dengan cara mencari referensi yang terkait dengan penelitian ini berdasarkan dari buku, jurnal, internet dan sebagainya.

3. Survey

Survey penelitian dilakukan dengan meminta ijin untuk pengambilan data mengenai data mahasiswa dan transkrip nilai mahasiswa program studi Teknik Informatika-S1 angkatan 2011-2013 kepada PSI Universitas Dian Nuswantoro.

### **3.1.3 Jenis Data**

Jenis data yang digunakan oleh peneliti dalam penyelesaian permasalahan pada penelitian ini ialah data kuantitatif. Data kuantitatif tersebut yaitu:

1. NIM
2. IPK
3. Indeks Prestasi Semester 1-4
4. Rerata nilai matakuliah penunjang
5. Rerata nilai matakuliah peminatan
6. Peminatan

### 3.2 Metode Analisis

Metodologi pengembangan *data mining* yang digunakan ialah CRISP-DM. Berikut ini merupakan langkah-langkah yang dilakukan peneliti berdasarkan dari fase CRISP-DM :

#### 1. *Business Understanding*

Pada tahapan pertama penulis mencoba untuk memahami permasalahan yang ada dalam pemilihan jalur peminatan pada program studi Teknik Informatika-S1 Universitas Dian Nuswantoro. Sehingga dapat menentukan tujuan dan pola yang akan didapatkan dengan *data mining*. Adapun tujuan penelitian ialah mengolah data mahasiswa dalam jumlah besar sehingga didapatkan informasi yang berguna sebagai rekomendasi dalam pemilihan peminatan.

#### 2. *Data Understanding*

Pada tahap ini penulis melakukan pemahaman terhadap data yang dibutuhkan, untuk kemudian mengambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian. Adapun data yang digunakan yaitu data mahasiswa, KHS, dan transkrip nilai.

#### 3. *Data Preparation*

Pada tahap ini penulis mengolah data yang didapat dengan beberapa tahapan pada KDD seperti; melakukan pembersihan terhadap data / *data cleaning*, melakukan integrasi data / *data integration*, melakukan pemilihan data / *data selection* dan transformasi pada data / *data transformation*. *Tools* yang digunakan dalam menunjang pengolahan datanya ialah *microsoft excel*. Pada tahap mempersiapkan data, peneliti juga melakukan proses inisialisasi dan juga perhitungan rerata IPS 1-4, rerata nilai matakuliah penunjang, dan rerata nilai matakuliah peminatan. Berikut ini penjelasan mengenai tahap dalam mempersiapkan data :

a. Pembersihan Data dan Integrasi Data

Pembersihan data bertujuan untuk menghilangkan data yang tidak memiliki nilai (*null*), data yang salah *input*, data yang tidak relevan, dan data yang tidak konsisten. Sedangkan integrasi data bertujuan menggabungkan beberapa dataset hingga menjadi satu dataset.

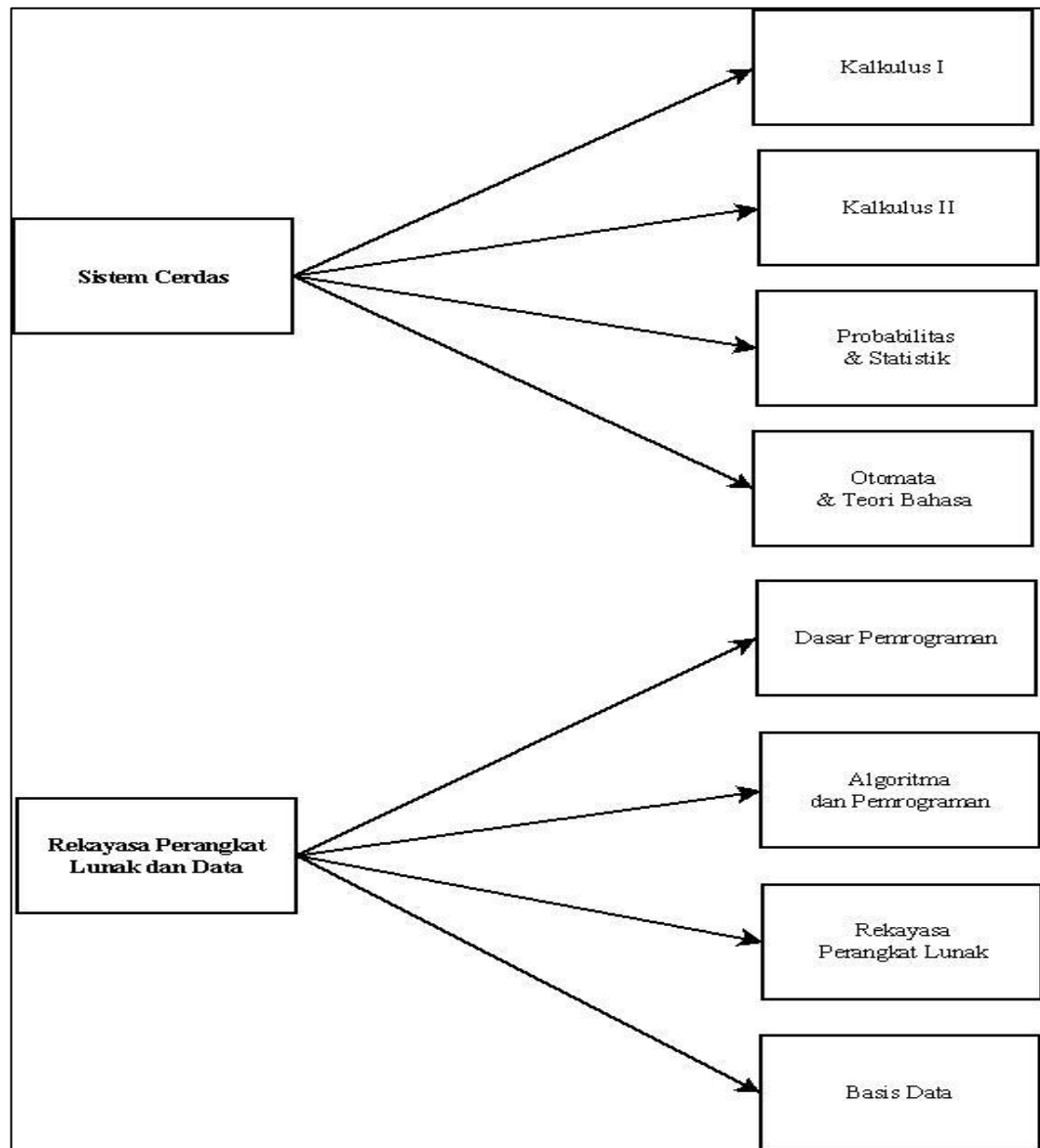
b. Seleksi Data dan Transformasi Data

Seleksi data dilakukan dengan cara pemilihan atribut-atribut yang sesuai dan terkait pada tujuan dari penelitian. Sedangkan transformasi data dilakukan dengan memberikan inisialisasi terhadap data yang memiliki nilai nominal menjadi bernilai numerik. Inisialisasi tersebut dilakukan pada atribut nilai matakuliah pada data transkrip nilai. Sebagai contoh; nilai A akan diberi inisialisasi menjadi angka 5. Dengan begitu semua nilai A akan diubah menjadi angka 5. Pada Tabel 3.4 menjelaskan mengenai perubahan terhadap data nominal menjadi data numerik pada atribut nilai matakuliah.

**Tabel 3.4 Inisialisasi pada Atribut Nilai Matakuliah**

A	:	5
B	:	4
C	:	3
D	:	2
E	:	1

Setelah itu akan dicari rerata nilai matakuliah penunjang dari masing-masing peminatan. Dalam perhitungan rerata nilai matakuliah penunjang, sebelumnya harus dilihat terlebih dahulu hubungan dari matakuliah penunjang dengan masing-masing peminatan. Berikut ini pada Gambar 3.1 menjelaskan hubungan dari matakuliah penunjang dengan masing-masing peminatan.



**Gambar 3.1 Hubungan Jalur Peminatan dan Matakuliah Penunjang**

Berikut ini merupakan perhitungan rata-rata yang dilakukan pada nilai matakuliah penunjang SC dan nilai matakuliah penunjang RPLD :

$$\text{Rerata MK Penunjang Sistem Cerdas} = \frac{\text{Kalkulus I} + \text{Kalkulus II} + \text{Probabilitas \& Statistik} + \text{Otomata}}{4} \quad (3.1)$$

Rerata MK Penunjang Rekayasa Perangkat Lunak dan Data

$$= \frac{\text{Dasar Pemrograman} + \text{Algoritma} + \text{Pemrograman} + \text{RPL} + \text{Basis Data}}{4} \quad (3.2)$$

Dalam mendapatkan atribut peminatan, atribut rerata nilai MK peminatan SC, dan atribut rerata nilai MK peminatan RPLD diketahui berdasarkan matakuliah peminatan yang dipilihnya dan juga nilai yang diperoleh pada matakuliah peminatan tersebut. Berikut pada Tabel 3.5 merupakan beberapa matakuliah peminatan yang digunakan dalam perhitungan rerata nilai MK peminatan :

**Tabel 3.5 Daftar Matakuliah Peminatan**

<b>Peminatan</b>	<b>Matakuliah Peminatan</b>	<b>Kode Matakuliah</b>
Sistem Cerdas	Strategi Algoritma	54508
	Komputer Grafik	54503
	Pengolahan Citra Digital	54521
Rekayasa Perangkat Lunak dan Data	Rekayasa Perangkat Lunak Lanjut	54504
	Sistem Basis Data	54505
	Pemrograman Internet	54816

Berikut ini merupakan perhitungan rata-rata yang dilakukan pada nilai matakuliah peminatan SC dan nilai matakuliah peminatan RPLD :

Rerata MK Peminatan Sistem Cerdas

$$= \frac{\text{Strategi Algoritma} + \text{Komputer Grafik} + \text{Pengolahan Citra Digital}}{n} \quad (3.3)$$

Rerata MK Peminatan Rekayasa Perangkat Lunak dan Data

$$= \frac{\text{RPL Lanjut} + \text{Sistem Basis Data} + \text{Pemrograman Internet}}{n} \quad (3.4)$$

Sedangkan dalam menghitung rata-rata IPS ke-1 sampai dengan ke-4 dilakukan seperti berikut :

$$\text{IPS 1 - 4} = \frac{\text{IPS 1} + \text{IPS 2} + \text{IPS 3} + \text{IPS 4}}{4} \quad (3.5)$$

#### 4. *Modeling*

Pada tahap ini penulis menentukan teknik *data mining* yang digunakan untuk mengolah data yang sudah disiapkan sebelumnya. Teknik yang dilakukan yaitu dengan *clustering* menggunakan algoritma *K-means*. Data yang sudah melalui proses pengolahan kemudian akan dilakukan perhitungan dengan menggunakan *tools rapidminer*. Dua langkah yang dilakukan pada tahap ini ialah :

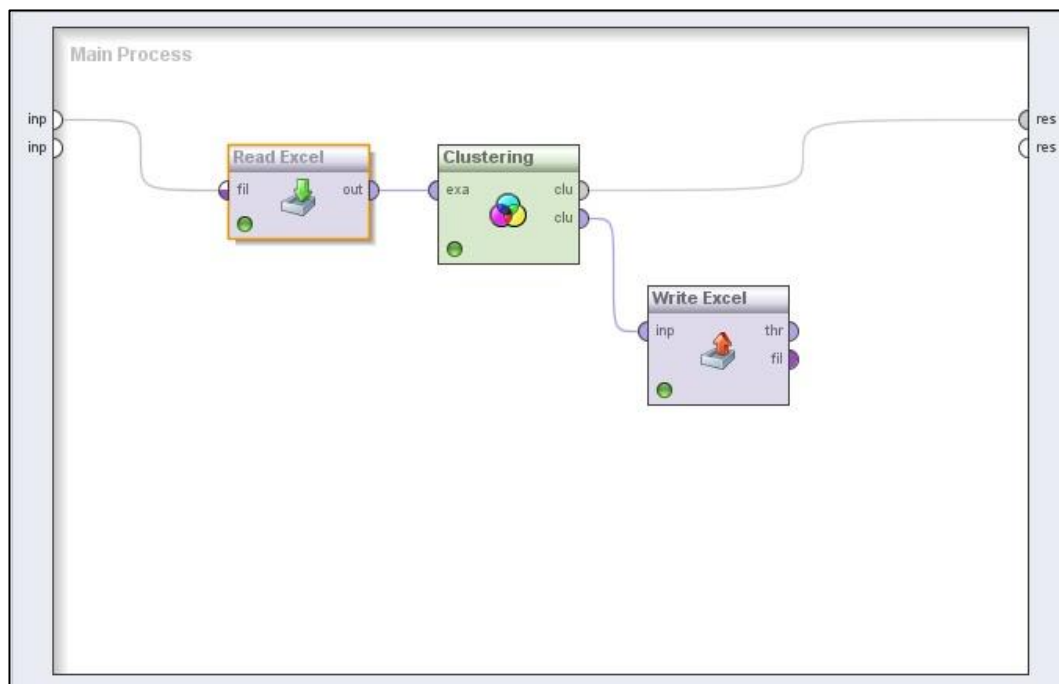
##### a. Perhitungan *K-Means* secara manual

Data yang akan digunakan dalam perhitungan *K-Means* secara manual yaitu 50 sampel data. Dimana data tersebut merupakan data yang sama dengan perhitungan menggunakan *rapidminer*, yang diambil secara acak oleh peneliti.

##### b. Penerapan *K-Means clustering* menggunakan *rapidminer*

Penerapan dengan menggunakan *rapidminer* bertujuan untuk memudahkan dalam pemrosesan data yang berjumlah besar. Proses perhitungan *K-Means* menggunakan *rapidminer* dilakukan dengan model seperti pada Gambar 3.2 berikut ini.





**Gambar 3.2** Proses *Clustering* pada *Rapidminer*

Berikut merupakan penjelasan dari model proses perhitungan *K-Means* menggunakan *rapidminer*.

- 1) *Read Excel* : merupakan operator yang digunakan dalam mengambil data yang sudah dipersiapkan dan memiliki format xls ataupun xlsx.
- 2) *Clustering* : merupakan operator yang digunakan dalam melakukan proses *clustering* dengan algoritma yang dipilih.
- 3) *Write Excel* : merupakan operator yang digunakan untuk menyimpan *output* yang dihasilkan dalam format xls ataupun xlsx. Penggunaan operator *write excel* dapat membantu dalam menganalisa lebih lanjut data yang didapatkan, dengan menggunakan *microsoft excel*.

## 5. *Evaluation*

Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama. Untuk itu akan dilakukan profilisasi pada setiap *cluster* yang telah terbentuk, untuk diketahui karakteristik pada kelompok tersebut. Disamping itu untuk diketahui kesesuaian dengan jalur peminatan akan dilakukan analisis lebih lanjut untuk dihubungkan dengan atribut peminatan, rerata nilai MK peminatan SC, dan rerata nilai MK peminatan RPLD. Sehingga diharapkan mendapatkan informasi atau pola yang berguna sebagai rekomendasi pemilihan jalur peminatan yang sesuai dengan kemampuan akademis mahasiswa. Langkah yang dilakukan pada tahap ini ialah :

### a. Profilisasi pada *cluster*

Data hasil pengelompokan tersebut, selanjutnya akan dianalisis lebih lanjut dengan bantuan *microsoft excel*. Profilisasi tersebut bertujuan untuk mengidentifikasi dan menampilkan karakteristik yang didapatkan dari setiap *cluster* yang terbentuk.

### b. Menentukan pemilihan jalur peminatan

Berdasarkan hasil dari pengelompokan kemampuan akademis mahasiswa beserta karakteristik yang didapatkan dari setiap *cluster* yang terbentuk, kemudian akan dianalisis lebih lanjut untuk dihubungkan dengan atribut peminatan, atribut rerata nilai MK peminatan SC, dan atribut rerata nilai MK peminatan RPLD sehingga dapat diketahui kesesuaian kemampuan akademis mahasiswa dengan masing-masing jalur peminatan. Informasi dan pola yang didapatkan tersebut untuk kedepannya dapat digunakan sebagai rekomendasi pemilihan jalur peminatan yang sesuai dengan kemampuan akademis mahasiswa.

### c. Mengukur tingkat akurasi rekomendasi

Dilakukan agar dapat mengetahui tingkat akurasi dari rekomendasi yang dihasilkan dengan peminatan yang dipilih mahasiswa, jika dibandingkan dengan sistem rekomendasi lama.

## 6. *Deployment*

Pada tahap ini penulis mencoba menyajikan hasil dari penerapan *data mining* metode *clustering* dengan algoritma *K-Means* yang telah dilakukan.