

BAB 3

METODE PENELITIAN

3.1 Koleksi Data

Dalam menyusun skripsi ini, diperlukan data-data informasi yang relatif lengkap sebagai bahan yang mendukung kebenaran materi pembahasan sehingga dilakukan pengumpulan data untuk mendapatkan informasi atau materi yang diperlukan. Adapun parameter yang digunakan dalam penelitian ini pada tabel 3.1. Kemudian, data yang diambil ialah data mahasiswa angkatan 2015, 2016 dan 2017. Angkatan 2015 dan 2016 ada 16 data yang peneliti ambil secara langsung menanyakan nilai semester 1 sampai nilai semester 7 dan 4 data yang mengisi di kuesioner sedangkan angkatan 2017 37 mahasiswa mengisi di kuesioner.

Tabel 3.1 Parameter

^	Keterangan	Jenis data
Nama	Nama mahasiswa	Text
Nilai Semester 1	Nilai Semester 1	Kategori
Nilai Semester 2	Nilai Semester 2	Kategori
Nilai Semester 3	Nilai Semester 3	Kategori
Nilai Semester 4	Nilai Semester 4	Kategori
Nilai Semester 5	Nilai Semester 5	Kategori
Nilai Semester 6	Nilai Semester 6	Kategori
Nilai Semester 7	Nilai Semester 7	Kategori

3.2 *PreProcessing*

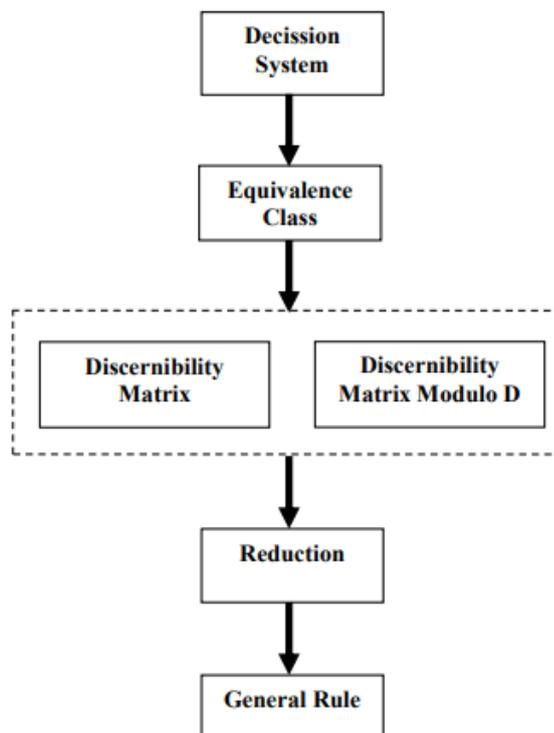
Pada tahap ini data yang di ubah menjadi kategori adalah nilai semester 1 sampai nilai semester 7 menggunakan excel yang mana di ubahnya secara manual dari numerik ke kategori. Indeks Prestasi Semester (IPS) 4 masuk ke kategori A, Indeks Prestasi Semester (IPS) 3,5 sampai 3,9 masuk ke kategori AB, Indeks Prestasi Semester (IPS) 3 sampai 3,4 masuk ke kategori B, Indeks Prestasi Semester (IPS) 2,5 sampai 2,9 masuk ke kategori BC, Indeks Prestasi Semester (IPS) 2 sampai 2,4 masuk ke kategori C, Indeks Prestasi Semester (IPS) 1 sampai 1,9 masuk ke kategori D, Indeks Prestasi Semester (IPS) 0 sampai 0,9 masuk ke kategori E.

3.2.1 Data Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal. Dalam penelitian ini proses *clening* ada 1 data yang tidak bisa di pakai maka dari itu data tersebut peneliti hapus di dataset.

3.3 Model Metode Rough Set

Untuk menyelesaikan permasalahan penelitian dengan menggunakan metode *rough set*, maka dapat mengikuti proses metode *rough set* seperti pada Gambar 3.1.



Gambar 3.1 Model Metode *Rough Set*

3.3.1 Decision System

Dalam penggunaan *information system*, terdapat *outcome* dari klasifikasi yang telah diketahui yang disebut dengan atribut keputusan. *Information system* tersebut disebut dengan *Decision system* merupakan fungsi yang mendeskripsikan *information system*, yaitu dimana $DS=\{U,(A,C)\}$, U = objek, A = atribut kondisi, C = atribut keputusan (Prajana 2016).

3.3.2 Equivalence Class

Equivalence class adalah mengelompokan objek-objek yang sama untuk *attribute* $A = (U,A)$ and $B \subset A$, *discernibility matrix* dari A adalah MB , dimana tiap-tiap *entry* $MB(I,j)$ terdiri dari sekumpulan *attribute* yang berbeda antara objek X_i dan X_j (Sembiring and Azhar 2013).

3.3.3 Discernible Matrix Modulo D

Didefinisikan seperti berikut dimana $MB(I,j)$ adalah sekumpulan *attribute* yang berbeda antara objek X_i dan X_j dan juga berbeda *attribute* keputusan. Diberikan sebuah I *discernibility Matrix Module D* dari A , MBd . (Sembiring and Azhar 2013)

3.3.4 Reduct

Reduct adalah penyeleksian *attribut* minimal (*interesting attribute*) dari sekumpulan *attribute* kondisi dengan menggunakan *Prime Implicant* fungsi *Boolean*. Kumpulan dari semua *Prime Implicant* mendeterminasikan *sets of reduct* (Sembiring and Azhar 2013).

3.3.5 Derated Rules

Proses selanjutnya yaitu mendapatkan pengetahuan dalam *database* melalui ekstraksi aturan dari sistem keputusan. Hasil keputusan tersebut didasarkan pada proses *reduct* (Saifudin 2017).

3.4 Validasi

Validasi adalah pemeriksaan dan memastikan bahwa data tersebut telah sesuai kriteria yang ditetapkan. Penelitian ini data akan dibagi menjadi 2 bagian

yakni 70% untuk data *training* dan 30 % data *testing*. Kemudian, untuk mengevaluasi model, maka digunakan akurasi, *presisi* dan *recall*.