

BAB 3

METODOLOGI 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.

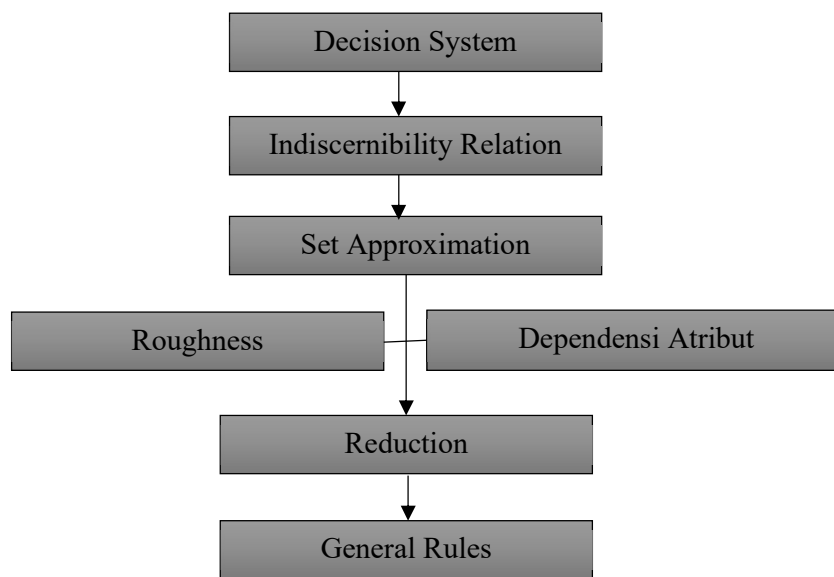
Table 3.1 Parameter Penelitian

Parameter	Keterangan	Jenis Data
Nama	Nama Mahasiswa	<i>Text</i>
NIM	NIM Mahasiswa	Numerik
Jenis Kelamin	Jenis Kelamin Mahasiswa	Kategori
Jumlah Keluarga	Jumlah Keluarga Mahasiswa	Kategori
Tempat Tinggal Asal	Tempat Tinggal Asal Mahasiswa	Kategori
Saudara Laki-Laki	Saudara Laki-Laki Mahasiswa	Kategori
Saudara Perempuan	Saudara Perempuan Mahasiswa	Kategori
Tempat Tinggal Sekarang	Tempat Tinggal Sekarang Mahasiswa	Kategori
Transportasi Ke Kampus	Transportasi Ke Kampus Mahasiswa	Kategori
Jenis Pendidikan SMA/SMK	Jenis Pendidikan SMA/SMK Mahasiswa	Kategori
Nilai Pendidikan SMA/SMK	Nilai Pendidikan SMA/SMK Mahasiswa	Kategori/Numeric
Lokasi SMA/SMK	Lokasi SMA/SMK Mahasiswa	Kategori
Olahraga/Atlet	Olahraga/Atlet Mahasiswa	Kategori
Jenis Olahraga	Jenis Olahraga Mahasiswa	Kategori
Pekerjaan Ayah	Pekerjaan Ayah Mahasiswa	Kategori
Pendapatan Ayah	Pendapatan Ayah Mahasiswa	Kategori/Numeric

Pekerjaan Ibu	Pekerjaan Ibu Mahasiswa	Kategori
Pendapatan Ibu	Pendapatan Ibu Mahasiswa	Kategori/ Numeric
IP Semester 2	IP Semester 2 Mahasiswa	Kategori/Numeric
IPK	Indek Prestasi Mahasiswa	Kategori/Numeric

Tabel 3.1 parameter penelitian tabel ini menunjukkan data parameter yang di gunakan untuk mempredeksi mahasiswa, data yang diambil dari data mahasiwa teknik informatika angkatan 2019

3.2 Model Metode Rough Set



Gambar 3.2 Model Metode Rough Set

Di dalam Gambar 3.1 Metode *Rough set* terdapat beberapa langkah - langkah penyelesaian masalah, yaitu sebagai berikut:

1. *Decision System* tersebut dilakukan teknik klasifikasi kriteria yang disebut "*Indiscernibility Relation*"
2. Menentukan perkiraan (prediksi) disebut "*Set Approximation*"
3. Kemudian dilakukan proses *Discernibility* "*Roughset* dan *Dependensi Atribut*"
4. Proses "*Reduction*"
5. Untuk memperoleh hasil akhir dilakukan proses "*General Rules*" Langkah-langkah dalam menjalankan metode *rough set* di atas.

3.2.1 Decision System

Decision System adalah sistem informasi dengan atribut tambahan yang disebut atribut pengambilan keputusan, yang disebut kelas atau target dalam data mining. Atribut ini mewakili hasil klasifikasi yang diketahui.

Sebuah sistem pengambilan keputusan adalah fungsi yang menggambarkan sebuah sistem informasi. Artinya, terlihat seperti ini:

$$DS = \{U, (A, C)\} \dots\dots\dots (1)$$

U= objek

A=atribut kondisi

C=atribut keputusan

3.2.2 Indiscernibility Relation

Dalam *decision system*, sebuah objek dapat memiliki nilai yang sama untuk sebuah atribut kondisionalnya. Contohnya, pasien P1, P4, dan P6 memiliki nilai atribut kondisional *headache* yang sama, yaitu "no". Hubungan tersebut disebut dengan indiscernible (tidak dapat dipisah).

Misalkan $SS = (UU, AA)$ adalah information system, dan $BB \subseteq AA$. Maka sebuah *indiscernibility relation* objek-objek menurut atribut B yang dilambangkan dengan $IIIII_{SS}(BB)$, dapat didefinisikan sebagai: $IIIII_{SS}(BB) = \{(xx, xx') \in UU \mid \forall aa \in BBB (xx = aa(xx'))\}$ (4) disebut sebagai *B-indiscernibility relation*. $IIIII_{SS}(BB)$ merupakan *equivalence relation*. Jika $(xx, xx') \in IIIII_{SS}(BB)$ maka objek x dan x' adalah objek yang tidak dapat dibedakan (*indiscernible*) satu sama lain oleh atribut B. Kelas-kelas yang ekuivalen dengan *B-indiscernibility relation* dinotasikan dengan $[xx]$

3.2.3 Set Approximation

Set approximation bertujuan untuk menentukan apakah objek-objek yang terbentuk memiliki kepastian keputusan atau mungkin dapat diklasifikasikan kedalam kelas keputusan atau mungkin dapat diklasifikasikan kedalam kelas keputusan tertentu atas dasar pengetahuan yang ada. Untuk *decision system*, sangat penting menemukan seluruh subset menggunakan kelas yang ekuivalen yaitu yang mempunyai nilai kelas yang sama. Tetapi, subset ini tidak selalu didefinisikan dengan tepat.

3.2.4 Roughness

Roughness dari himpunan ini memberikan inkarnasi baru pada ukuran kesamaan untuk berbagai informasi yang melibatkan informasi ketidakpastian dalam *set universal* yang sama. Selain itu, teori *multiset neutrofik* masih kasar juga diterapkan dalam pengambilan keputusan untuk diagnosis medis. Hasil perbandingan menunjukkan bahwa memperkirakan kekasaran informasi penting untuk diperoleh hasil terbaik untuk ukuran kemiripan terdekat.

3.2.5 Dependensi Atribut

1. Nilai indistinguishable pertama yang dicari adalah 1 indistinguishability, yang merupakan kombinasi atribut terkecil.
2. Kemudian lakukan pencarian untuk atribut dependensi. Jika nilai atribut ketergantungan yang diperoleh adalah 1, itu adalah variabel yang membuat himpunan minimum variabel tidak bisa dibedakan.
3. Jika pada proses pencarian kombinasi atribut tidak ditemukan dependency attributes = 1, maka lakukan pencarian kombinasi yang lebih besar, di mana kombinasi variabel yang dicari adalah kombinasi dari variabel di tahap sebelumnya yang nilai dependency attributes paling besar.
4. Lakukan proses (3), sampai didapat nilai dependency attributes = 1.

3.2.6 Reduction

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

3.2.7 Generating Rules

Generating rules adalah suatu metode *rough set* untuk menghasilkan *rules/knowledge* berdasarkan equivalence class dan reduct. Proses selanjutnya yaitu mendapatkan pengetahuan dalam *database* melalui ekstraksi aturan dari sistem keputusan. Hasil keputusan tersebut didasarkan pada proses *reduct*.

3.3 Validasi

Validasi adalah langkah pemeriksaan untuk memastikan bahwa data tersebut telah sesuai kriteria yang ditetapkan. Dengan tujuan bahwa data yang akan

dimasukkan ke dalam *rough set* telah diketahui dan dapat dijelaskan kebenaran datanya untuk penelitian ini. Data akan dibagi menjadi 2 bagian yaitu 70 % data *training* dan 30 % data *testing*. Kemudian untuk mengevaluasi model, maka digunakan akurasi, presisi dan *recall*.