

## **BAB 3**

### **METODE PENELITIAN**

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

Penelitian dilakukan pada Sekolah Menengah Atas (SMA) Pangudi Luhur Don Bosko Semarang untuk mengelompokkan dalam bidang pendidikan yang berdasarkan nilai raport dengan metode clustering algoritma *Fuzzy C-Means*.

#### **3.2 Metode Analisis**

Data yang digunakan pada penelitian diperoleh berdasarkan data sekunder yaitu dataset yang diperoleh berupa data yang sudah jadi. Dataset yang didapatkan berasal dari Sekolah Menengah Atas (SMA) Pangudi Luhur Donbosko. Dataset bersifat kuantitatif karena dataset yang didalamnya berupa informasi dalam bentuk angka atau bilangan. Teknik pengambilan data dengan melakukan dokumentasi ke Sekolah Menengah Atas (SMA) Pangudi Luhur Don Bosko.

Instrumen penelitian dalam penelitian ini merupakan alat yang dibutuhkan untuk mengumpulkan dan mengolah dataset. Instrumen penelitian ini menggunakan teknik dokumentasi yaitu teknik pengambilan data yang dilakukan dengan mencari garis besar dari data yang dibutuhkan yaitu berupa data responden yang berjumlah 1000 data untuk dijadikan sampel dalam penelitian yang dilakukan.

#### **3.3 Metode Analisis**

##### **3.3.1 Data Mining**

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan *knowledge discovery in databases* (KDD) yang terdiri dari Database, Data Cleaning, Data integration, Task relevan data, Data transformation, Data mining, Pattern evolution, knowledge :

1. Database

Kumpulan data yang saling berkaitan digunakan untuk melakukan perancangan guna memenuhi kebutuhan informasi organisasi:

2. Data Cleaning

Dataset yang diperoleh baik dari database perusahaan ataupun dari hasil melakukan penelitian memiliki hasil yang kurang sempurna seperti data yang hilang, tidak valid atau sekedar salah ketik. Dataset yang tidak relevan itu juga seharusnya dibuang karena mengurangi mutu atau keakuratan dari hasil akhir dalam data mining. Pembersihan data juga berguna untuk mempengaruhi hasil dari system data mining karena data yang dikerjakandapat berkurang jumlah kesulitannya.

3. Data integration

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi tujuan unik berupa atribut nama, nomor pegawai, tempat lahir, agama dan lain sebagainya. Integrasi dilakukan secara cermat karena apabila terjadi kesalahan dapat menghasilkan data yang menyimpang, bahkan menyesatkan pengambilan tindakan selanjutnya.

4. Task relevan data

Setelah semua sumber data diintegrasikan menjadi gabungan pada database maka tahap selanjutnya yaitu task relevant data. Pada tahap ini dilakukan relevansi atribut dari data relevant sesuai target atau output yang akan dihasilkan.

5. Data transformation

Coding yaitu tahap mentransformasi data yang dipilih sehingga data sesuai untuk diproses ke dalam data mining. Tahap data transformasi dalam KDD merupakan proses kreatif dan mengikuti jenis atau pola informasi yang akan dicari dalam database.

6. Data mining

Data mining merupakan tahap meneliti dan menganalisa data dalam jumlah besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dengan menggunakan teknik atau metode tertentu. Metode atau algoritma sangat bergantung pada tujuan dan proses dalam keseluruhan KDD.

7. Pattern evaluation

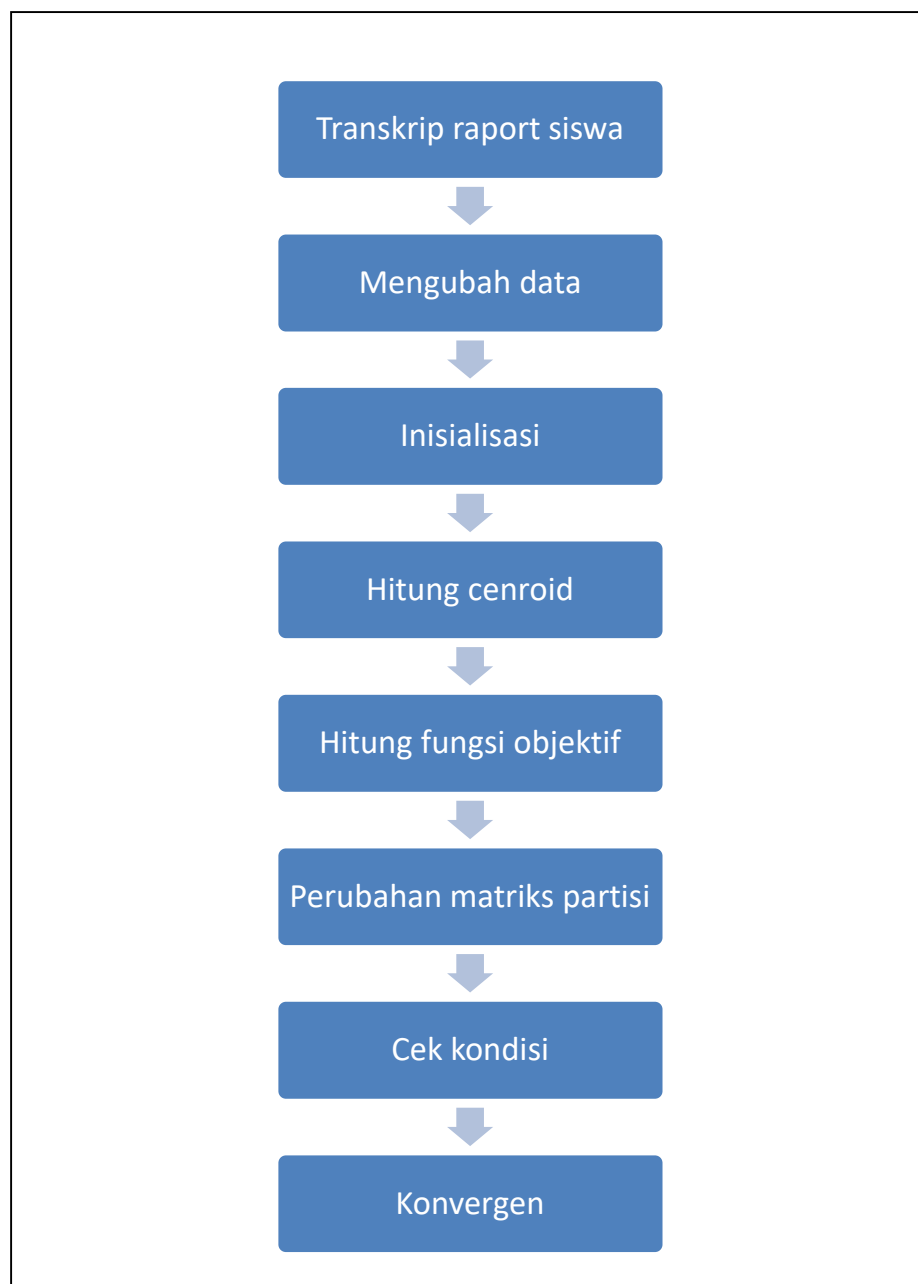
Tahap ini adalah hasil dari teknik data mining yang berupa pola-pola khusus yang dievaluasi untuk menilai hasil pencapaian hipotesa. Apabila hasil yang diperoleh tidak sesuai dengan hipotesamaka dapat dilakukan beberapa alternatif yang bisa digunakan seperti, menjadikannya sebagai umpan balik untuk memperbaiki data mining lain yang lebih sesuai.

8. Knowledge

Metode terakhir proses data mining yaitu menyusun keputusan dan tindakan dari hasil analisis yang diperoleh.

### 3.3.2 Algoritma *Fuzzy C-Means*

Berdasarkan tahapan KDD selanjutnya dimasukkan dalam tahapan blok diagram, hal ini akan membahas garis besar langkah-langkah penghitungan menggunakan metode algoritma *Fuzzy C-Means* yang dijabarkan sebagai berikut :



**Gambar 3.1** Blog Diagram

### Uraian Dalam Block Diagram

Penjelasan langkah-langkah penghitungan secara garis besar pada block diagram, tahap ini akan menguraikan secara detail dari masing-masing prosesnya, berikut uraian dari paparan di atas :

#### 1. Data Raport Siswa

Penelitian ini dilakukan pada Sekolah Menengah Atas (SMA) Pangudi Luhur Don Bosko Semarang dengan memanfaatkan data nilai raport siswakesel X. Data nilai raport siswa yang digunakan sebagai pertimbangan diantaranya adalah Sistem Informasi, Basis Datayang nantinya dapat digunakan untuk mengetahui kompetensi siswa agar dapat dilakukan pengelompokan jurusan pada Sekolah Menengah Atas (SMA) Pangudi Luhur Don Bosko Semarang.

#### 2. Konversi Data

Dataset Raport Siswa yang berupa huruf akan menyusahkan pada saat tahap pemrosesan, maka dari itu data transkrip nilai harus dirubah kedalam angka terlebih dahulu.

#### 3. Inisialisasi

Pemberian nilai pada *pseudo-partition* (derajat keanggotaan setiap data pada setiap cluster) dalam jangkauan  $[0,1]$ , seperti persamaan 2.1.

#### 4. Hitung Centroid

Penghitungan centroid digunakan pada tiap-tiap cluster dengan rumus persamaan 2.3 dimana  $n$  adalah jumlah data,  $w$  adalah bobot pangkat,  $\mu_{ik}$  adalah nilai derajat keanggotaan data, dan  $X_{ij}$  adalah data yang dicluster.

#### 5. Hitung Fungsi Objektif

Selanjutnya dilakukan penghitungan fungsi objektif menggunakan rumus persamaan 2.4.

#### 6. Perubahan Matrik Partisi

Menghitung perubahan matrik partisi digunakan rumus persamaan 2.5 dimana  $V_{kj}$  adalah pusat *cluster* ke- $k$  untuk atribut ke- $j$ ,  $X_{ij}$  adalah data ke- $i$  atribut ke- $j$ , sedangkan  $\mu_{ik}$  adalah derajat keanggotaan untuk data sample ke- $i$  pada *cluster* ke- $k$ .

#### 7. Cek Kondisi

Tahapan ini adalah tahap dilakukannya pengecekan kondisi berhenti pada nilai

fungsi objektif dengan cara sebagai berikut :

- a.  $|P_t - P_{t-1}| < \zeta$  atau  $(t > \text{MaxIter})$  maka berhenti;
- b. Jika tidak, iterasi dinaikkan  $t=t+1$ , ulangi langkah ke-4.

Pengecekan kondisi dilakukan sesuai kebutuhan, karena semakin kecil selisih maka semakin tinggi akurasi dan cluster yang didapat dinyatakan sudah convergen.

#### 8. Konvergen

Tahap terakhir dari proses clustering yaitu mencari nilai terbesar dari derajat keanggotaan yang diperoleh dari iterasi terakhir. Derajat keanggotaan terbesar menunjukkan bahwa kecenderungan tertinggi lulus untuk masuk menjadi anggota *cluster* tersebut.