

BAB 3

METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini dilaksanakan pada Universitas Dian Nuswantoro Semarang untuk meng-*clustering* bidang kerja berdasarkan lulusan dengan memanfaatkan data mahasiswa dengan menggunakan algoritma *Fuzzy C-Means*.

3.2 Metode Analisis

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

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi organisasi

2. Data Cleaning

Pada umumnya, data yang diperoleh, baik dari database suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data yang tidak relevan itu juga lebih baik dibuang karena keberadaanya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performasi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. Data integration

Integrasi data dilakukan pada atribut-attribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, nomor pegawai, tempat lahir, agama dan lain sebagainya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

4. Task relevan data

Setelah semua sumber data digabung atau diintegrasikan menjadi satu keseluruhan database, maka tahap selanjutnya adalah melakukan task relevant data. Pada tahap ini melakukan relevansi atribut dari data yang relevant atau yang sesuai dengan target atau output yang akan dihasilkan.

5. Data transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.

6. Data mining

Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan.

7. Pattern evaluation

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah

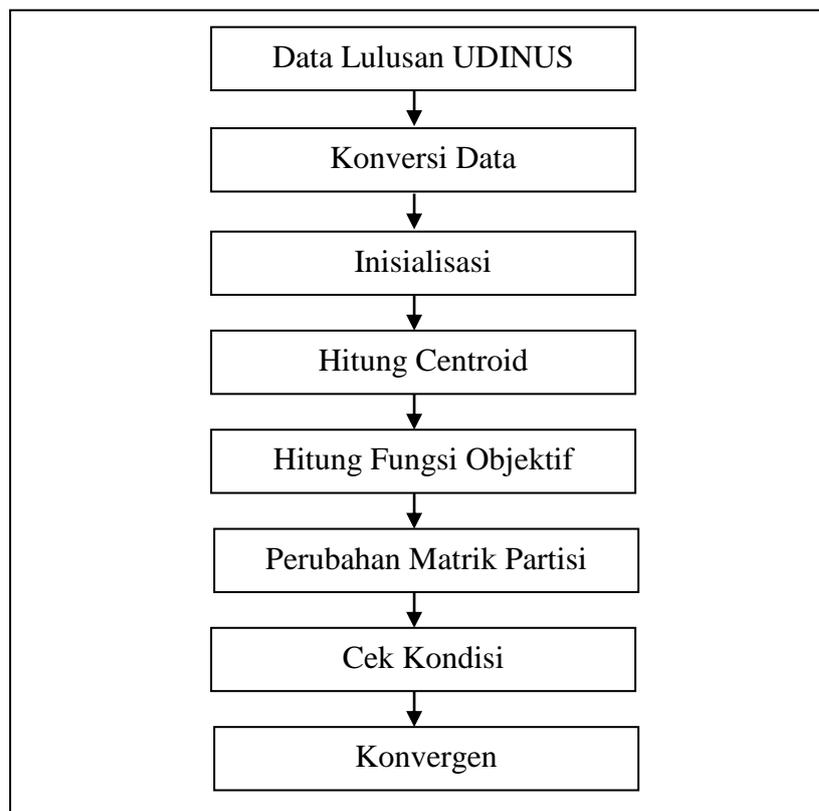
hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. Knowledge

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

3.3 Blok Diagram Proses Algoritma *Fuzzy C-Means*

Berdasarkan pada tahapan KDD berikutnya dimasukkan kedalam tahapan blok diagram, bagian ini akan membahas secara garis besar langkah-langkah penghitungan menggunakan metode algoritma *Fuzzy C-Means* yang dipaparkan pada gambar 3.1.



Gambar 3.1 Blok Diagram Penelitian Algoritma Fuzzy C-Means

3.4 Uraian Block Diagram

Setelah penjelasan langkah-langkah penghitungan secara garis besar pada block diagram, bagian ini akan menguraikan secara detail masing-masing prosesnya, berikut adalah uraian gambar 3.1 :

1. Data Lulusan UDINUS

Penelitian ini dilaksanakan pada Universitas Dian Nuswantoro Semarang dengan memanfaatkan data transkrip nilai lulusan mahasiswa TI (Teknik Informatika) strata satu (S-1) dan SI (Sistem Informasi) strata satu (S-1) yang ke 53, dan 54 pada tahun 2014. Data nilai mata kuliah yang digunakan sebagai pertimbangan diantaranya adalah Rekayasa Perangkat Lunak, rekayasa Perangkat Lunak Lanjut, Software Quality dan Testing, Interaksi Manusia dan Komputer, Dasar Pemrograman, Sistem Operasi, Jaringan Komputer, Jaringan Komputer Lanjut, Sistem Terdistribusi, Keamanan Sistem Komputer, Kriptografi, Pengantar Multimedia, Pemrograman Animasi dan Multimedia, Musik/Audio for Multimedia, Video Processing, Dasar Entrepreneurship, Entrepreneurship Lanjut, Start Up Bussiness, Bussiness Management, Sistem Informasi, Basis Data, Management Proyek, dan Konsep dan Aplikasi GIS yang nantinya dapat digunakan untuk mengetahui bidang pekerjaan yang sesuai yaitu bidang pekerjaan sebagai Programmer, IT Support, Multimedia, Wirausaha, dan Sistem Informasi sesuai dengan kompetensi yang ada di UDINUS.

2. Konversi Data

Data transkrip nilai mahasiswa yang berupa huruf akan menyusahkan pada saat tahap pemrosesan, maka dari itu data transkrip nilai harus dikonversi kedalam angka terlebih dahulu.

3. Inisialisasi

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

4. Hitung Centroid

Penghitungan centroid dilakukan pada tiap-tiap cluster dengan menggunakan rumus persamaan 2.3 dimana n adalah jumlah data, w adalah bobot pangkat, μ_{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 μ_{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 konvergen.

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 lulusan untuk masuk menjadi anggota *cluster* tersebut.