

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian ini dilakukan menggunakan tinjauan dari beberapa sumber referensi yang terkait berupa jurnal nasional, buku, dan website sebagai sumber landasan teori. Berikut adalah beberapa penelitian terkait yang sudah pernah dilakukan untuk topik penelitian.

Penelitian yang terkait dengan penelitian yang dilakukan oleh penulis yaitu Ambar Tri Hapsari pada tahun 2016. Sistem yang berjalan saat ini pada sekolah SMA PGRI 3 Jakarta masih sederhana dengan dilakukan pengelompokan secara manual. Sehingga untuk menyelesaikan masalah dalam penulisan jurusan dibutuhkan suatu sistem pendukung keputusan yang mampu melakukan perhitungan nilai yang mampu menentukan jurusan yang tepat. Sistem menggunakan algoritma *Fuzzy C-Means*. Algoritma *Fuzzy C-Means* dikatakan akurat jika hasil FCM dan peminatan yang dipilih hasilnya sama, dan rata-rata nilai hasil UN di atas 70, kondisi selain itu algoritma FCM dinyatakan tidak akurat. Diperoleh keakuratan peminatan IPA menggunakan algoritma FCM sebesar 92%, Diperoleh keakuratan peminatan IPS menggunakan algoritma FCM sebesar 63% [1].

Penelitian yang dilakukan oleh Fajar Akbar pada tahun 2015 dengan. Penulis melakukan penelitian untuk mengolah data hasil raport siswa yang digunakan untuk pemilihan jurusan pada SMAN 84 Jakarta karena penentuan jurusan harus sesuai dengan hasil nilai, kemampuan dan kuota kelas jurusan yang tersedia. Maka sehubungan dengan permasalahan yang dijelaskan penulis menggunakan *Fuzzy C-Means* untuk menentukan jurusan dan untuk mendistribusikan siswa sesuai dengan kuota kelas yang ada dengan menggunakan algoritma genetika.

Maka dengan menggunakan metode *Fuzzy C-Means* memperoleh tingkat akurasi sebesar 72,26 % dalam penerapan penentuan jurusan pada SMA 84 Jakarta [2].

Cary Lineker Simbolon melakukan penelitian pada tahun 2013 yang memiliki tujuan pengelompokan lulusan mahasiswa Matematika FMIPA UNTAN Pontianak yang dibatasi dari tanggal 24 Februari 2007 sampai 31 Maret 2012 dan mencari pusat *cluster* dengan *Fuzzy C-Means clustering* berdasarkan nilai IPK dan lama studi. Hasil ini dapat dijadikan sebagai bahan pertimbangan para staff pengajar pada Untan Pontianak jurusan Matematika pada Fakultas MIPA untuk meningkatkan IPK mahasiswa agar dapat menyelesaikan masa studinya dengan lebih cepat [4].

Erfan Agil Putranto melakukan penelitian sistem pendukung keputusan penjurusan siswa kelas X maka menghasilkan pengujian algoritma *Fuzzy C-Means* (FCM) dalam penentuan jurusan di Sekolah Menengah Atas pada 138 sampel data siswa yang diuji dalam penelitian ini menunjukkan bahwa tingkat kemiripan perhitungan FCM dibandingkan secara manual dengan FCM dengan penggunaan daya dukung minat masing-masing siswa didapatkan hasil rata-rata sebesar 83.33 %, sedangkan jika dibandingkan dengan perhitungan FCM tanpa penggunaan daya dukung minat dibandingkan juga secara manual, didapatkan hasil rata-rata sebesar 78.26 %. Pengujian selanjutnya adalah membandingkan perhitungan algoritma FCM dengan penggunaan daya dukung minat dan FCM tanpa penggunaan daya dukung minat dibandingkan dengan hasil nilai peminatan yang didapatkan oleh masing-masing siswa di kelas XI semester 1 [5].

Aidina Ristiawan melakukan penelitian proses pengclusteran keahlian yang dilakukan pada penelitian ini terbukti dapat dilakukan dengan baik. Hal ini bisa ditunjukkan pada saat bahwa nilai fungsi objektif sudah berada di bawah ambang batas yang ditetapkan 0,05 yaitu : -5,6053. Dan untuk perhitungan index XBI pada salah satu keahlian yaitu "ANIMASI" mencapai 3,04901E-14 pada saat cluster berjumlah 5. Sehingga pada penelitian ini susunan pengclusteran paling optimal adalah dengan jumlah cluster 5. Hasil dari penelitian ini dapat digunakan untuk melihat posisi atau keadaan kinerja akademik mahasiswa, dan dapat pula

dikembangkan sebagai dasar pertimbangan dalam memilih profesi dalam penelitian berikutnya. Adapun saran pada penelitian ini adalah perlu dilakukannya pengklasifikasian matakuliah kedalam kelompok – kelompok keahlian. Sehingga diharapkan hasil pengclusteran keahlian dapat lebih aktual [6].

Menurut penelitian yang telah dilakukan, permasalahan yang terjadi didalam sekolah yaitu pengelompokan jurusan pada sekolah yang belum sistematis maka, penulis membuat model penentuan peminatan siswa untuk mengolah data nilai siswa menggunakan metode clustering dan algoritma fuzzy c-means sehingga menghasilkan informasi yang berguna bagi para guru untuk menentukan pengelompokan peminatan jurusan.

Tabel 1.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Ambar Tri Hapsari	Masih menggunakan sistem yang sederhana dalam pengelompokan data.	Menggunakan metode <i>Fuzzy C-Means</i>	Setelah menggunakan metode FCM diperoleh hasil diperoleh keakuratan peminatan IPA menggunakan algoritma FCM sebesar 92%. Diperoleh keakuratan peminatan IPS menggunakan algoritma FCM

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
				sebesar 63%.
2.	Fajar Akbar	Pengolahan data raport siswa untuk penjurusan yang masih dilakukan secara sederhana.	Algoritma <i>Fuzzy C-Means</i> (FCM)	Maka hasil dari penerapan metode <i>Fuzzy C-Means</i> diperoleh tingkat akurasi sebesar 72,26 % dalam penerapan penentuan jurusan pada SMA 84 Jakarta
3.	Cary Lineker Simbolon, et al	Untuk mengelompokkan lulusan data mahasiswa FMIPA UNTAN yang masih manual	Algoritma <i>Fuzzy C-Means</i>	Dari hasil keempat <i>cluster</i> tersebut diketahui bahwa pada <i>cluster4</i> memiliki anggota lulusan yang paling banyak dengan 33 lulusan. <i>Cluster 4</i> terdiri dari lulusan dengan kisaran lama studi 5,91 tahun

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
4.	Erfan Agil Putranto, etc	untuk menentukan pengelompokkan penjurusan siswa kelas X SMA Negeri 2	Algoritma <i>Fuzzy C-Means</i>	hasil pengujian algoritma <i>Fuzzy C-Means</i> (FCM) dalam penentuan jurusan di Sekolah Menengah Atas dalam penelitian ini menunjukkan bahwa tingkat kemiripan perhitungan FCM dibandingkan secara manual dengan FCM dengan penggunaan daya dukung minat masing-masing siswa didapatkan hasil rata-rata sebesar 83.33 %.
5.	Aidina Ristyawan, etc	Untuk melakukan pengelompokkan kinerja Akademik Mahasiswa pada	Algoritma <i>Fuzzy C-Means</i>	Dari penelitian yang dilakukan akan menghasilkan

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
		STMIK AMIKOM		data pengelompokkan jurusan yang secara sistematis dengan menggunakan penerapan metode <i>Fuzzy C-Means</i> .

2.2 Pendekatan Penjurusan

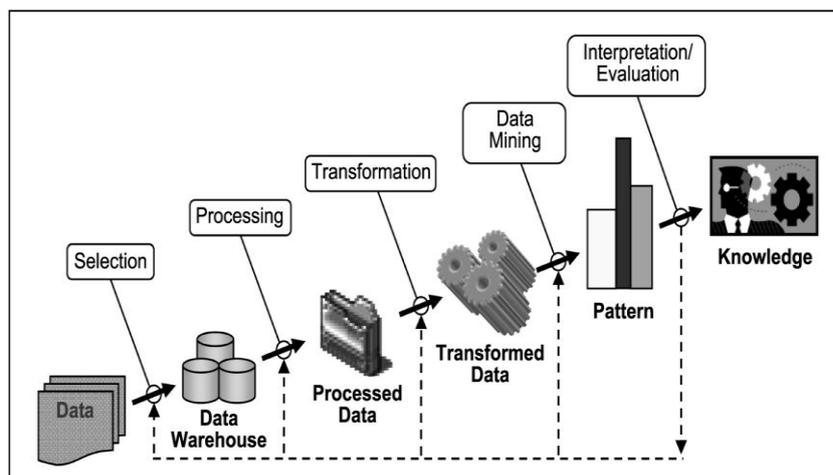
Pendekatan penjurusan menurut pasal 12 ayat 2 undang-undang No. 20 Tahun 2003 yang mengatur tentang kaidah Pendidikan yang mengemukakan bahwa setiap siswa berhak mendapat pelayanan pendidikan yang sesuai dengan bakat dan kemampuannya. Setiap siswa terutama pada sekolah menengah atas ataupun sederajat mempunyai hak dalam menentukan pilihan pembelajaran atau jurusan yang sesuai dengan kemampuan yang dimiliki setiap siswa untuk mengembangkan potensi masing-masing individu.

Proses pemilihan peminatan penjurusan padasaat kenaikan kelas X ke kelas XI proses penjurusan yang akan dilakukan berdasarkan dari nilai raport setiap mata pelajaran sebagai pertimbangan dan juga di lihat dari minat siswa yang bersangkutan [2].

2.3 Data Mining

Data mining adalah sebuah serangkaian proses pemecahan data dengan cara mengelompokkan untuk menghasilkan informasi yang ada didalam database. Dan data mining juga dapat diartikan sebagai penemuan untuk menemukan hubungan yang bermanfaat seperti pola dengan melihat dan memeriksa kumpulan data yang tersimpan didalam penyimpanan dengan menggunakan teknik pengenalan pola seperti tabel statistik dan matematika didalam teknik identifikasi yang berguna dari berbagai database dalam jumlah besar dari beberapa bagian data mining di olah secara otomatis dari yang telah dibuat sebelumnya. Didalam Data Mining data yang diolah merupakan data yang berjumlah besar yang bermaksud untuk dapat di buat pola supaya dapat memberikan hasil atau petunjuk yang berguna.

Pada jurnal Data mining dipopulerkan dengan sebutan (KDD) Knowledge Discovery in Database. kedua istilah itu saling berkaitan untuk menemukan informasi yang tersembunyi. Keduanya memiliki konsep yang tidak sama tetapi saling berkaitan. Data mining termasuk didalam salah satu dari tahapan KDD, gambaran dari KDD sebagai berikut [3] :



Gambar 2.1 Tahapan KDD (*Knowledge Discovery in Database*)

1. Data Selection

Merupakan tahapan pada sistem informasi di mana data yang relevan untuk diseleksi dan akan analisis untuk dikembangkan dan dikembalikan ke dalam penerapan data mining.

2. *Processed Data / Cleaning*

Proses yang dilakukan untuk menghilangkan adanya penggandaan data , mengoreksi adanya kesalahan didalam pendataan, dan juga untuk membenahan kesalahan data.

3. *Transformed Data*

Dalam tahap ini dimana data berupa proses coding yang dirubah atau disatukan menjadi bentuk yang tepat untuk digunakan dalam proses data mining.

4. *Data Mining*

Tahap dasar metode yang diperlukan untuk menghasilkan pola data untuk diolah menjadi informasi dengan menggunakan teknik tertentu.

5. *Pattern evolution*

Tahap ini digunakan untuk mengenali pola yang benar-benar mewakili pengetahuan berdasarkan dari beberapa tindakan yang menarik.

6. *Knowledge presentation*

Proses yang digambarkan dengan teknik visualisasi dan pengetahuan digunakan untuk memeriksa apakah informasi yang dihasilkan pada suatu pola sesuai dengan fakta atau tidak.

Didalam Datamining itu sendiri dikelompokkan menjadi beberapa berdasarkan tugas yang di lakukan :

1. *Deskripsi*

Metode yang dilakukan untuk mengetahui cara dalam mempresentasikan pola atau rancangan yang ada didalam data

2. *Estimasi*

Proses yang dilakukan untuk memperkirakan nilai yang berdasarkan dari nilai dan sampel-sampel.

3. *Prediksi*

Metode yang diterapkan dalam penelitian saat akan melakukan prediksi sebuah nilai yang belum diketahui sebelumnya

4. Klasifikasi

Proses untuk mendapatkan model yang dapat digunakan untuk mempresentasikan atau membedakan *class* data yang ditujukan untuk memperkirakan *class* dari sebuah objek dengan label yang belum di ketahui sebelumnya.

5. Pengklasteran

Metode yang digunakan untuk melakukan pengelompokkan data dan hasil-hasil penelitian ke dalam *class* yang memiliki kemiripan.

6. Asosiasi

Metode yang digunakan untuk menemukan atribut yang muncul dalam satu waktu.

Didalam Data Mining itu sendiri dikelompokkan menurut jenisnya menjadi beberapa diantaranya:

1. Klasifikasi

Klasifikasi adalah suatu teknik yang di lakukan untuk menentukan record datas yang baru kedalam salah satu kategori atau kelas yang telah di definisikan sebelumnya. Atau bisa disebut juga dengan *supervised learning*.

2. Regresi

Regresi yaitu suatu teknik yang di gunakan untuk melakukan prediksi suatu variabel kontinyu yang diperoleh nilai berdasarkan variabel lain dengan cara mengkonsumsi dari model yang bergantung pada linier atau nonlinier.

3. Klasterisasi (*Clustering*)

Cluster merupakan suatu metode yang di gunakan untuk menemukan sekumpulan objek sampai objek-objek terkumpul didalam satu kelompok yang sama dan dapat dibedakan antara yang mempunyai hubungan satu dengan yang lain [7].

Beberapa tipe yang penting dari *clustering* adalah :

- a. **Partitional**
Adalah pembagian suatu objek data dalam sub bagian didalamhimpunanatau cluster yang tidak overlap sehingga setiap objekdata berada di dalamsatu bab himpunan.
- b. **Hierarchical**
Merupakan sebuah himpunan cluster yang bersarang dan diatur sebagai suatu pohon hierarki yang setiap simpul didalam pohon merupakan gabungan dari anak sub cluster dan simpul akar yang berisi semua objek.
- c. **Fuzzy**
Dalam fuzzy clustering adalah sebuah titik yang termasuk dalam setiap cluster dengan suatu nilai bobot antara 0 dan 1.
- d. **Complate**
Complate merupakan penempatan setiap objek ke dalam sebuah cluster atau pengelompokkan.
- e. **Asosiasi**
Merupakan suatu metode yang digunakan dalam mendeteksi kumpulan atributyang sering muncul secara bersamaan didalam frekuensi yang sering sehingga membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.

2.4 Fuzzy C-Means (FCM)

Fuzzy C-Means yaitu sebuah metode yang digunakan dalam menentukan sebuah *cluster* atau pemilihan yang optimal dalam suatu bidang ruang vector berdasarkan bentuk normal *Euclidian* untuk jarak antar vector. *Fuzzy C-Means* merupakan suatu algoritma yang berguna untuk pemodelan *fuzzy* terutama didalam mengidentifikasi sebuah aturan pada *fuzzy*.

Dari beberapa algoritma *clustering* data salah satunya adalah *fuzzy C-Means* (FCM) merupakan teknik *pengcluster* data dimana setiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan titik tiap data.

Konsep dasar dalam algoritma *Fuzzy C-Means* (FCM) pertama kali yaitu menentukan pusat *cluster* yang akan digunakan untuk menandai lokasi rata-rata untuk setiap *cluster*. Pada kondisi awal pusat *cluster* masih belum akurat setiap titik data memiliki derajat keanggotaan untuk setiap *cluster* [8]. Kelebihan dari *Fuzzy C-Means* adalah salah satu metode yang memiliki tingkat akurasi yang tinggi dan waktu yang komputasi lebih cepat [3].

Didalam Algoritma *Fuzzy C-Means* (FCM) berdasarkan pada minimasi fungsi objektif maka dapat di formulasikan ke dalam persamaan seperti berikut :

$$J = \sum_{i=1}^n \sum_{j=1}^c (U_{ij})^m \|z_i - c_j\|^2$$

Keterangan :

U_{ij} : level keanggotaan dari Z_i dalam *cluster* j

Z_i : merupakan nilai data ke i dari d -dimensi data

C_j : merupakan nilai ke j dari d -dimensi *cluster center*

m : sembarang bilangan real lebih besar dari 1

Langkah-langkah pada penyusunan Algoritma *Fuzzy C-Means* (FCM) berdasarkan pada meminimalisasi fungsi objektif yang diformulasi dalam persamaan:

1. Tentukan himpunan data Z , tentukan jumlah *cluster* yang di harapkan $1 < c < N$, nilai pembobot $m > 1$, toleransi penghentian $\epsilon > 0$.

(2.1)

2. Inisialisasi matriks partisi secara acak, $U(0) \in M_{fc}$

Ulangi untuk $l = 3,4,5$ (2.2)

3. Hitung pusat *cluster* (*means*)

$$v_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, 1 \leq i \leq c \quad (2.3)$$

4. Menghitung jarak

$$D_{ikA}^2 = (z_k - v_i^{(l)})^T A (z_k - v_i^{(l)}), 1 \leq i \leq c, 1 \leq k \leq N \quad (2.4)$$

5. Perbaharui matriks partisi

Untuk $1 \leq k \leq N$

Jika $D_{ikA} > 0$ untuk semua $i=1,2,\dots,c$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA}/D_{jkA})^{2/(m-1)^3}}$$

Atau dengan arti lain :

$$\mu_{ik}^{(l)} = 0 \text{ jika } D_{ikA} > 0 \text{ dan } \mu_{ik}^{(l)} \in [0,1] \text{ dengan } \sum_{i=1}^c \mu_{ik}^{(l)} = 1$$

$$\text{Ulangi sampai } \|U^{(l)} - U^{(l-1)}\| \quad [8] \quad (2.5)$$

2.5 Matlab

Matlab adalah salah satu bahasa pemrograman level tinggi yang dikhususkan untuk komputasi teknis. Bahasa ini mengintegrasikan kemampuan komputasi, visualisasi dan pemrograman dalam sebuah lingkungan yang tunggal dan mudah digunakan. Matlab merupakan program yang ditujukan untuk orang-orang yang tidak ingin disibukkan dengan rumitnya sintak dan alur logika pemrograman, sementara pada saat yang sama membutuhkan hasil komputasi dan visualisasi yang maksimal untuk mendukung pekerjaan [9].

2.6 Java

Java tidak lain merupakan salah satu bahasa pemrograman yang dapat dijalankan di semua komputer dan tidak membutuhkan spesifikasi tinggi. Bahkan saat ini sudah merambah ke telepon genggam. Karena fungsionalitasnya yang memungkinkan aplikasi java mampu berjalan di beberapa platform sistem operasi yang berbeda, java dikenal pula dengan slogannya, "*Tulis sekali, jalankan di mana pun*". Saat ini java merupakan bahasa pemrograman yang paling populer digunakan, dan secara luas dimanfaatkan dalam pengembangan berbagai jenis perangkat lunak aplikasi ataupun aplikasi. Bahasa ini awalnya dibuat oleh James

Gosling saat masih bergabung di Sun Microsystems yang merupakan bagian dari Oracle dan dirilis tahun 1995. Java banyak mengadopsi sintaksis yang terdapat pada C dan C++ namun dengan sintaksis model objek yang lebih sederhana. Aplikasi-aplikasi berbasis java umumnya dikompilasi ke dalam *bytecode* dan dapat dijalankan pada berbagai Mesin Virtual Java atau sering di sebut dengan JVM [10].

2.7 MySQL

Perangkat lunak sistem manajemen basis data atau *database management system* atau lebih sering di kenal dengan nama DBMS yang *multithread, multi-user*, dengan sekitar 6 juta instalasi di seluruh dunia. Membuat MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi GNU General Public License (GPL), tetapi mereka juga menjual dibawah lisensi komersial untuk kasus-kasus di mana penggunaannya tidak cocok dengan penggunaan GPL. Tidak sama dengan proyek-proyek seperti Apache, di mana perangkat lunak dikembangkan oleh komunitas umum, dan hak cipta untuk kode sumber dimiliki oleh penulisnya masing-masing, MySQL dimiliki dan disponsori oleh sebuah perusahaan komersial Swedia [11].