

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Beberapa penelitian terdahulu telah banyak yang menerapkan data mining, yang bertujuan dalam menyelesaikan beberapa permasalahan seputar dunia pendidikan. Khususnya dalam melakukan pengelompokan terhadap mahasiswa, sehingga didapatkan informasi dan pola menarik yang dapat berguna dalam pengambilan keputusan. Berikut ini merupakan beberapa penelitian terdahulu yang memiliki permasalahan, tujuan, ataupun metode yang hampir sama dan terkait dengan penelitian ini.

2.1.1 Klasifikasi Data Mahasiswa Menggunakan Metode *K-Means* Untuk Menunjang Pemilihan Strategi Pemasaran

Penelitian yang dilakukan oleh [6] bertujuan untuk analisis *cluster* dengan *data mining* terhadap data mahasiswa sehingga hasilnya dapat digunakan untuk menunjang strategi pemasaran. Penelitian ini menggunakan algoritma *K-Means*, untuk dapat mengelompokkan data mahasiswa berdasarkan dari nilai rerata UN dan IPK. Menggunakan data mahasiswa angkatan 2014/2015 yang telah melakukan registrasi. Atribut data yang digunakan adalah nama mahasiswa, jurusan SLTA, program studi, kota asal mahasiswa, nilai UAN, dan IPK. Langkah-langkah dalam penelitian ini yaitu; menghimpun data dari bagian admisi dan akademik, menentukan dan menetapkan *cluster* dan titik pusatnya, melakukan pengolahan data yang berdasarkan dari nilai rata-rata UN dan IPK, hasil yang dinyatakan valid lalu dianalisis lebih lanjut dengan mencocokkan berdasarkan asal kota sekolah. Hasil penelitian menampilkan profil mahasiswa, keterkaitan nilai UN terhadap IPK, serta asal sekolahnya. Sehingga hasil penelitian dapat membantu pihak admisi perguruan tinggi dalam menyusun strategi promosi program studi yang menjadi target. Persamaan penelitian [6] dengan penelitian

sekarang yaitu menggunakan algoritma *K-Means clustering* dalam mengelompokkan mahasiswa. Sedangkan perbedaannya adalah dari tujuan penelitian dan atribut data yang digunakan dalam mengelompokkan mahasiswa berbeda.

2.1.2 Model *Data Mining* Dalam Pengklasifikasian Ketertarikan Belajar Mahasiswa Menggunakan Metode *Clustering*

Tujuan dari penelitian [4] untuk membuat model data mining dalam mengelompokkan mahasiswa berdasarkan ketertarikan mahasiswa terhadap matakuliah yang disenanginya. Sehingga dapat menghasilkan model data mining yang dapat dimanfaatkan dalam menentukan konsentrasi minat belajar pada mahasiswa program studi Sistem Informasi Universitas Binadarma Palembang. Pada penelitian [4] menggunakan algoritma *K-Means*. Menggunakan SDLC (*System Development Life Cycle*) sebagai metode pengembangan sistem. Menggunakan tools SSAS (*SQL Server 2008 Analysis Service*) dalam melakukan proses *mining*. Data yang digunakan yaitu data mahasiswa Sistem Informasi angkatan 2011-2013 Universitas Bina Darma. Hasil penelitian didapatkan 3 *cluster* berdasarkan konsentrasi matakuliah pilihan. *Cluster* 1 didapatkan 400 mahasiswa minat matakuliah konsentrasi A, *cluster* 2 didapatkan 186 mahasiswa minat matakuliah konsentrasi B, dan *cluster* 3 didapatkan 188 mahasiswa minat terhadap matakuliah konsentrasi C. Perbedaan antara penelitian terdahulu dengan yang sekarang ialah pada objek penelitian dan atribut yang akan diolah. Objek penelitian terdahulu dilakukan pada mahasiswa Sistem Informasi Universitas Bina Darma, sedangkan penelitian sekarang dilakukan terhadap mahasiswa Teknik Informatika Universitas Dian Nuswantoro. Penelitian terdahulu tidak menggunakan atribut rerata nilai dari matakuliah penunjang, pada penelitian yang sekarang melakukan pengelompokan berdasarkan atribut indeks prestasi (IP) semester ke-1 sampai ke-4 dan rerata nilai MK penunjang dari masing-masing peminatan. Sehingga dapat mengelompokkan mahasiswa berdasarkan kemampuan akademisnya, untuk dapat diketahui kesesuaiannya dengan masing-masing jalur peminatan berdasarkan tingkat kemampuan akademisnya.

2.1.3 Metode *Clustering* Dengan Algoritma *Fuzzy C-Means* Untuk Rekomendasi Pemilihan Bidang Keahlian Pada Program Studi Teknik Informatika

Pada penelitian yang dilakukan oleh [2] permasalahan yang terjadi ialah adanya tiga pilihan terhadap bidang keahlian yaitu keahlian *Networking*, *Informatic*, dan *Software Develop and Engineer* pada program studi Teknik Informatika. Sehingga mahasiswa kesulitan dalam menentukan keahlian yang akan diambil. Oleh karena itu penelitian tersebut bertujuan untuk menerapkan metode *clustering* dengan algoritma *Fuzzy C-Means* dalam kasus pengelompokan mahasiswa yang didasarkan dari transkrip nilai mata kuliah prasyarat sebagai rekomendasi pemilihan bidang keahlian. Data yang digunakan dalam penelitian ini ialah data mahasiswa dan nilai mata kuliah prasyarat. Perbedaan antara penelitian terdahulu [2] dengan penelitian sekarang yaitu dari metode yang digunakan. Metode pada penelitian terdahulu menggunakan algoritma *Fuzzy C-Means* sedangkan penelitian sekarang menggunakan algoritma *K-Means*.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Totok Suprawoto, 2016	Diperlukan adanya strategi pemasaran yang tepat, dengan menggunakan <i>data mining</i> sehingga dapat membantu dalam menyusun suatu strategi promosi pada program studi	Menggunakan Algoritma <i>K-Means</i>	<ul style="list-style-type: none"> • Algoritma <i>K-Means</i> dapat digunakan untuk pengelompokan mahasiswa berdasarkan nilai UN dan IPK • Dari hasil pengelompokan

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
				, lalu dianalisa lebih lanjut agar dapat diketahui relasi antara prestasi dengan sebaran asal kota sekolahnya
2.	Marlindawati & Andri, 2015	Perlu adanya sebuah keputusan sehingga dapat dimanfaatkan untuk mengarahkan konsentrasi mahasiswa, <i>data mining</i> dapat digunakan untuk mengetahui informasi dari ketertarikan belajar mahasiswa pada Universitas Binadarma Palembang	Menggunakan <i>tools</i> SSAS dalam proses <i>mining</i> . Menggunakan SDLC (<i>system development life cycle</i>) sebagai metode pengembangan sistem. Menggunakan Algoritma <i>K-Means</i> .	<ul style="list-style-type: none"> • Penelitian menghasilkan sebuah model <i>data mining</i> dalam pengelompokan minat matakuliah konsentrasi • Model <i>data mining</i> digunakan untuk melakukan prediksi minat matakuliah konsentrasi
3.	Muhammad Faisal Mirza,	Adanya tiga bidang keahlian	Menggunakan Algoritma <i>Fuzzy C-</i>	<ul style="list-style-type: none"> • Algoritma Fuzzy C-Means

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
	2013	pada program studi Teknik Informatika Universitas Dian Nuswantoro, sehingga mahasiswa mengalami kesulitan dalam menentukan keahlian yang diambil	<i>Means</i> , diaplikasikan dengan matlab 7.10	<p>dapat diterapkan dalam pengelompokan mahasiswa untuk rekomendasi penjurusan</p> <ul style="list-style-type: none"> • Hasil pengelompokan dapat diketahui rerata nilai mahasiswa pada setiap jurusan

2.2 Landasan Teori

2.2.1 Minat Studi

Minat studi mahasiswa bertujuan sebagai wadah dalam edukasi untuk mengarahkan mahasiswa agar dapat lebih fokus dan juga terarah dengan pembelajaran yang sesuai peminatan dari masing-masing mahasiswa [13]. Pada semester ke-lima mahasiswa program studi Teknik Informatika-S1 Universitas Dian Nuswantoro diharuskan mengambil salah satu peminatan yang ditawarkan. Berdasarkan jenisnya, peminatan tersebut dapat dikelompokkan menjadi dua yaitu peminatan Sistem Cerdas dan peminatan Rekayasa Perangkat Lunak dan Data.

2.2.1.1 Penentuan Minat Studi

Terdapat dua cara dalam memilih dan menentukan minat studi mahasiswa [13]:

1. Mahasiswa tersebut menyadari minat dan kemampuannya berdasarkan dari perolehan nilai yang didapatkan pada beberapa matakuliah tertentu, yang dijadikan matakuliah penunjang dan berkaitan dengan peminatan yang ditawarkan.
2. Mahasiswa dapat melakukan konsultasi dan berdiskusi terhadap dosen wali.

2.2.1.2 Jalur Peminatan

Terdapat beberapa matakuliah khusus dan matakuliah umum pada masing-masing peminatan tersebut. Matakuliah khusus merupakan matakuliah yang dapat diambil, jika mahasiswa tersebut merupakan bagian dari salah satu jalur peminatan. Berikut ini pada Tabel 2.2 merupakan daftar matakuliah pada masing-masing peminatan [1].

Tabel 2.2 Daftar Matakuliah pada Masing-Masing Peminatan

Peminatan Sistem Cerdas			Peminatan Rekayasa Perangkat Lunak dan Data		
Semester	Kode MK	Nama MK	Semester	Kode MK	Nama MK
V	54501	Jaringan Komputer	V	54501	Jaringan Komputer
	54507	Sistem Operasi		54507	Sistem Operasi
	54506	Sistem Informasi		54506	Sistem Informasi
	54508	Strategi Algoritma		54504	RPL Lanjut
	54503	Komputer Grafik		54505	Sistem Basis Data
	54521	Pengolahan Citra		54816	Pemrograman

Peminatan Sistem Cerdas			Peminatan Rekayasa Perangkat Lunak dan Data		
Semester	Kode MK	Nama MK	Semester	Kode MK	Nama MK
		Digital			Internet
	54505	Agama		54305	Agama
VI	54603	Interaksi Manusia dan Komputer	VI	54603	Interaksi Manusia dan Komputer
	54605	Kecerdasan Buatan		54605	Kecerdasan Buatan
	54609	Sistem Terdistribusi		54609	Sistem Terdistribusi
	54607	Metodologi Penelitian		54607	Metodologi Penelitian
	54606	Kriptografi		54606	Kriptografi
	54602	Pendidikan Kewarganegaraan		54602	Pendidikan Kewarganegaraan
	54806	Data Mining		54806	Data Mining
VII	54702	Kerja Praktek	VII	54702	Kerja Praktek
	54704	Representasi Pengetahuan dan Penalaran		54703	Manajemen Projek
	54701	Bimbingan Karier		54701	Bimbingan Karier
	54705	Tugas Akhir I		54705	Tugas Akhir I
		Pilihan I			Pilihan I

Peminatan Sistem Cerdas			Peminatan Rekayasa Perangkat Lunak dan Data			
Semester	Kode MK	Nama MK	Semester	Kode MK	Nama MK	
		Pilihan II			Pilihan II	
VIII	54801	Tugas Akhir II	VIII	54801	Tugas Akhir II	
		Pilihan III				Pilihan III
		Pilihan IV				Pilihan IV

2.2.1.3 Matakuliah Peminatan

Ada beberapa perbedaan matakuliah pada semester 5-8 yang ditawarkan dalam masing-masing peminatan diantaranya ialah:

1. Peminatan Sistem Cerdas
 - a. Strategi Algoritma
 - b. Komputer Grafik
 - c. Pengolahan Citra Digital
 - d. Representasi Pengetahuan dan Penalaran
2. Peminatan Rekayasa Perangkat Lunak dan Data
 - a. Rekayasa Perangkat Lunak Lanjut
 - b. Sistem Basis Data
 - c. Pemrograman Internet
 - d. Manajemen Proyek

2.2.1.4 Matakuliah Penunjang

Berdasarkan dari buku panduan minat studi mahasiswa program studi Teknik Informatika-S1 Universitas Dian Nuswantoro, berikut ini merupakan matakuliah

yang ditawarkan pada semester 1-4 yang dapat dijadikan matakuliah penunjang sebagai dasar rekomendasi dalam peminatan [13]:

1. Peminatan Sistem Cerdas
 - a. Kalkulus I
 - b. Kalkulus II
 - c. Probabilitas & Statistik
 - d. Otomata & Teori Bahasa
2. Peminatan Rekayasa Perangkat Lunak dan Data
 - a. Dasar Pemrograman
 - b. Algoritma dan Pemrograman
 - c. Rekayasa Perangkat Lunak
 - d. Basis Data

2.2.2 Data Mining

Data Mining merupakan proses eksplorasi secara otomatis maupun semiotomatis yang menganalisa data berjumlah besar, dan memiliki tujuan menemukan pola yang bermakna [3]. Sedangkan menurut Turban dalam buku “Algoritma *Data Mining*” [12], *Data mining* ialah proses dengan menggunakan teknik kecerdasan buatan, *machine learning*, matematika dan statistik dalam mengidentifikasi dan mengekstraksi informasi dan pengetahuan yang bermanfaat dalam berbagai *database* besar.

Data mining berisi pola yang diinginkan ataupun pencarian trend di dalam *database* yang besar yang dapat membantu pengambilan keputusan di waktu yang akan datang [5]. Pola yang diekstrak menggunakan *data mining* dapat membantu organisasi dalam mengambil keputusan terbaik [11].

2.2.2.1 Proses Data Mining

Data mining merupakan bagian dalam proses KDD (*Knowledge Discovery in Database*). KDD merupakan proses keseluruhan dalam melakukan konversi data mentah untuk dijadikan pengetahuan yang berguna. Berikut ini merupakan tahapan dalam KDD (Han & Kamber, 2006):

1. *Data cleaning*

Tahap ini dilakukan pemilihan data, dengan memisahkan data yang tidak konsisten dan tidak relevan

2. *Data integration*

Tahap ini bertujuan untuk menggabungkan berbagai sumber data menjadi satu sumber.

3. *Data selection*

Tahap ini dilakukan pemilihan data yang relevan dengan analisa yang akan dilakukan.

4. *Data transformation*

Tahap ini untuk mengubah format data menjadi format data yang sesuai untuk diproses *data mining*

5. *Data mining*

Pada tahap ini dilakukan dengan menerapkan metode tertentu dalam melakukan proses *mining* agar didapatkan informasi yang tersembunyi dari data yang ada.

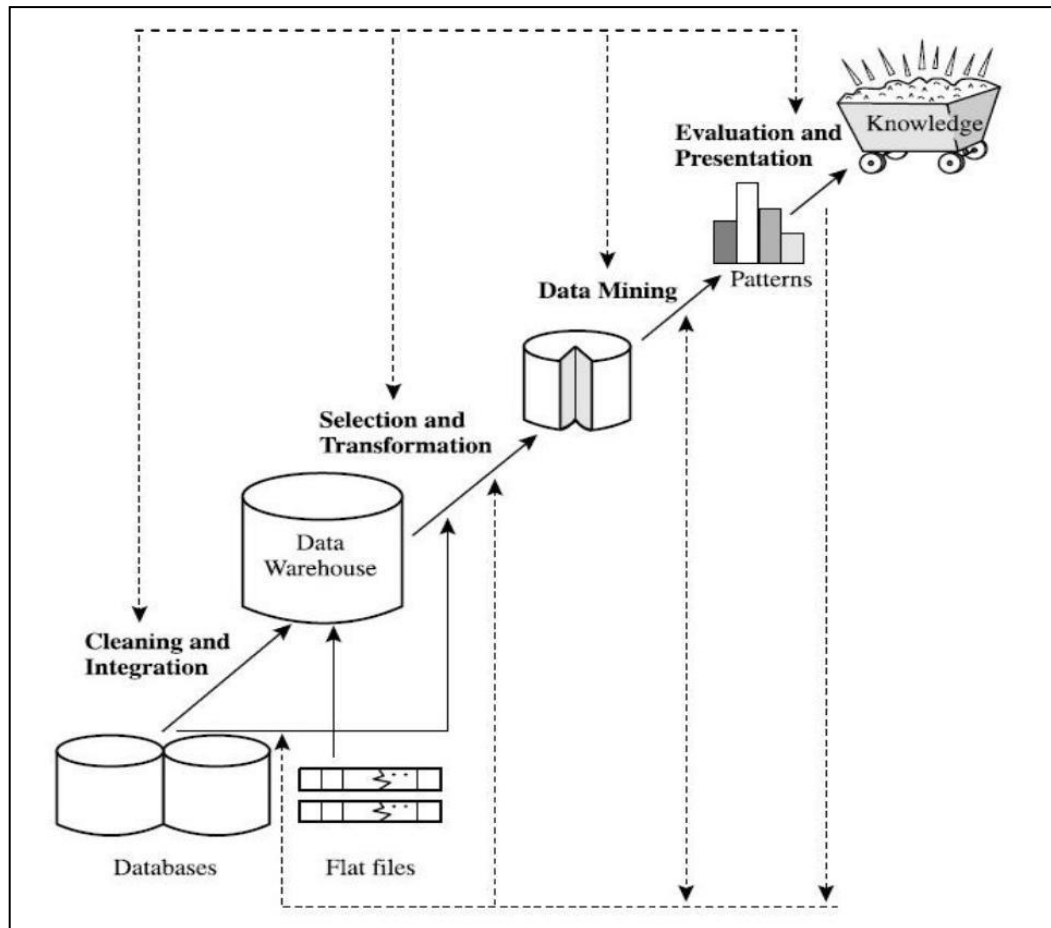
6. *Pattern Evaluation*

Tahap ini dilakukan identifikasi pada pola-pola yang menarik yang telah didapatkan dari *data mining*, untuk kemudian direpresentasikan.

7. *Knowledge Presentation*

Tahap ini dilakukan visualisasi dan penyajian pada teknik yang digunakan.

Tahap ke-1 sampai 4 merupakan beberapa tahapan dari *preprocessing* data, yang bertujuan dalam menyusun data untuk selanjutnya dilakukan proses *data mining*. Berikut ini pada Gambar 2.1 menjelaskan mengenai tahapan dalam KDD.



Gambar 2.1 Tahapan KDD

(Sumber: Han & Kamber, 2006)

2.2.2.2 Tugas Utama Data Mining

Pada umumnya terdapat empat tugas utama dalam data mining, yaitu [3]:

1. Model Prediksi (*Prediction Modelling*)

Tugas tersebut berkaitan dalam membangun sebuah model sehingga dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, model tersebut digunakan dalam memberikan nilai target pada himpunan baru yang didapat. Klasifikasi dan regresi termasuk dalam model prediksi, dimana

klasifikasi digunakan untuk variabel diskrit sedangkan regresi digunakan pada variabel kontinyu.

2. Analisis Asosiasi (*Association Analysis*)

Merupakan penemuan pola yang menggambarkan kondisi dari nilai atribut yang secara bersamaan muncul dalam sebuah himpunan data. Pola tersebut merepresentasikan bentuk aturan asosiasi.

3. Analisis *Cluster* (*Cluster Analysis*)

Analisis *cluster* mengelompokkan data berdasarkan dari kesamaan karakteristik ke dalam sejumlah kelompok. Data yang memiliki tingkat kesamaan yang tinggi akan bergabung dalam kelompok, dan jika diluar dari batas kesamaan maka akan terpisah dalam kelompok yang berbeda.

4. *Anomaly Detection / Outlier Mining*

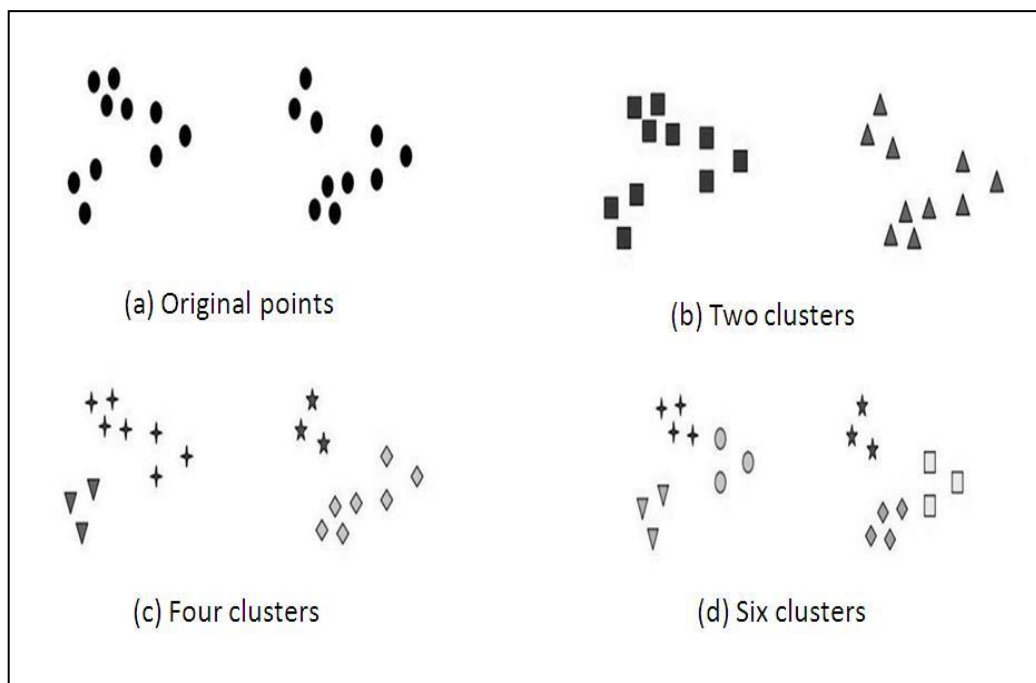
Tugas deteksi anomali berkaitan dalam menemukan dan mendeteksi data yang memiliki karakteristik yang menyimpang dan berbeda dari sebagian data lainnya. Data-data yang menyimpang dan berbeda tersebut biasa disebut dengan *outlier*. Algoritma dalam deteksi anomali harus mempunyai tingkat kesalahan yang rendah dan mempunyai tingkat deteksi yang tinggi.

2.2.3 *Clustering*

Menurut [11] menyatakan bahwa “*Clustering* membagi dataset ke dalam kelompok yang berbeda. Proses pengelompokan objek ke dalam kelompok tersebut bahwa objek dalam kelompok yang sama memiliki kemiripan dan objek dalam kelompok yang berbeda tidak memiliki kemiripan”. Pada umumnya tujuan *clustering* ialah berusaha untuk memaksimalkan variasi antar suatu *cluster* dan meminimalisasi variasi dalam satu *cluster* [7].

Clustering sering disebut dengan pembelajaran tidak terbimbing (*unsupervised learning*), dimana pengelompokan data berdasarkan kemiripan / ketidakmiripan antar-data tanpa adanya label kelas yang diketahui sebelumnya. Karena nantinya label baru bisa diberikan ketika *cluster* sudah terbentuk [3]. Analisa *cluster* dapat dianggap sebagai suatu bentuk klasifikasi yang memberi label objek-objek dengan label kelasnya. Hal ini merupakan perbedaan dengan analisa klasifikasi atau yang

sering disebut *supervised classification*, dimana jika terdapat satu data yang tidak diketahui kelasnya akan diberi label kelas menggunakan model dari objek-objek yang diketahui label kelasnya [5]. Pada Gambar 2.2 merupakan ilustrasi yang menunjukkan beberapa macam *clustering* dari *dataset* yang sama. Dimana Gambar 2.2 (a) merupakan titik awal sebelum dilakukan *clustering*; Gambar 2.2 (b) merupakan *clustering* yang memisahkan kelompok menjadi dua; Gambar 2.2 (c) merupakan pengelompokan menjadi empat *cluster*; dan Gambar 2.2 (d) merupakan pengelompokan menjadi enam *cluster*.



Gambar 2.2 Ilustrasi *Clustering* [5]

Menurut strukturnya, *clustering* dapat dibedakan menjadi dua yaitu *partitional clustering* dan *hierarchical clustering*.

1. Pengelompokan berbasis partisi (*partitional clustering*)

Merupakan pengelompokan yang membagi dataset, dimana data yang berada pada setiap *cluster* tidak bertumpang-tindih dan hanya menjadi anggota dari satu *cluster*. Metode yang masuk dalam kategori *partitional clustering* ialah *K-Means*, *DBSCAN*, dan *Self-Organizing Map (SOM)* [3].

2. Pengelompokan berbasis hierarki (*hierarchical clustering*)

Merupakan pengelompokan data, dimana tiap *cluster* merupakan gabungan dari *subcluster*. Dan satu data tunggal dapat menjadi sebuah *cluster*. Metode yang dapat digunakan ialah *Agglomerative Hierarchical Clustering* (AHC) [3].

2.2.4 Algoritma *K-Means*

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *k cluster* yang sudah ditetapkan di awal [3]. Berikut merupakan langkah-langkah dari algoritma *K-Means* [3][8]:

1. Tentukan *k* sebagai jumlah *cluster* yang dibentuk.

Dalam menentukan banyaknya *cluster k* dapat dilakukan dengan pertimbangan konseptual ataupun teoritis dan sesuai dengan tujuan dari pengelompokan.

2. Tentukan titik pusat (*centroid*) dari setiap *cluster*.

Untuk menentukan *centroid* awal setiap *cluster* dapat dilakukan secara *random* / acak.

3. Alokasikan semua data ke *centroid* terdekat dengan matrik jarak yang sudah ditetapkan.

Rumus dalam menghitung jarak antara lain : jarak *Euclidean*, jarak *Manhattan* / *Block City*, dan jarak *Minkowski*. Dalam penelitian ini untuk menghitung jarak antara objek dengan *centroid* menggunakan *Euclidian Distance* dengan perhitungan sebagai berikut :

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3 \dots n \quad (2.1)$$

dimana x_i : objek x ke- i

y_i : daya y ke- i

n : banyaknya objek

4. Hitung kembali titik pusat (*centroid*) berdasarkan data yang mengikuti *cluster* masing-masing.

Untuk menghitung titik pusat (*centroid*) *cluster* ke-*i* berikutnya, menggunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1,2,3 \dots n \quad (2.2)$$

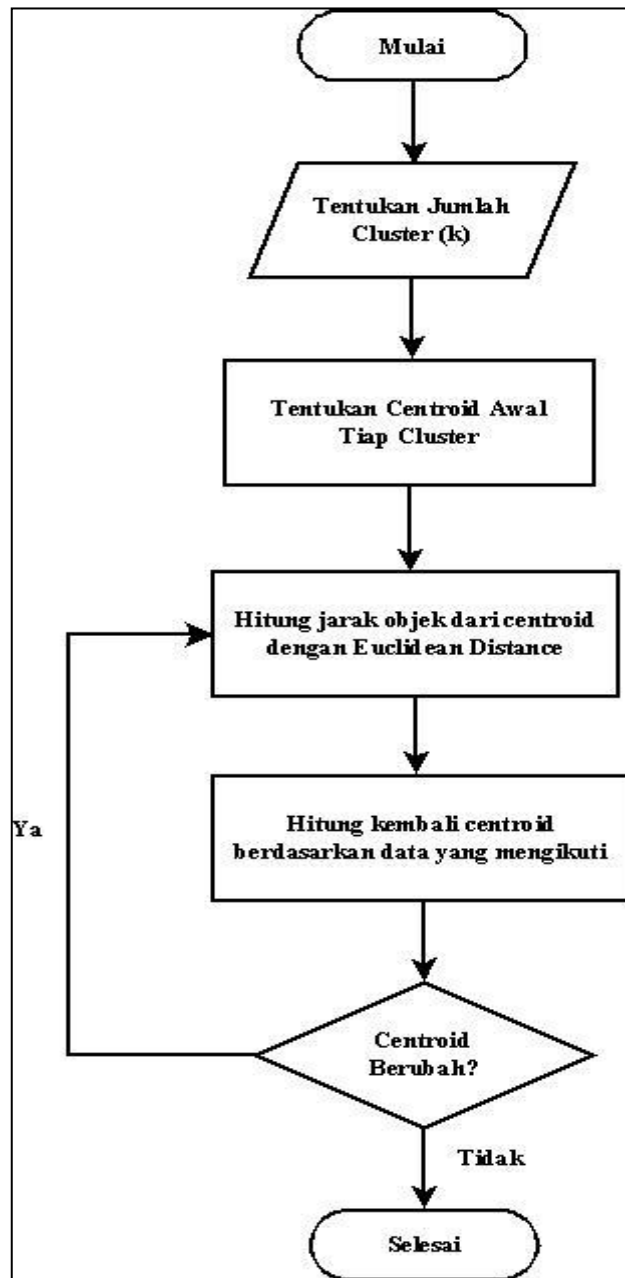
dimana *v* : *centroid* pada *cluster*

x_i : objek ke-*i*

n : banyaknya objek

5. Ulangi langkah 3 dan 4 jika posisi *centroid* baru berubah / tidak sama.

Algoritma *K-Means* mencapai kondisi *konvergen* ketika pengalokasian kembali titik data dan juga lokasi *centroid* tidak berubah [3]. Pengecekan *konvergensi* dilakukan dengan membandingkan iterasi sebelumnya pada iterasi yang sedang berjalan. Jika hasilnya berbeda maka kondisi belum *konvergen* untuk itu perlu dilakukan iterasi berikutnya, dan jika hasilnya sama maka algoritma *K-Means* sudah *konvergen*. Gambar 2.3 menjelaskan alur dari algoritma *K-Means*.



Gambar 2.3 Flowchart Algoritma K-Means

2.2.5 CRISP-DM

Cross-Industry Standart Process for Data Mining (CRISP-DM) menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [12]. Berikut ini merupakan fase CRISP-DM, yang telah ditetapkan dalam pengembangan *data mining* [9]:

1. *Business understanding*

Pada fase pertama dimulai dengan memahami tujuan dan kebutuhan bisnis, kemudian menterjemahkan ke dalam pendefinisian masalah dalam *data mining*. Selanjutnya tentukan rencana dan strategi dalam mencapai tujuan tersebut.

2. *Data understanding*

Fase kedua bertujuan untuk pembelajaran pada data, pengumpulan dan melakukan penyeleksian data.

3. *Data preparation*

Pada fase ketiga dilakukan pemilihan data, pembersihan data, mengintegrasikan data, dan transformasi data.

4. *Modeling*

Fase ini bertujuan untuk memilih dan menerapkan teknik pemodelan yang sesuai.

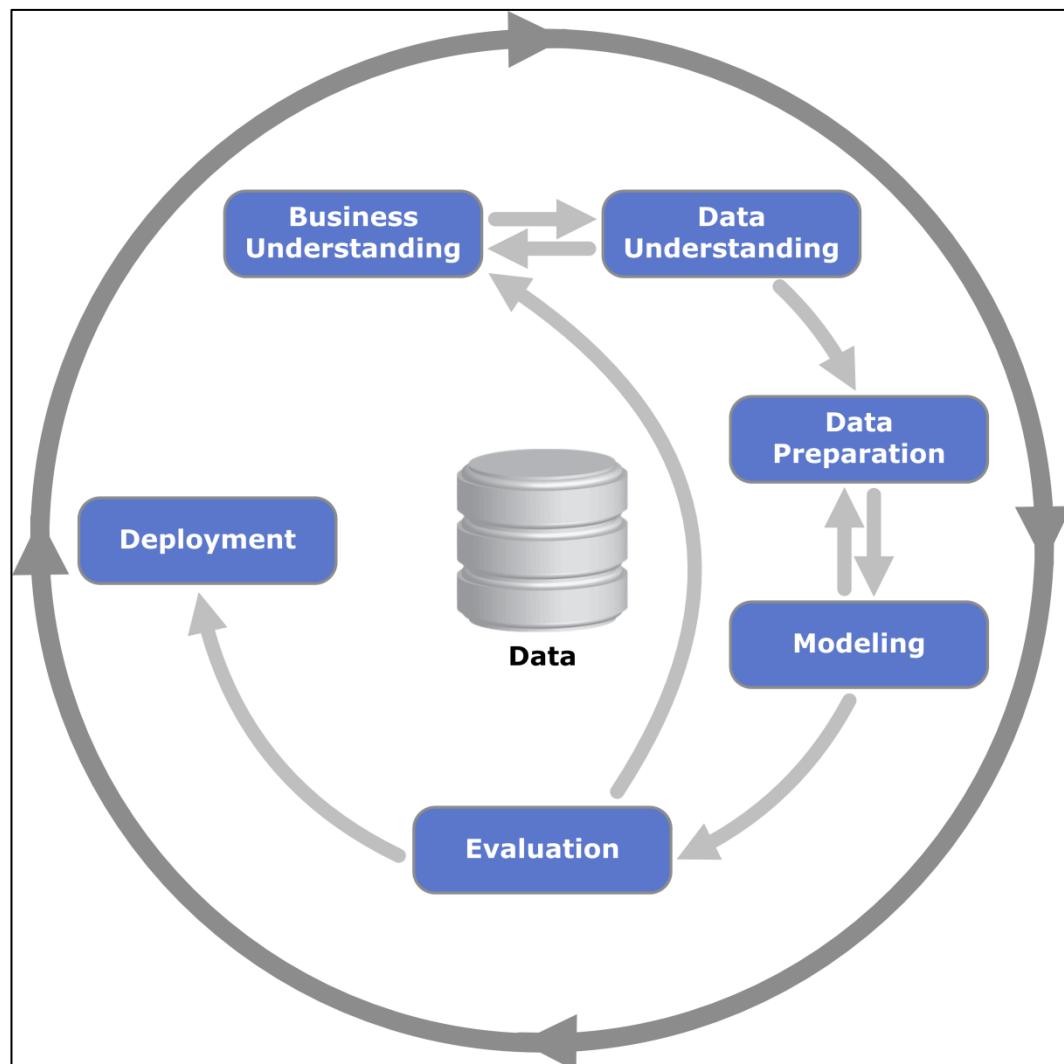
5. *Evaluation*

Pada fase ini akan dilakukan proses evaluasi dari fase sebelumnya.

6. *Deployment*

Fase ini dilakukan untuk penyusunan laporan atau presentasi hasil pengetahuan yang didapatkan dari proses *data mining*.

Gambar 2.4 menjelaskan tentang siklus hidup pengembangan *data mining* yang telah ditetapkan oleh CRISP-DM.



Gambar 2.4 Proses CRISP-DM

(Sumber : Kenneth Jensen, 2012)

2.2.6 *Rapidminer*

Rapidminer merupakan *software* berlisensi gratis yang digunakan sebagai alat bantu dalam *data mining* dan *knowledge discovery*. Pada *rapidminer* terdapat 400 prosedur *data mining*, termasuk operator untuk masukan, *output*, *data preprocessing* dan visualisasi [10]. Berikut ini merupakan beberapa fungsi menu pada *rapidminer*:

1. *Process control* : untuk mengontrol aliran proses
2. *Utility* : untuk mengelompokkan subprocess
3. *Repository access* : untuk membaca data dan menulis data di repository
4. *Import* : untuk membaca data
5. *Export* : untuk menulis data
6. *Data Transformation* : untuk mengubah bentuk data
7. *Modelling* : untuk melakukan teknik *data mining* seperti *clustering*, klasifikasi, asosiasi, regresi, dll.
8. *Evaluation* : digunakan untuk menghitung kualitas *modelling*.