

BAB II

LANDASAN TEORI

2.1 Data Mining

Secara sederhana, data mining merupakan ekstraksi informasi yang tersirat dalam sekumpulan data. Data mining merupakan sebuah proses untuk menggali kumpulan data dan menemukan informasi di dalamnya. (Turban, dkk, 2005). Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan tehnik gambar dari statistik, mesin pembelajaran dan sistem manajemen database. Penggalan data ini dilakukan pada sekumpulan data yang besar untuk menemukan pola atau hubungan yang ada dalam kumpulan data tersebut (Kusrini & Luthfi, 2009). Hasil penemuan yang diperoleh setelah proses penggalan data ini, kemudian dapat digunakan untuk analisis yang lebih lanjut.

Data mining yang disebut juga dengan *Knowledge-Discovery in Database* (KDD) adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*). Proses KDD ini terdiri dari langkah-langkah sebagai berikut (Han, J. & Kamber, M, 2001):

1. *Data Cleaning*, proses menghapus data yang tidak konsisten dan kotor
2. *Data Integration*, penggabungan beberapa sumber data
3. *Data Selection*, pengambilan data yang akan dipakai dari sumber data
4. *Data Transformation*, proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining

5. *Data Mining*, suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data
6. *Pattern Evaluation*, proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada didalam data itu sendiri
7. *Knowledge Presentation*, proses visualisasi dan teknik menyajikan knowledge digunakan untuk menampilkan knowledge hasil mining kepada user.

2.2 Clustering

Clustering merupakan sebuah teknik penggalian data yang bersifat *unsupervised*, karena tidak ada satu atributpun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama. Pengklasteran merupakan satu dari sekian banyak fungsi proses data mining untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Pengelompokan data ini didasarkan pada kesamaan karakter atau kriteria dari data-data yang dianalisis. Data-data yang ada dalam cluster yang sama memiliki karakter atau kriteria yang sama, sementara data-data yang berada dalam cluster yang berbeda juga memiliki karakter atau kriteria yang berbeda (Agusta, Y, 2007).

Analisis kelompok (*cluster analysis*) adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Tan, 2006). Tujuan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan dan mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya (Prasetyo, E, 2012).

Metode klasterisasi secara umum dapat dibagi menjadi dua yaitu (Tan, 2006):

- a. *Hierarchical clustering* yaitu pengelompokkan data melalui suatu bagan yang berupa hirarki, dimana terdapat penggabungan dua grup yang terdekat di setiap iterasinya ataupun pembagian dari seluruh set data kedalam kluster-kluster.
- b. *Partitional clustering* yaitu pengelompokkan ke dalam sejumlah kluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada metode ini setiap kluster memiliki titik pusat kluster (*centroid*) dan secara umum metode ini memiliki fungsi tujuan yaitu meminimumkan jarak (*dissimilarity*) dari seluruh data ke pusat kluster masing-masing.

2.3 Rekayasa Perangkat Lunak

Istilah Rekayasa Perangkat Lunak (RPL) secara umum disepakati sebagai terjemahan dari istilah *Software Engineering*. Istilah *Software Engineering* mulai dipopulerkan tahun 1968 pada *Software Engineering Conference* yang diselenggarakan oleh NATO. Rekayasa perangkat lunak (RPL) adalah suatu disiplin ilmu yang membahas semua aspek produksi perangkat lunak, mulai dari tahap awal yaitu analisa kebutuhan pengguna, menentukan spesifikasi dari kebutuhan pengguna, desain, pengkodean, pengujian sampai pemeliharaan sistem setelah digunakan (Mulyanto, R, 2008).

Sebagian orang mengartikan RPL hanya sebatas pada bagaimana membuat program computer. Padahal ada perbedaan mendasar antara perangkat lunak (*software*) dan program computer. Perangkat lunak adalah seluruh perintah yang digunakan untuk memproses informasi. Perangkat lunak dapat berupa program atau prosedur. Program adalah kumpulan perintah yang dimengerti oleh computer sedangkan prosedur adalah perintah yang dibutuhkan oleh pengguna dalam memproses informasi (O'Brien, 1999)

Beberapa tujuan yang dilakukan rekayasa perangkat lunak antara lain (Mulyanto, R, 2008):

1. Memperoleh biaya produksi perangkat lunak yang rendah.

2. Menghasilkan perangkat lunak yang kinerjanya tinggi, andal dan tepat waktu.
3. Menghasilkan perangkat lunak yang dapat bekerja pada berbagai jenis *platform*.
4. Menghasilkan perangkat lunak yang biaya perawatannya rendah.

2.4 Sistem Cerdas

Kecerdasan Buatan (*Artificial Intellegence*) merupakan sebuah studi tentang bagaimana membuat komputer melakukan hal-hal yang pada saat ini dapat dilakukan lebih baik oleh manusia (Rich and Knight, 1991). Kecerdasan Buatan (AI) merupakan cabang dari ilmu komputer yang dalam merepresentasi pengetahuan lebih banyak menggunakan bentuk simbol-simbol daripada bilangan, dan memproses informasi berdasarkan metode heuristic atau dengan berdasarkan sejumlah aturan (Encyclopedia Britannica).

Beberapa Tujuan dari kecerdasan buatan antara lain (Winston dan Prendergast, 1984) :

1. Membuat mesin menjadi lebih pintar (tujuan utama)
2. Memahami apa itu kecerdasan (tujuan ilmiah)
3. Membuat mesin lebih bermanfaat (tujuan entrepreneurial)

2.5 Buku Panduan Akademik UMG 2012/2013

Buku panduan akademik merupakan suatu buku yang menjadi petunjuk bagi seorang mahasiswa selama melaksanakan kegiatan perkuliahan. Buku panduan akademik bisa jadi berbeda untuk setiap universitas. Buku panduan akademik yang akan digunakan dalam penelitian ini merupakan buku panduan akademik Universitas Muhammadiyah Gresik tahun 2013/2014. Buku ini berisi informasi umum, peraturan akademik, prosedur-prosedur kerja dan kurikulum untuk setiap program studi di Universitas Muhammadiyah Gresik.

2.6 K-Means

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* (Agusta, Y, 2007).

Langkah-langkah pengelompokan data dengan metode K-Means, sebagai berikut (Prasetyo, Eko, 2012):

1. Tentukan jumlah kelompok.
2. Alokasikan data ke dalam kelompok secara acak.
3. Hitung pusat kelompok (sentroid/rata-rata) dari data yang ada di masing-masing kelompok dengan menggunakan rumus korelasi antar dua objek yaitu Euclidean.

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \dots\dots\dots (2.1)$$

Dimana : $D(x_2, x_1)$ = jarak antara data x_2 dan x_1

x_1 = data ke-1

x_2 = data ke-2

4. Alokasikan masing-masing data ke sentroid/rata-rata terdekat.
5. Kembali ke Langkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi obyektif yang digunakan masih di atas nilai ambang yang ditentukan. Rumus perhitungan lokasi sentroid (titik pusat) setiap kelompok yang diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya :

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad \dots\dots\dots (2.2)$$

Dimana : C_i = sentroid fitur ke-i
M = jumlah data dalam sebuah kelompok
i = fitur ke-i dalam sebuah kelompok
p = dimensi data

Adapun karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal klaster karena K-Means membangkitkan titik pusat klaster awal secara random. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang atributnya bernilai numeric (Pena, J. M., Lozano, J. A. and Larranaga, P,1999).

2.7 K – Harmonic Means (KHM)

K-Harmonic Means (KHM) pertama kali diperkenalkan oleh Zhang, Hsu, dan Dayal (1999) dari HP Laboratories Palo Alto yang kemudian dikembangkan oleh Hammerly dan Elkan pada tahun 2002. Tujuan pengembangan metode KHM adalah untuk menangani masalah utama dalam K-Means yang hasil clusteringnya sangat sensitif dengan inisialisasi data yang dijadikan sebagai centroid awal. Hasil yang sering berbeda (lokal optima) dari proses clusteringnya (pada set data yang sama) disebabkan oleh inisialisasi centroid yang berbeda.

KHM juga salah satu metode clustering berbasis partisi yang menggunakan rata-rata harmonic (*harmonic average*) jarak dari setiap titik data ke centroid sebagai komponen dalam fungsi kinerja (fungsi objektif). KHM secara signifikan meningkatkan kualitas hasil clustering dibandingkan dengan metode seperti K-Means maupun Expectation Maximization (EM). Kualitas yang lebih baik tersebut adalah bahwa hasil cluster yang didapat berusaha mendekati hasil yang global optima (hasil cluster yang didapat selalu sama).

Secara prinsip, KHM menggunakan jumlah semua titik data dari rata-rata harmonik kuadrat jarak dari titik data ke semua centroid sebagai fungsi objektifnya, seperti disajikan dalam persamaan (2.3). Persamaan (2.3) sebagai fungsi objektif KHM sangat berbeda terhadap K-Means yang menggunakan total varian data dalam cluster. Dalam KHM, $C = \{c_j | j = 1, \dots, K\}$ adalah K centroid, dan $X = \{x_i | i=1, \dots, N\}$ adalah N data yang dicluster, fungsi objektif KHM diberikan oleh persamaan (2.3)

$$J = \sum_{i=1}^N \frac{K}{\sum_{j=1}^K \frac{1}{\|x_i - c_j\|^2}} \dots\dots\dots (2.3)$$

- Dimana :
- J = fungsi objektif KHM
 - N = data yang dicluster
 - K = centroid
 - x_i = data ke-i
 - c_j = sentroid ke-j

Kuantitas didalam ruas kanan adalah rata-rata harmonik dari K jarak kuadrat, $\{\|x_i - c_j\|^2 | j = 1, \dots, K\} \dots\dots\dots (2.4)$

Algoritma *clustering* dengan *K-Harmonic Means* sebagai berikut (Zhang, B. *et al.*, 1999):

1. Tentukan Nilai K sebagai jumlah kelompok / *cluster*.
2. Inisialisasi posisi centroid awal dimana $C = \{c_j | j = 1, \dots, K\}$ sebanyak K *centroid* secara acak dari data yang ada.
3. Hitung Jarak data terhadap masing-masing centroid. Misalnya menggunakan rumus jarak *euclidean* seperti persamaan berikut :

$$d_{i,j} = \|x_i - c_j\|_2 = \sqrt{(x_i - c_j)^2} \dots\dots\dots (2.5)$$

Dimana $X = \{x_i | i=1 \dots N\}$, N adalah jumlah data yang akan diklaster dengan metode KHM.

4. Cari jarak terdekat $d_{i,min}$ dan masukkan X kedalam *cluster* sesuai dengan kelompok/*centroid* tersebut.
5. Cari *centroid* baru sebanyak K dengan persamaan KHM seperti berikut :

$$m_{i,k} = \frac{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{j=1}^K \frac{1}{d_{i,j}^2} \right)^2} \cdot x_i}{\sum_{i=1}^N \frac{1}{d_{i,k}^3 \left(\sum_{j=1}^K \frac{1}{d_{i,j}^2} \right)^2}} \quad \dots \quad (2.6)$$

Dimana : $m_{i,k}$ = sentroid baru metode KHM

N = data yang dicluster

$d_{i,k}$ = jarak antara i ke k

$d_{i,j}$ = jarak antara i ke j

x_i = data ke- i

Catatan : $d_{i,\min} = 0$ maka vektor m_k diset menjadi 0.

6. Lakukan langkah nomor 2 - 4 hingga posisi anggota *cluster* tidak berubah.

2.8 Contoh Perhitungan KHM

Contoh perhitungan KHM menggunakan data set yang terdiri dari 10 data yang memiliki 2 atribut yaitu mata kuliah1 (MK1) dan mata kuliah2 (MK2). Berikut tabel data tersebut :

Tabel 2.1 Data Set Nilai Mata Kuliah

No.	Data	MK1	MK2
1.	A	3	4
2.	B	2	8
3.	C	8	1
4.	D	1	7
5.	E	4	5
6.	F	5	3
7.	G	2	5
8.	H	2	6
9.	I	6	3
10.	J	2	7

Langkah pertama, melakukan perhitungan menggunakan metode *K-Harmonic Means* (KHM) adalah menentukan nilai *K* sebagai jumlah kelompok/*cluster*. Nilai *K* pada contoh perhitungan ini telah ditentukan sebanyak 2.

Iterasi 1

Langkah kedua, melakukan inisialisasi posisi centroid awal dimana $C = \{c_j | j = 1, \dots, K\}$ sebanyak *K* centroid secara acak dari data yang ada, yaitu :

Tabel 2.2 Centroid Awal

No.	Data	MK1	MK2
3.	A	3	4
2.	F	5	3

Ketiga, menghitung jarak data terhadap masing-masing centroid menggunakan rumus jarak *euclidean* seperti persamaan (2.5)

$$d_{1,1} = \sqrt{(3-3)^2 + (4-4)^2} = 0$$

$$d_{1,2} = \sqrt{(3-5)^2 + (4-3)^2} = 2.24$$

$$d_{2,1} = \sqrt{(2-3)^2 + (8-4)^2} = 4.12$$

$$d_{2,2} = \sqrt{(2-5)^2 + (8-3)^2} = 5.83$$

$$d_{3,1} = \sqrt{(8-3)^2 + (1-4)^2} = 5.83$$

$$d_{3,2} = \sqrt{(8-5)^2 + (1-3)^2} = 3.61$$

$$d_{4,1} = \sqrt{(1-3)^2 + (7-4)^2} = 3.61$$

$$d_{4,2} = \sqrt{(1-5)^2 + (7-3)^2} = 5.66$$

$$d_{5,1} = \sqrt{(4-3)^2 + (5-4)^2} = 1.41$$

$$d_{5,2} = \sqrt{(4-5)^2 + (5-3)^2} = 2.24$$

$$d_{6,1} = \sqrt{(5-3)^2 + (3-4)^2} = 2.24$$

$$d_{6,2} = \sqrt{(5-5)^2 + (3-3)^2} = 0$$

$$d_{7,1} = \sqrt{(2-3)^2 + (5-4)^2} = 1.41$$

$$d_{7,2} = \sqrt{(2-5)^2 + (5-3)^2} = 3.61$$

$$d_{8,1} = \sqrt{(2-3)^2 + (6-4)^2} = 2.24$$

$$d_{8,2} = \sqrt{(2-5)^2 + (6-3)^2} = 4.24$$

$$d_{9,1} = \sqrt{(6-3)^2 + (3-4)^2} = 3.16$$

$$d_{9,2} = \sqrt{(6-5)^2 + (3-3)^2} = 1$$

$$d_{10,1} = \sqrt{(2-3)^2 + (7-4)^2} = 3.16$$

$$d_{10,2} = \sqrt{(2-5)^2 + (7-3)^2} = 5$$

Langkah keempat, mencari jarak terdekat $d_{i,min}$ dan masukkan *X* kedalam *cluster* sesuai dengan kelompok/*centroid* tersebut.

Tabel 2.3 Hasil Perhitungan Jarak dan Pengelompokan Data

No.	Data	Jarak C1	Jarak C2	Jarak Min	Cluster
1.	A	0.00	2.24	0.00	1
2.	B	4.12	5.83	4.12	1
3.	C	5.83	3.61	3.61	2
4.	D	3.61	5.66	3.16	1
5.	E	1.41	2.24	1.41	1
6.	F	2.24	0.00	0.00	2
7.	G	1.41	3.61	1.41	1
8.	H	2.24	4.24	2.24	1
9.	I	3.16	1.00	1.00	2
10.	J	3.16	5.00	3.16	1

Kelima, mencari *centroid* baru sebanyak K sesuai dengan persamaan (2.6):

$$\begin{aligned}
 M_{1,1} = & \frac{0 + \left(\frac{1}{4.12^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \cdot 2 \right) + \left(\frac{1}{5.83^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \cdot 8 \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \cdot 1 \right)}{0 + \left(\frac{1}{4.12^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \right) + \left(\frac{1}{5.83^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \right)} \\
 & + \frac{\left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \cdot 4 \right) + 0 + \left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \cdot 2 \right) + \left(\frac{1}{2.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \cdot 2 \right)}{\left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \right) + 0 + \left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{2.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \right)} \\
 & \frac{\left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \cdot 6 \right) + \left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \cdot 1 \right)}{\left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \right) + \left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \right)} = 2.27
 \end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{1,2} = & \frac{0 + \left(\frac{1}{4.12^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \cdot 8 \right) + \left(\frac{1}{5.83^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \cdot 1 \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \cdot 7 \right)}{0 + \left(\frac{1}{4.12^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \right) + \left(\frac{1}{5.83^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \right)} \\
& + \frac{\left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \cdot 5 \right) + 0 + \left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \cdot 5 \right) + \left(\frac{1}{2.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \cdot 6 \right)}{\left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \right) + 0 + \left(\frac{1}{1.41^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{2.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \right)} \\
& \frac{\left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \cdot 3 \right) + \left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \cdot 7 \right)}{\left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \right) + \left(\frac{1}{3.16^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \right)} = \mathbf{6.34}
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_{2,1} = & \frac{0 + \left(\frac{1}{5.83^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \cdot 2 \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \cdot 8 \right) + \left(\frac{1}{5.66^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \cdot 1 \right)}{0 + \left(\frac{1}{5.83^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{5.66^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \right)} \\
& + \frac{\left(\frac{1}{2.24^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \cdot 4 \right) + 0 + \left(\frac{1}{3.61^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \cdot 2 \right) + \left(\frac{1}{4.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \cdot 2 \right)}{\left(\frac{1}{2.24^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \right) + 0 + \left(\frac{1}{3.61^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{4.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \right)} \\
& \frac{\left(\frac{1}{1^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \cdot 6 \right) + \left(\frac{1}{5^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \cdot 1 \right)}{\left(\frac{1}{1^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \right) + \left(\frac{1}{5^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \right)} = \mathbf{5.10}
\end{aligned}$$

$$\begin{aligned}
M_{2,2} = & \frac{0 + \left(\frac{1}{5.83^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \cdot 8 \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \cdot 1 \right) + \left(\frac{1}{5.66^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \cdot 7 \right)}{0 + \left(\frac{1}{5.83^3 \left(\frac{1}{4.12^2} + \frac{1}{5.83^2} \right)^2} \right) + \left(\frac{1}{3.61^3 \left(\frac{1}{5.83^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{5.66^3 \left(\frac{1}{3.61^2} + \frac{1}{5.66^2} \right)^2} \right)} \\
& + \frac{\left(\frac{1}{2.24^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \cdot 5 \right) + 0 + \left(\frac{1}{3.61^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \cdot 5 \right) + \left(\frac{1}{4.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \cdot 6 \right)}{\left(\frac{1}{2.24^3 \left(\frac{1}{1.41^2} + \frac{1}{2.24^2} \right)^2} \right) + 0 + \left(\frac{1}{3.61^3 \left(\frac{1}{1.41^2} + \frac{1}{3.61^2} \right)^2} \right) + \left(\frac{1}{4.24^3 \left(\frac{1}{2.24^2} + \frac{1}{4.24^2} \right)^2} \right)} \\
& + \frac{\left(\frac{1}{1^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \cdot 3 \right) + \left(\frac{1}{5^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \cdot 7 \right)}{\left(\frac{1}{1^3 \left(\frac{1}{3.16^2} + \frac{1}{1^2} \right)^2} \right) + \left(\frac{1}{5^3 \left(\frac{1}{3.16^2} + \frac{1}{5^2} \right)^2} \right)} = 3.87
\end{aligned}$$

Tabel 2.4 Centroid Baru

No.	MK1	MK2
1.	2.27	6.34
2.	5.10	3.87

Lakukan langkah nomor 2 - 4 hingga posisi anggota *cluster* tidak berubah.

Menghitung jarak data terhadap centroid baru menggunakan rumus jarak *euclidean* seperti persamaan (2.5)

Tabel 2.5 Hasil Perhitungan Jarak dan Pengelompokan Data dengan Centroid Baru

No.	Data	Jarak C1	Jarak C2	Jarak Min	Cluster
1.	A	2.45	2.10	2.10	2
2.	B	1.68	5.16	1.68	1
3.	C	7.83	4.08	4.08	2

4.	D	1.43	5.16	1.43	1
5.	E	2.19	1.57	1.57	2
6.	F	4.31	0.88	0.88	2
7.	G	1.37	3.30	1.37	1
8.	H	0.43	3.76	0.43	1
9.	I	5.01	1.25	1.25	2
10.	J	0.71	4.40	0.71	1

Iterasi 2

Mencari *centroid* baru sebanyak K menggunakan metode KHM seperti persamaan (2.6)

Tabel 2.6 Centroid Baru Iterasi 2

No.	MK1	MK2
1.	2.34	6.12
2.	6.03	2.58

Menghitung jarak data terhadap centroid baru menggunakan rumus jarak *euclidean* seperti persamaan (2.5)

Tabel 2.7 Hasil Perhitungan Jarak dan Pengelompokan Data dengan Centroid Baru Iterasi 2

No.	Data	Jarak C1	Jarak C2	Jarak Min	Cluster
1.	A	2.22	3.34	2.22	1
2.	B	1.91	6.75	1.91	1
3.	C	7.63	2.53	2.53	2
4.	D	1.60	6.69	1.60	1
5.	E	2.00	3.15	2.00	1
6.	F	4.10	1.11	1.11	2

7.	G	1.17	4.70	1.17	1
8.	H	0.36	5.28	0.36	1
9.	I	4.81	0.42	0.42	2
10.	J	0.95	5.98	0.95	1

Iterasi 3

Mencari *centroid* baru sebanyak K menggunakan metode KHM seperti persamaan (2.6)

Tabel 2.8 Centroid Baru Iterasi 3

No.	MK1	MK2
1.	2.29	6.13
2.	6.33	2.27

Menghitung jarak data terhadap centroid baru menggunakan rumus jarak *euclidean* seperti persamaan (2.5)

Tabel 2.9 Hasil Perhitungan Jarak dan Pengelompokan Data dengan Centroid Baru Iterasi 3

No.	Data	Jarak C1	Jarak C2	Jarak Min	Cluster
1.	A	2.25	3.75	2.25	1
2.	B	1.89	7.18	1.89	1
3.	C	7.68	2.10	2.10	2
4.	D	1.55	7.13	1.55	1
5.	E	2.05	2.90	2.05	1
6.	F	4.14	1.52	1.52	2
7.	G	1.17	5.12	1.17	1
8.	H	0.32	5.72	0.32	1
9.	I	4.86	0.80	0.80	2

10.	J	0.92	6.41	0.92	1
-----	---	------	------	------	---

Dari hasil perhitungan, didapatkan hasil :

- Data A, B, D, E, G, H dan J termasuk *cluster*/kelompok 1
- Data C, F dan I termasuk *cluster*/kelompok 2
- Cluster tetap berhasil ditemukan pada iterasi ke-3.

2.9 Hasil Uji Konsistensi Metode KHM

Pengujian dilakukan pada 10 data acak yang memiliki 2 fitur seperti pada **tabel 2.1**, kemudian data-data tersebut diclustering menggunakan metode KHM dengan nilai $k=2$. Proses pengujian dilakukan sebanyak 25 kali percobaan dengan inisialisasi centroid secara acak di setiap proses clustering. Hasil yang didapatkan dapat dilihat pada **tabel 2.10**. Dari 25 kali percobaan anggota cluster tidak tetap didapatkan hanya pada 2 kali percobaan.

Tabel 2.10 Hasil Uji Konsistensi Metode KHM

No	Centroid		Cluster		Iterasi
	1	2	1	2	
1	A	B	C, F, I	A, B, D, E, G, H, J	2
2	A	C	A, B, D, E, G, H, J	C, F, I	3
3	A	D	C, F, I	A, B, D, E, G, H, J	2
4	A	E	C, F, I	A, B, D, E, G, H, J	2
5	A	F	A, B, D, E, G, H, J	C, F, I	3
6	A	G	C, F, I	A, B, D, E, G, H, J	2
7	A	H	C, F, I	A, B, D, E, G, H, J	2
8	A	I	A, B, D, E, G, H, J	C, F, I	3
9	A	J	C, F, I	A, B, D, E, G, H, J	5
10	B	C	C, F, I	A, B, D, E, G, H, J	2
11	B	D	-	A, B, C, D, E, F, G, H, I, J	10
12	B	E	C, F, I	A, B, D, E, G, H, J	2
13	B	F	A, B, D, E, G, H, J	C, F, I	2

14	B	G	A, B, D, E, G, H, J	C, F, I	4
15	B	H	-	A, B, C, D, E, F, G, H, I, J	8
16	B	I	A, B, D, E, G, H, J	C, F, I	2
17	B	J	B, D, G, H, J	A, C, E, F, I	6
18	C	D	C, F, I	A, B, D, E, G, H, J	1
19	C	E	C, F, I	A, B, D, E, G, H, J	3
20	C	F	C, F, I	A, B, D, E, G, H, J	3
21	C	G	C, F, I	A, B, D, E, G, H, J	2
22	C	H	C, F, I	A, B, D, E, G, H, J	1
23	C	I	C, F, I	A, B, D, E, G, H, J	3
24	C	J	C, F, I	A, B, D, E, G, H, J	1
25	D	E	A, B, D, E, G, H, J	C, F, I	2

Berdasarkan hasil pengujian dapat disimpulkan bahwa anggota clustering metode KHM memiliki tingkat konsistensi yang sangat baik yaitu 92%.

2.10 Penelitian Sebelumnya

Penelitian tentang metode *K-Harmonic Means* (KHM) pertama kali dilakukan oleh Zhang Bin, Hsu Meichun, dan Dayal Umeshwar pada tahun 1999 dari HP Laboratories Palo Alto dengan judul penelitian “**K-Harmonic Means, A Data Clustering Algorithm**”. Penelitian ini membuat pandangan terpadu dari dua algoritma clustering yang paling banyak digunakan dan populer yaitu K-Means (KM) dan Expectation Maximization (EM) dengan metode algoritma K-Harmonic Means yang merupakan hasil penyempurnaan dari algoritma K-Means.

Dari hasil penelitian didapatkan kesimpulan bahwa KHM bekerja sangat baik dan cepat dibandingkan EM dan KM meskipun inisialisasi dilakukan secara acak bahkan ketika diberikan inisialisasi sangat buruk. Namun, pada data skala besar KHM memerlukan waktu yang sedikit lebih lama dibandingkan dengan KM untuk mendapatkan hasil yang baik.

Penelitian lain tentang metode *K-Harmonic Means* (KHM) dilakukan oleh I Made Widiartha seorang mahasiswa dari jurusan Ilmu Komputer,

Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana pada tahun 2011. Penelitian ini berjudul “**Studi Komparasi Metode Klasterisasi Data K-Means dan K-Harmonic Means.**” Sesuai dengan judulnya penelitian ini melakukan perbandingan hasil clustering antara dua metode clustering yaitu metode *partitional clustering* yang sangat populer yaitu K-Means (KM) dan K-Harmonic Means (KHM) yang merupakan hasil penyempurnaan dari K-Means. Penelitian ini ditujukan untuk melihat bagaimana performa metode KHM dalam menyempurnakan metode KM. Studi komparasi ini menggunakan lima buah data set.

Dari hasil penelitian ini didapatkan hasil bahwa metode KHM telah terbukti berhasil mengoptimalkan posisi titik pusat kluster dengan mengarahkan hasil kluster menuju solusi global optimal. Hal ini dibuktikan dengan hasil penelitian yang menunjukkan nilai fungsi tujuan *objective function* dari metode KHM memiliki nilai yang lebih kecil dari metode KM di semua percobaan. Dari sisi penilaian hasil kluster secara eksternal menggunakan F-measure, metode KHM terlihat mendominasi daripada metode KM. Dari sisi waktu yang dibutuhkan untuk melakukan proses klasterisasi data, metode KHM membutuhkan waktu lebih lama dibandingkan dengan metode KM. Hal ini disebabkan oleh proses dalam KHM yang lebih kompleks daripada proses dalam KM.