Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 3, No. 1, Januari 2007 ISSN 0216 - 0544

Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering

Junaidillah Fadlil
Wayan Firdaus Mahmudy, (wayanfm@ub.ac.id)

Jurusan Matematika, FMIPA
Universitas Brawijaya

ABSTRAK

Sistem rekomendasi adalah suatu sistem yang digunakan untuk melakukan prediksi terhadap sesuatu objek. Sistem ini merupakan salah satu bentuk dari personalisasi web yang digunakan dalam sistem *e-commerce*. Pada tulisan ini dilakukan analisa terhadap penggabungan 2 metode yang diterapkan dalam sistem rekomendasi yaitu metode klasifikasi dengan menggunakan *decision tree* dan algoritma *k-means clustering*. Pengukuran tingkat akurasi dilakukan dengan membandingkan data asli dengan hasil prediksi yang didapatkan. Pada cluster sebanyak 210, rata-rata beda hasil sebesar 0,484. Semakin besar banyaknya cluster maka tingkat akurasi semakin baik.

Kata kunci: Sistem rekomendasi, *decision tree, k-means clustering*

1. PENDAHULUAN

Salah satu perkembangan teknologi pada sistem *e-commerce* adalah dengan adanya *personalisasi*. *Personalisasi* web adalah proses untuk mendapatkan atau mengumpulkan kecenderungan dari seorang user dalam melakukan belanja secara online dengan *e-commerce*. Salah satu keunggulan *personalisasi web* adalah adanya *sistem rekomendasi* kepada usernya, sehingga dengan adanya sistem tersebut seorang user akan mudah dalam memilih barang.

Beberapa contoh website yang telah menerapkan metode sistem rekomendasi adalah ebay.com, yahoo.com yang kemudian dikenal dengan *myYahoo*, amazon.com[4] dan masih banyak yang lainnya. *Collaborative filtering* adalah salah satu metode yang diguanakan untuk sistem rekomendasi. Salah satu contoh penerapan metode ini adalah pada sistem rekomendasi dalam memilih pembelian musik atau film, metode ini melakukan prediksi pada seorang user tentang musik atau film yang disenanginya. Prediksi yang dilakukan oleh sistem ini spesifik pada setiap user namun infomasi yang didapatkan berasal dari user-user yang lain. Metode lain yang dapat digunakan dalam sistem rekomendasi adalah metode *klasifikasi*, *aturan asosisasi* serta *data clustering* [3].

Pada metode klasifikasi metode yang dikenal dalam melakukan rekomendasi antara lain adalah pohon keputusan atau yang lebih dikenal dengan *decision tree*. Pada metode *clustering* banyak algoritma yang telah dikembangkan untuk diterapkan dalam sistem rekomendasi, antara lain adalah Algoritma *K-means clustering*, *Hierarchical*, *ROCK* dan lain sebagainya.

2. SISTEM REKOMENDASI

Sistem rekomendasi adalah suatu program yang melakukan prediksi sesuatu item, seperti rekomendasi film, musik, buku, berita dan lain sebagainya yang menarik user. Sistem ini berjalan dengan mengumpulkan data dari user secara langsung maupun tidak [2].

Pengumpulkan data secara langsung dapat dilakukan sebagai berikut :

1. Meminta user untuk melakukan rating pada sebuah item.
2. Meminta user untuk melakukan rangking pada item favorit setidaknya memilih satu item favorit.
3. Memberikan beberapa pilihan item pada user dan memintanya memilih yang terbaik.
4. Meminta user untuk mendaftar item yang paling disukai atau item yang tidak disukainya.

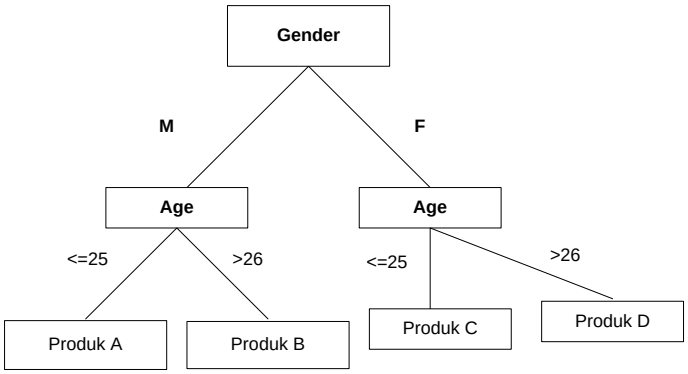
Pengumpulan data dengan tidak langsung berhubungan dengan seorang user, dilakukan dengan cara seperti berikut:

1. Mengamati item yang dilihat oleh seorang user pada sebuah web *e-commerce*.
2. Mengumpulkan data transaksi pada sebuah toko online.

Data hasil pengumpulan, kemudian dilakukan perhitungan dengan algoritma tertentu yang kemudian hasil tersebut dikembalikan lagi kepada user sebagai sebuah rekomendasi item dengan parameter dari user tersebut. Sistem rekomendasi juga merupakan salah satu alternatif sebagai mesin pencari suatu item yang dicari oleh user.

3. KLASIFIKASI

Klasifikasi adalah proses untuk menemukan sebuah model berdasarkan kelas-kelas yang digunakan sebagai pembeda antara kelas satu dengan kelas yang lain. *Decision tree* adalah penerapan metode klasifikasi yang paling populer, dengan metode ini sebuah item dapat dikelompokkan dan dimodelkan pada sebuah pohon keputusan, sehingga dapat dengan mudah dimengerti[3]. Contoh *decision tree* dapat dilihat pada Gambar 2 yang menerangkan bagaimana sebuah produk biasa dipilih oleh pembeli.



Gambar 1. Contoh *Decision tree*

4. CLUSTERING

Clustering adalah pengklasifikasian pada objek-objek yang sama menjadi kelompok-kelompok yang berbeda, dengan menjadikan partisi–partisi data yang ada menjadi kelompok yang baru, dan setiap kelompok data tersebut dikenali dengan tingkat jarak tertentu.[3]. "Data Clustering" adalah proses untuk mengelompokkan data–data yang sama berdasarkan tingkat kesamaan atau tingkat ketidaksamaan antar data atau item[1].

Data Clustering adalah suatu teknik yang biasa digunakan pada berbagai bidang, seperti data mining, kecerdasan buatan, pengenalan pola, penganalisaan gambar. Tujuan utama melakukan data *clustering* adalah untuk mengurangi jumlah data yang besar dengan memberikan kategori-kategori atau dengan mengelompokkan data–data yang memiliki tingkat kesamaan yang tinggi.

Algoritma *K-Means* merupakan satu salah algoritma yang banyak digunakan untuk melakukan pengelompokan data. Algoritma ini memecah kumpulan data menjadi *k* cluster dan mencari titik tengah dari tiap cluster. Titik tengah yang telah ditetapkan harus berada pada tempat yang tepat karena jika titik tengah suatu cluster memiliki koordinat yang berbeda maka hasil yang didapatkan juga akan mengalami perubahan. Sehingga jalan yang terbaik untuk mendapatkan nilai yang maksimal adalah dengan menempatkan setiap titik tengah setiap cluster pada jarak yang cukup jauh. Formula untuk algoritma K-Means adalah sebagai berikut [2]:

$$\text{Minimize } J = \sum_{j=1}^{k}\sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2 \qquad (1)$$

$\| x_i^{(j)} - c_j \|$ adalah jarak antara titik data dengan titik tengah dari cluster c_j.

Langkah–langkah algoritma *K-Means* adalah sebagai berikut:

1. Menentukan banyaknya jumlah *cluster*.
2. Menentukan pusat *cluster* pada setiap *cluster* yang telah dibentuk.
3. Menentukan jarak pada setiap titik ke setiap titik tengah *cluster*.

4. Untuk setiap titik yang telah ditandai dan dihitung jaraknya pada setiap titik tengah, maka ditentukan titik tersebut masuk pada cluster yang paling dekat jaraknya, sehingga membentuk anggota cluster yang baru.
5. Pada langkah yang ke empat dilihat apakah anggota pada setiap cluster tersebut berubah, jika masih berubah maka kembali ke Langkah 2
6. Selesai.

5. METODE PENELITIAN

5.1. Persiapan Data

Pada makalah ini, data yang digunakan adalah data dari *www.movielens.org*. kumpulan data pada website ini dikelola oleh sebuah kelompok belajar yang berasal dari Universitas Minnesota Amerika Serikat. Data yang akan diteliti ini berupa data rating film dari 934 responden pada 1682 film dan data rating yang terkumpul sebanyak 100.000 rating. Rating film yang diberikan antara skala 1 sampai dengan skala 5 yang terbesar. Setiap responden minimal melakukan rating pada 20 film. *Grouplens Reseach Group* menjelaskan data tersebut dikumpulkan melalui website Movielens (movielens.umn.edu) selama tujuh bulan pada periode 19 september 1997 sampai dengan 22 April 1998 pada sebuah proyek penelitian yang diberi nama *Grouplens Reseach Project*. Data tersebut telah dibersihan dari data yang kurang lengkap seperti responden yang melakukan rating pada film kurang dari 20, serta informasi kurang lengkap seorang responden baik umur, gender dan lain sebagainya. Berikut ini adalah keterangan pada setiap file.

1. **u.data**, file ini berisi seluruh kumpulan data 100.000 rating film oleh 934 responden pada 1628 film, yang setiap user melakukan rating minimal 20 film, data yang dikumpulkan adalah data acak untuk setiap user. Urutan data pada file ini adalah kode responden, film yang dirating oleh responden berupa kode film, kemudian rating yang diberikan oleh responden pada film, dan waktu responden tersebut memberikan rating pada film tersebut.
2. **u.info**, file ini memberika informasi tentang banyaknya responden, rating yang diberikan dan banyaknya film yang dirating oleh responden.
3. **u.item**, file ini berisi seluruh informasi film yang akan dilakukan rating, sedangkan urutan datanya adalah kode film, judul film, tanggal film ini dibuat, tanggal film ini diluncurkan ke pasaran, alamat website yang memuat tentang informasi film ini, kemudian genre film tersebut, apakah tidak diketahui, film laga, film tentang petualangan, apakah termasuk kedalam animasi, film untuk anak-anak, film lucu atau komedi, film tentang kejahatan, film dokumentasi, drama, fantasi, *film-noir*, film horror, film *musical*, misteri, drama romantis, sci-fi, trailer film, film tentang perang dan film barat. Pada genre film tersebut ditandai 1 untuk *genre* yang diketahui dan 0 untuk selain itu. Film yang ada bisa memiliki lebih dari satu *genre* pada file ini. Kode film pada file ini adalah kode yang digunakan pada file u.data.
4. **u.genre**, file ini berisi kumpulan *genre* film, atau keterangan lengkap untuk *genre* film yang digunakan pada file u.item.
5. **u.user**, file ini berisi tentang informasi tentang responden yang melakukan rating pada film. Urutan data pada file ini adalah kode responden, umur responden, pekerjaan responden, kode zip. Kode user pada file ini adalah kode user yang digunakan pada file u.data.
6. **u.occupation**, file ini berisi data tentang daftar pekerjaan dari responden.
7. **u.base** dan **u.test**, file tersebut merupakan u.data yang dipecah.
8. **allbut.pl**, file ini berisi bahasa pemrograman perl yang digunakan untuk *generate* data pada u.test
9. **mku.sh**, file ini berisi bahasa *shell script bash* yang digunakan untuk melakukan eksekusi pada file u.data

5.2 Langkah Penelitian

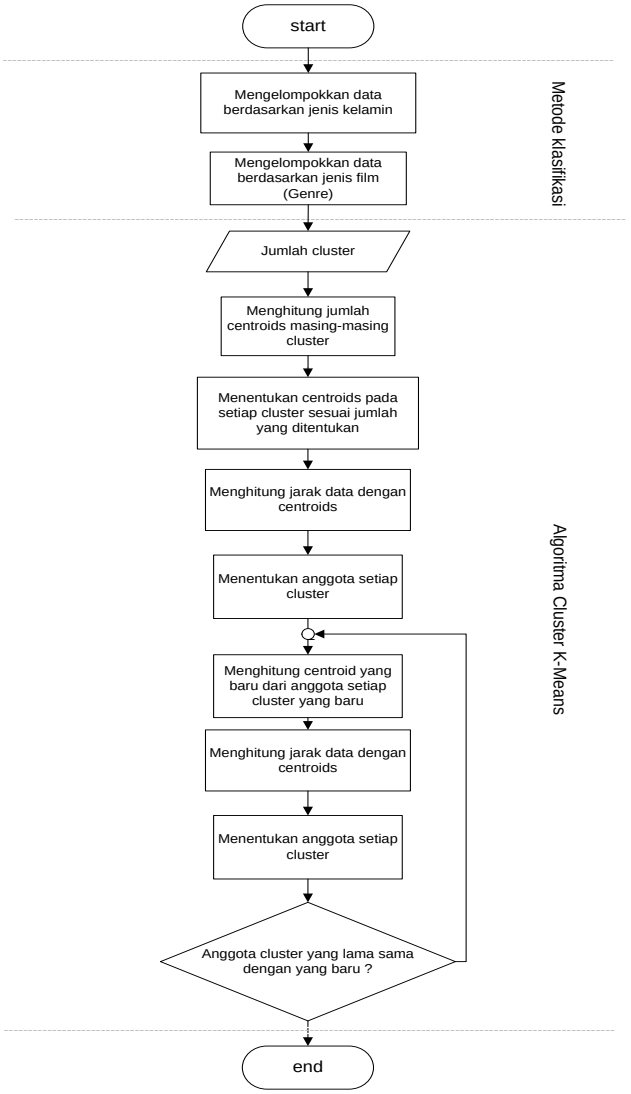
Langkah – langkah yang dilaksanakan pada penelitian ini adalah :

1. Klasifikasi / Pemecahan data
2. Clustering data
3. Perhitungan beda hasil.

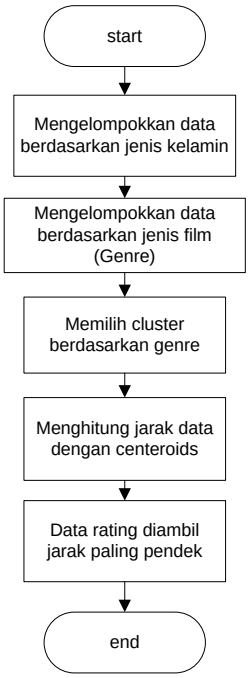
Langkah di atas digunakan untuk menentukan acuan data yang akan dilakukan pemeriksaan terhadap akurasi, sedangkan langkah untuk memeriksa akurasi data rating adalah:

1. Menentukan data dengan *decision tree*.
2. Menghitung titik terdekat dengan cluster
3. Menghitung beda.

Gambar 3 menunjukkan langkah–langkah membangun sistem rekomendasi. Pada gambar tersebut metode yang digunakan telah dibagi berdasarkan garis, yaitu metode klasifikasi dan metode *clustering*. Gambar 4 adalah flowchart yang digunakan untuk melakukan test pada data. Langkah yang digunakan pada test pada data adalah melakukan perbandingan antara rating pada data asli dengan rating yang dihasilkan sistem.



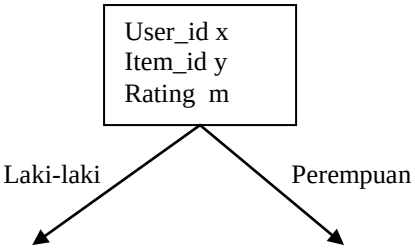
Gambar 3. Flowchart membentuk cluster



Gambar 4. Flowchart melakukan test pada data.

5.3 Klasifikasi Data

Langkah ini menggunakan metode *decission tree* yaitu data dipecah berdasarkan jenis kelamin, sehingga pada langkah awal didapatkan dua kelompok besar untuk data. Langkah awal ini juga digunakan ketika sistem ini digunakan untuk memberikan prediksi terhadap rating film, ketika sebuah data akan dilakukan prediksi tentang rating yang diberikan maka data tersebut diperiksa apakah user tersebut memiliki jenis kelamin laki-laki atau perempuan.



Gambar 5. *Decision tree* jenis kelamin.

Kemudian data dikelompokkan berdasarkan jenis filmnya. Jenis film tersebut didapatkan dari kode film pada tabel data. Sebuah film bisa memiliki lebih dari satu genre. Pada data utama ada genre 19 film yang terdaftar, *genre* film tersebut ada pada Tabel 1.

Tabel 1. Daftar *genre* film

1. Unknown	6. Comedy	11. Film-noir	16. Sci-fi
2. Action	7. Crime	12. Horror	17. Thriller
3. Advanture	8. Documentary	13. Musical	18. War
4. Animation	9. Drama	14. Mystery	19. Western
5. Children	10. Fantasy	15. Romance	

Sebuah film bisa memiliki lebih dari satu buah genre seperti pada Tabel 2 berikut.

Tabel 2. Contoh daftar film dan genre

No	Kode film	Judul film	Jenis Film(Genre)
1	Fm-01	Batman Forever	Action, advanture, crime, thriller
2	Fm-02	Bad Boys	Action, comedy
3	Fm-03	Braveheart	Action, drama, war
4	Fm-04	Die Hard	Action
5	Fm-05	Space Jam	Advanture, animation, children, fantasy

Dari Tabel 2 dilakukan ‘normalisasi’ sehinga didapatkan Tabel 3 berikut.

Tabel 3. Hasil pengolompokan berdasarkan *genre* film.

Jenis Film(Genre)	Kode Film	Judul Film
Action	Fm-01	Batman Forever
	Fm-02	Bad Boys
	Fm-03	Braveheart
	Fm-04	Die Hard
Advanture	Fm-01	Batman Forever
	Fm-05	Space Jam
Crime	Fm-01	Batman Forever
Thriller	Fm-01	Batman Forever
Comedy	Fm-02	Bad Boys
Drama	Fm-03	Braveheart
War	Fm-03	Braveheart
Animation	Fm-05	Space Jam
Children	Fm-05	Space Jam
Fantasy	Fm-05	Space Jam

Jika seorang user merating satu atau beberapa film, maka dicatat sebagai berikut:

Tabel 4. Data rating film.

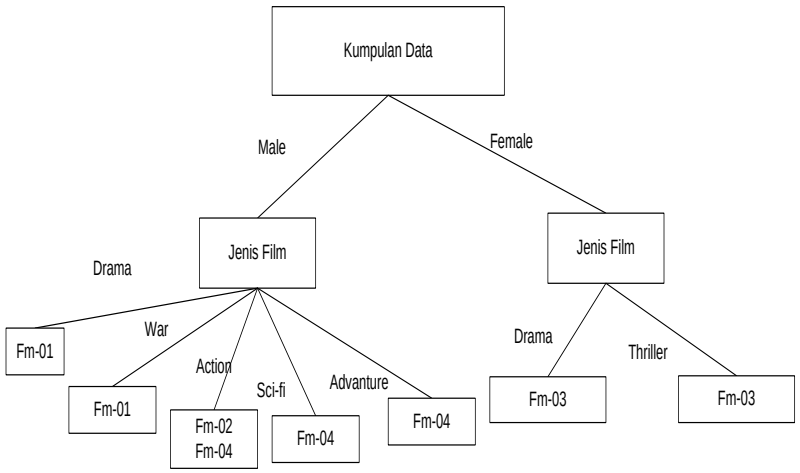
No	Kode User	Jenis Kelamin	Kode Film	Genre Film
1	U1	L	Fm-01	Drama, war
2	U2	L	Fm-02	Action
3	U3	P	Fm-03	Drama, thriller
4	U4	L	Fm-04	Action, Sci-fi, Advanture

Dari Tabel 3 dan Tabel 4 dihasilkan pengelompokan seperti pada Tabel 5 berikut.

Tabel 5. Data hasil pengelompokan

Jenis Kelamin	Jenis Film	Kode User	Kode Film
L	Drama	U1	Fm-01
	War	U1	Fm-01
	Action	U2	Fm-02
		U4	Fm-04
	Sci-fi	U4	Fm-04
	Advanture	U4	Fm-04
P	Drama	U3	Fm-03
	Thriller	U3	Fm-03

Berdasarkan Tabel 5, digambarkan *decision tree* sebagai berikut:



Gambar 6. Bagan *decision tree*

5.4. Clustering Data

Setelah didapatkan data hasil pengelompokkan berdasarkan jenis kelamin dan jenis film, langkah berikutnya adalah menentukan banyaknya *cluster* yang akan digunakan sebagai parameter algoritma *K-Mens* dalam membagi data. *Clustering* dilakukan pada masing–masing data yang telah dikelompokan berdasarkan jenis film, sedangkan untuk jumlah cluster yang dimaksudkan adalah jumlah *cluster* untuk seluruh data. Langkah pertama pada *clustring* data adalah dengan menentukan jumlah titik pusat (centroids) pada setiap *genre* film dimasing – masing *gender*. Perhitungan banyaknya titik pusat pada masing–masing genre adalah

$$\text{Jumlah centroids} = \frac{jumlah\_cluster}{seluruh\_data} \text{ x jumlah data per bagian.}$$

Jumlah titik pusat dimungkinkan untuk mengalami pembulatan sehingga banyaknya cluster hasil masukkan dapat bertambah. Berikut adalah contoh perhitungan dalam menentukan banyaknya centroids pada masing-masing genre:

```
Data keseluruhan = 300 buah data.
Jumlah cluster = 20

Perbandingan responden
1. Laki-laki : 200 data
2. Perempuan : 100 data.

Data gender laki-laki. Daftar genre yang dimiliki :
1. Action
2. Drama
3. Comedy
4. Thriller
```

```
5. Sci-fi

Perbandingan jumlah data pada jumlah genre
1. Satu genre ada 100 film
2. Dua genre ada 50 film
3. Tiga genre ada 50 film.

Pembagian data pada masing – masing genre
1. Action terdapat 100 data
2. Drama terdapat 40 data
3. Comedy terdapat 100 data
4. Thriller terdapat 40 data
5. Sci-fi terdapat 70 data

Data gender perempuan. Daftar genre yang dimiliki :
1. Action
2. Drama
3. Comedy
4. Thriller
5. Sci-fi

Pembagian data pada masing – masing genre
1. Action terdapat 20 data
2. Drama terdapat 80 data
3. comedy terdapat 30 data
4. thriller terdapat 30 data
5. Sci-fi terdapat 30 data
```

Dari data diatas maka dapat diketahui banyaknya data adalah 540, yaitu 350 pada data *gender* laki-laki dan 190 pada *gender* perempuan. Perhitungan jumlah titik pusat adalah sebagai berikut:

Data gender laki-laki

Action = $\frac{20}{540}$ x100 = 3.703 (dibulatkan 4.)	Drama = $\frac{20}{540}$ x40 = 1.48 (dibulatkan 2)	Sci-fi = $\frac{20}{540}$ x70 = 2.59 (dibulatkan 3)
Comedy = $\frac{20}{540}$ x100 = 3.703 (dibulatkan 4)	Thriller = $\frac{20}{540}$ x40 = 1.48 (dibulatkan 2)	

Data gender perempuan

Action = $\frac{20}{540}$ x20 = 0.703 (dibulatkan 1)	Drama = $\frac{20}{540}$ x80 = 2.96 (dibulatkan 3)	Sci-fi = $\frac{20}{540}$ x30 = 1.11 (dibulatkan 2)
Comedy = $\frac{20}{540}$ x30 = 1.11 (dibulatkan 2)	Thriller = $\frac{20}{540}$ x30 = 1.11 (dibulatkan 2)	

```
Jumlah centroids pada data jika dijumlahkan
total = 4+2+4+2+3+1+3+2+2+2 = 25
Jumlah cluster diupdate menjadi 25
```

Langkah selanjutnya adalah menentukan titik pusat (*centroids*) untuk setiap kumpulan data pada *genre* film sesuai dengan jumlah *centroids* yang telah dihitung. Titik pusat *cluster* itu dipilih dari kumpulan data dengan data yang telah dipecah dengan *gender* dan *genre* tertentu.

Langkah selanjutnya adalah menentukan *distance* data dengan *gender* dan *genre* tertentu dengan titik pusat cluster *tersebut*. Penentuan *distance* menggunakan *euclidean distance*.

$$d_{ij} = \sqrt{\sum_{k-1}^{n} (x_{ik} - x_{jk})^2} \qquad (2)$$

Data yang dipergunakan untuk perhitungan dalam menentukan *distance* antar titik data adalah kode film dan rating yang diberikan pada film tersebut. Data yang telah dihitung *distance* kemudian dicari nilai yang paling kecil pada cluster tertentu, jika titik *x* nilainya lebih kecil pada centoids *y* maka data tesebut masuk sebagai anggota cluster dengan centroids *y*. Parameter atau nilai yang digunakan dalam perhitungan adalah nilai umur dan rating. Langkah selanjutnya adalah penggabungan cluster, ini dilakukan apabila pada perhitungan *distance* terdapat nilai yang sangat dekat, dengan kata lain objek yang dihitung memiliki tingkat kesamaan yang tinggi, sehingga dapat dikatakan satu kelompok dan dilakukan penggabungan *cluster*.Penggabungan *cluter* ini akan terus dilakukan seiring dengan proses iterasi pada data sampai kondisi tertentu. Kondisi ini antara lain pada *K-Means* adalah sampai terbentuk jumlah cluster yang diinginkan dan tidak ada lagi *element* atau objek dari satu cluster berpindah kepada cluster yang lainnya.

Langkah selanjutnya adalah membandingkan anggota *cluster* yang lama dengan anggota *cluster* yang baru, jika anggota kedua *cluster* tersebut tidak sama maka dilakukan iterasi selanjutnya dengan mendefinisikan kembali *centroids*, dengan menjumlahkan seluruh data pada variabel yang sama, kemudian dibagi dengan banyaknya anggota.

$$\boldsymbol{c_2} = \frac{\sum_{x=0}^{n} data[umur]}{\sum_{x=0}^{n} x + 1} \qquad (3)$$

Hal yang dilakukan pada perhitungan untuk nilai rating, sehingga didapatkan nilai rata-rata untuk setiap parameter. Proses setelah mendapatkan *centroids* yang baru adalah menghitung kembali *distance* setiap data anggota *gender* dan *genre* tertentu ke *centroids* yang baru, kemudian ditentukan kembali anggota *cluster* dan dibandingkan dengan cluster yang lama jika sama maka iterasi dihentikan. Hasil dari cluster yang terakhir disimpan dalam database, digunakan sebagai acuan dalam melakukan analisa pada akurasi sistem rekomendasi.

5.6. Pengukuran Tingkat Akurasi

Tingkat keakurasian rekomendasi yang dihasilkan dapat diukur dengan menggunakan salah satu metode standar dalam statistika yang disebut dengan Mean Average Error atau MAE [2]. Secara mendasar, MAE menghitung kesalahan/ error absolut antara rating yang sebenarnya (p) dan rating hasil prediksi (q), semakin kecil nilai MAE yang didapat maka prediksi yang dihasilkan semakin akurat. Jika pengukuran dilakukan terhadap N data, maka MAE dapat dirumuskan dengan:

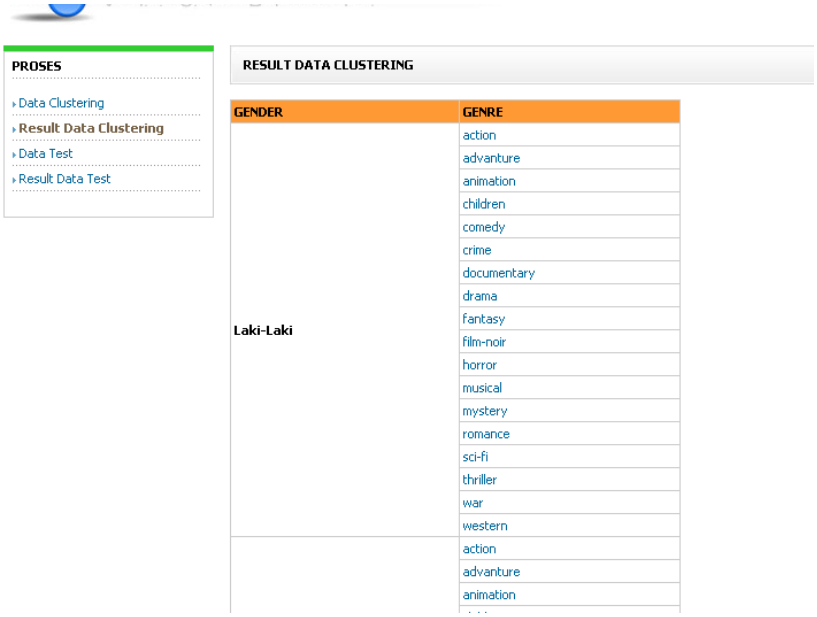
$$MAE = \sum_{i=1}^{N} \frac{|p_i - q_i|}{N} \qquad (4)$$

6. HASIL DAN PEMBAHASAN

Pada implementasi, bahasa pemrogaman yang digunakan adalah PHP dengan menggunakan web server apache.

6.1. Hasil Clustering

Pada proses ini keterangan yang muncul adalah jenis gender dan genre. Hasil yang ditampilkan adalah data centroids pada masing–masing cluster.



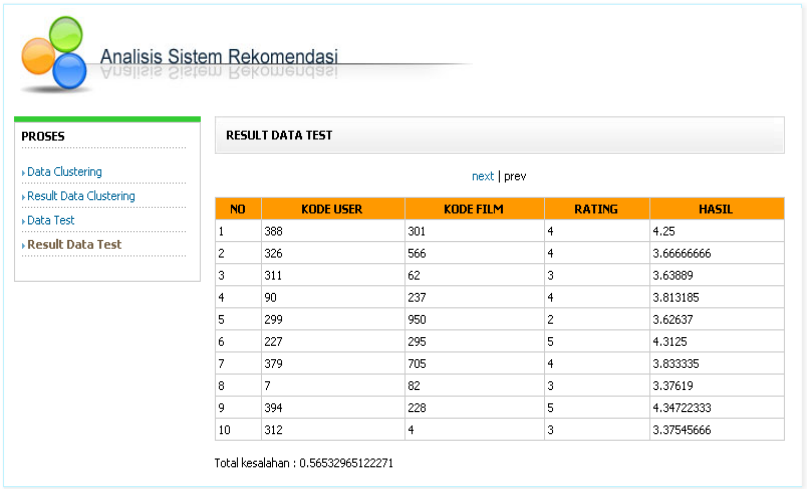
Gambar 7. Tampilan proses hasil clustering.

NO	UMUR	RATING
1	24.6667	3.16667
2	64	2.5
3	51.375	4.125
4	34	3.25
5	29	1
6	31.8571	4.42857
7	29.6667	4.66667
8	27	4.5
9	40.5	3.5
10	19.5	3.6

Gambar 8. Hasil centroids pada salah satu genre.

6.2. Hasil Pengujian

Pada hasil proses pengujian ini data ditampilkan dengan hasil rating yang didapatkan pada pengujian data sampel, sehingga dapat diketahui tingkat akurasi dengan menggunakan penggabungan ini.



Gambar 9. Tampilan hasil pengujian.

Berikut ini merupakan beberapa hasil pada 300 data sample yang diuji dengan beberapa jumlah cluster.

Tabel 6. Hasil rating

Jumlah Cluster	Deviasi / kesalahan (MAE)
51	0.75414836388646
112	0.71196101065502
150	0.53378495908297
210	0.48419914390279

Hasil yang didapatkan pada pengujian ini beda antara rating melalui *decision tree* dan algoritma *K-Means* dengan data asli rating rata – rata memiliki beda sebesar kurang dari 1, dari hasil tersebut ditemukan beberapa data dengan selisih beda yang agak besar hal ini kemungkinan disebabkan terlalu kecil dalam menentukan jumlah cluster. Pada Tabel 6 dapat diketahui bahwa semakin besar jumlah *cluster* yang diinputkan maka semakin kecil tingkat kesalahan dengan kata lain semakin besar tingkat akurasinya. Besarnya tingkat akurasi ditentukan dengan semakin banyaknya jumlah *cluster* karena semakin detail kelompok-kelompok data yang terbentuk. Pada saat pengujian ada kemungkinan suatu *cluster* tidak memiliki anggota atau titik data dikarenakan data-data tersebut memiliki perhitungan *distance* yang lebih kecil kepada *cluster* yang lain dari pada titik pusat cluster tersebut. Maka solusi dari masalah tersebut adalah dengan membagi anggota cluster yang memiliki anggota terbesar dengan cluster yang tidak memiliki anggota.

Pembagian anggota tersebut membuat perubahan pada titik pusat kedua *cluster* yang saling membagi anggota, maka dihitung kembali untuk titik pusat setiap cluster dengan keanggotaan titik data yang baru.

7. KESIMPULAN

Kesimpulan yang dapat diambil pada pengujian penelitian, adalah sebagai berikut:

1. Tingkat kesalahan yang kurang dari 1 mengindikasikan penggunaan metode *decision tree* dan algoritma K-Means clustering dapat digunakan sebagai salah satu metode dalam sistem rekomendasi dan tidak menutup kemungkinan untuk dikembangkan lebih lanjut.
2. Semakin besar banyaknya cluster, maka tingkat kesalahan semakin turun (tingkat akurasi naik).

DAFTAR PUSTAKA

[1] Mueller, C. 2005. **Data Clustering**. www.osl.iu.edu/~chemuell/new/oral-quals.php. Tanggal akses: 15 Agustus 2006.

[2] Scafer, J.B.; Konstan, J.A. dan Riedl, J. 2001. **Item-Based Collaborative Filtering Recommender Algorithms.** WWW10.

[3] Wikipedia. http://en.wikipedia.org/wiki/collaborative_filtering. Tanggal akses: 19 Agustus 2006.

[4] Wikipedia. **Personalisation**. http://en.wikipedia.org/wiki/collaborative_filtering. Tanggal akses: 19 Agustus 2006.