

SISTEM PEMILIHAN JADWAL MATAKULIAH BERBASIS MOBILE DENGAN METODE AGGLOMERATIVE NESTING DAN MULTI KRITERIA

Vera Tjahyono
Yetli Oslan, S.Kom., MT, Drs. Djoni Dwiyana, Akt., MT

Abstrak

Proses pemilihan jadwal matakuliah dapat menjadi hal yang sangat sulit dan rumit untuk dilakukan oleh mahasiswa karena pada umumnya mahasiswa melibatkan berbagai parameter di dalamnya. Parameter yang digunakan dapat meliputi waktu perkuliahan, dosen pengajar, jumlah sks maksimal yang dapat diambil, dan lain sebagainya. Masalah akan bertambah karena ketidaktahuan mahasiswa mengenai korelasi antara parameter-parameter tersebut dengan nilai yang akan diperolehnya. Hal ini menyebabkan seringkali mahasiswa salah mengambil keputusan dan akhirnya mendapatkan hasil yang kurang atau malah tidak maksimal.

Dalam hal ini, mahasiswa dapat terbantu dengan melihat data masa lalunya. Akan tetapi data masa lalu yang tidak sedikit dapat mempersulit proses analisa. Karenanya, data masa lalu tersebut perlu dikelompokkan terlebih dahulu. Proses pengelompokan data dapat dilakukan dengan banyak cara atau metode. Salah satunya adalah Agglomeratif Nesting (AGNES), yang digunakan dalam penelitian ini. Dalam mengelompokkan data, metode ini menggunakan dissimilarity matrix atau disebut juga matriks ketidaksamaan, yaitu mencari jarak minimum antar data. AGNES akan menghasilkan cluster yang berbentuk hierarki, mulai dari single object sampai menjadi single cluster.

Hasil pengelompokan tersebut ternyata dapat dijadikan sebagai pengetahuan untuk proses pengambilan keputusan. Karena hasil pengelompokan tadi adalah data-data di masa lalu yang telah terjadi, maka proses pengambilan keputusan pun akan menjadi lebih akurat. Salah satu metode pengambilan keputusan yang mudah dan cepat adalah perbandingan dengan multi kriteria, yang menggunakan sistem pembobotan kepada masing-masing pilihan yang tersedia.

Kata Kunci: *Clustering, Agglomerative Nesting, Multi Kriteria*

1. Pendahuluan

Pada setiap awal semester Universitas Kristen Duta Wacana (UKDW) selalu mengadakan registrasi mahasiswa. Mahasiswa dalam konteks ini adalah mahasiswa mulai semester ke dua karena mahasiswa baru UKDW telah diberikan jadwal kuliah yang diatur oleh program studi. Registrasi di UKDW sendiri dilakukan secara langsung di lab komputer UKDW dan juga secara *on line* pada sore hari yang sama. Pada registrasi inilah semua mahasiswa diijinkan untuk mendaftar untuk mengikuti matakuliah-matakuliah dengan jadwal yang dikehendaki, bahkan untuk melakukan pendaftaran cuti studi.

Meskipun berkesan bebas, sebenarnya dalam melakukan registrasi mahasiswa harus memperhatikan beberapa hal. Antara lain matakuliah-matakuliah apa saja yang ditawarkan untuknya semester itu, matakuliah apa saja yang wajib untuk semester itu, pra syarat masing-masing matakuliah yang ingin diambil, korelasi antar matakuliah, serta jumlah sks maksimal yang dapat diambil berdasarkan IPK dan IPSnya. Hal-hal tersebut merupakan hal yang sangat mendasar tapi justru sering diabaikan oleh mahasiswa UKDW dalam melakukan registrasi.

Biasanya mahasiswa kurang memperhatikan hal-hal di atas karena lebih mementingkan faktor-faktor yang dianggap lebih mendukung proses perkuliahan mereka. Beberapa faktor tersebut

adalah minat mahasiswa terhadap suatu jenis matakuliah, minat terhadap dosen pengajar, dan waktu kuliah. Sebagai contoh, sebagian mahasiswa akan menjadi enggan mengambil suatu matakuliah yang terlalu banyak hitungan matematika, terlalu menekankan pemrograman, diajar oleh dosen tertentu, berlangsung di hari Senin pukul 07.30, dan lain sebagainya.

2. Rumusan Masalah

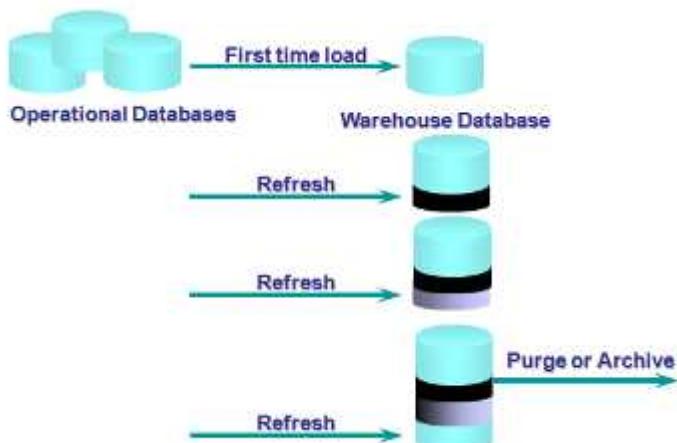
Memilihkan jadwal matakuliah bagi mahasiswa berdasarkan minat, dosen, dan waktu kuliah dengan metode Agglomerative Nesting dan Multi Kriteria yang dapat diakses oleh mahasiswa melalui perangkat bergeraknya.

3. Dasar Teori

3.1. Konsep Gudang Data

3.1.1. Pengertian Gudang Data

Menurut Bill Inmon, data warehouse merupakan koleksi data yang mementingkan penilaian atau pandangan pengguna, terintegrasi, memperhatikan waktu, dan tidak berubah untuk mendukung proses pengambilan keputusan manajemen. (Kimball, Ralph. 2004. Hal. 55)



Gambar 1. Daur hidup *data ware house*

Langkah-langkah dalam pemodelan data warehouse adalah:

- Bagaimana data diperoleh dan mendefinisikan aturan bisnis
- Membangun *dimensional model*
- Model-model resume
- Membangun *physical model*

3.1.2. Keuntungan dari penggunaan data warehouse

Keuntungan dari penggunaan *data warehouse*, antara lain adalah:

- Mengijinkan manajemen untuk menganalisa tren menggunakan data sebelumnya.
- Dapat mengakses data perusahaan secara luas.
- Melakukan analisis secara cepat.
- Dapat dilakukan pembersihan data.

3.1.3. Pengertian OLTP dan OLAP

Secara harafiah, OLTP (On-line Transaction Processing) dapat diartikan sebagai Sistem yang memproses suatu transaksi secara langsung melalui komputer yang terhubung dalam jaringan. Jadi setiap kali terjadi transaksi, data akan langsung tercatat dalam sistem ini. Contohnya adalah sistem transaksi penjualan dan pembelian. OLTP bersifat *data entry*, maksudnya adalah OLTP lebih cenderung membutuhkan inputan dan datanya belum dapat dibaca dengan baik karena belum diolah.

OLAP (On-line Analytical Processing) merupakan suatu proses sistem yang umumnya digunakan oleh pembuat kebijakan seperti analisis dan manajer, data yang didalam sebuah *data warehouse* umumnya berasal dari berbagai sistem OLTP dan memiliki karakteristik *data reading*, bukan *data entry*.

Syarat OLAP:

- a. Pandangan Multidimensional
- b. Transparan ke user
- c. Dapat diakses
- d. Pelaporan yang konsisten
- e. Arsitektur client-server
- f. Dapat membangkitkan dimensional
- g. Penanganan sparse matrix yang dinamis
- h. Mendukung Multiuser
- i. Operasi cross dimensi tanpa batasan
- j. Dapat dilakukan manipulasi data secara terus-menerus
- k. Pelaporan yang bersifat fleksibel
- l. Level dimensi dan fungsi agregasi yang tidak terbatas.

OLAP digunakan apabila terjadi kondisi seperti:

- a. Permintaan data untuk analisis, bukan transaksional
- b. Informasi yang dianalisis bukan pada titik awal pemasukkan ke perusahaan.
- c. Kalkulasi dan aggregat data dari level transaksi berbelit-belit
- d. Tipe utama elemen data yang dianalisis adalah numerik
- e. Pandangan cross-sectional dari data sering membutuhkan across-multiple-dimensions.
- f. Elemen yang identifikasi data menunjukkan relatif statis untuk sepanjang waktu.

3.1.4. Preprocessing

Sebelum suatu data dimasukkan ke dalam gudang data, pada umumnya dilakukan tiga langkah yaitu *extracting*, *transforming*, dan *loading* atau sering disebut dengan proses ETL. Proses ETL ini penting digunakan untuk mengintegrasikan sistem secara efektif, terutama untuk sistem-sistem yang memiliki perbedaan *Database Management Systems*, Sistem Operasi, perangkat keras, maupun protokol komunikasi. (Kimball, Ralph. 2004. Hal. 55)

Extracting adalah proses memetakan dan menangkap data mentah yang berasal dari sistem-sistem asal yang pada umumnya masih memiliki struktur sederhana. Biasanya pada tahap ini data akan berupa *flat file* atau tabel-tabel relasional. *Cleaning* adalah bagian proses *Transforming* melakukan pengecekan dan perbaikan terhadap nilai-nilai yang ada pada setiap sel tabel. Pada proses ini semua data yang akan diberi format yang sama dan dipastikan setiap selnya memiliki nilai agar data yang diperoleh menjadi valid. *Loading* merupakan proses memindahkan seluruh data yang telah diambil dari berbagai sumber dan telah dibersihkan ke dalam *data warehouse*. (Kimball, Ralph. 2005. Hal. 18-19)

3.2. Konsep Data Mining

3.2.1. Pengertian Data Mining

Data Mining adalah sekumpulan aktivitas untuk mengurutkan dan mengambil informasi khusus dari data yang tersimpan. Aktivitas data mining juga dapat dimanfaatkan untuk menemukan pola yang baru, tersembunyi, atau tak terduga. Untuk melakukan data mining, ada beberapa pendekatan atau algoritma, yaitu: (1) *classification*, (2) *association*, (3) *sequence*, (4) *cluster*.

Data mining telah banyak digunakan untuk menjawab berbagai pertanyaan dari para pembuat keputusan sehingga istilah data mining berkembang menjadi *knowledge data discovery* (KDD). Istilah ini dipakai untuk semua proses yang berhubungan dengan pencarian *knowledge* dari data agregasi. Data mining dapat menemukan pola penting dari kumpulan data dengan menggabungkan teknik-teknik seperti analisa statistik, neural dan logika fuzzy, analisa multidimensional, dan visualisasi data.

Cluster merupakan koleksi objek-objek data. Data-data yang mirip satu dengan lainnya berada dalam satu *cluster*, dan yang berbeda diletakkan pada *cluster* lainnya. *Cluster analysis* adalah proses pengelompokan satu set objek data ke dalam *cluster-cluster*. *Clustering* termasuk *unsupervised classification*, yaitu tidak ada pendefinisian awal kelas-kelasnya.

Metode *clustering* yang baik akan menghasilkan kualitas *cluster* yang tinggi dengan tingkat kemiripan yang tinggi dalam satu kelas dan tingkat kemiripan yang rendah antar kelas. Kualitas hasil *clustering* sangat tergantung pada ukuran kemiripan yang digunakan oleh metode dan implementasinya. Kualitas metode *clustering* juga diukur dari kemampuannya untuk menemukan beberapa atau semua pola yang tersembunyi.

Syarat-syarat yang harus dimiliki oleh *clustering* dalam data mining adalah:

- a. Scalability.
- b. Mampu menangani berbagai tipe atribut.
- c. *Discovery of clusters with arbitrary shape (not just spherical clusters)*
- d. *Minimal requirements for domain knowledge to determine input parameters (such as # of clusters)*
- e. Mampu menangani *noise* dan outliers.
- f. Tidak sensitive terhadap urutan inputan data.
- g. Incremental *clustering*
- h. *High dimensionality*
- i. Menyesuaikan aturan-aturan yang ditetapkan oleh pengguna.
- j. Mudah dipahami dan digunakan.

3.2.2. Pemanfaatan Data Mining

Aplikasi yang umum menggunakan *data mining* adalah:

- a. Pattern Recognition
- b. Spatial Data Analysis
 - Membuat peta tematik pada SIG dengan meng-*cluster*-kan fitur-fitur spasialnya.
 - Mendeteksi *cluster* spasial dan menjelaskan *cluster* tersebut dalam data mining spasial.
 - Contoh: penggunaan lahan, perencanaan kota, pembelajaran gempa bumi, dan lain-lain.
- c. Economic Science (riset pasar) – contoh: *marketing*, asuransi
- d. Image Processing
- e. WWW
 - Pengklasifikasian dokumen
 - Pengelompokan data Weblog untuk mencari kelompok dari pola akses yang sama.

3.2.3. Agglomerative Nesting

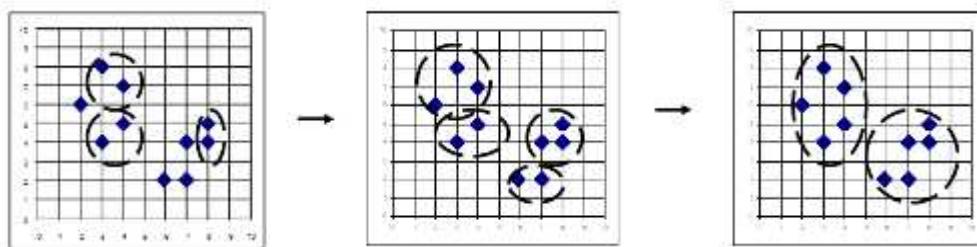
Agglomerative Nesting (AGNES) merupakan salah satu algoritma *clustering* yang diperkenalkan oleh Kaufmann dan Rousseeuw pada 1990. Algoritma ini menggunakan metode Single-Link dengan rumus pencarian jarak:

$$D_{sl}(C_i, C_j) = \min_{x,y} \{d(x, y) \mid x \in C_i, y \in C_j\}$$

dengan matriks ketidakmiripan:

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ : & : & : & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$

Inti dari algoritma ini adalah bagaimana menggabungkan titik-titik yang memiliki kemiripan paling tinggi. Terkadang semua titik dapat masuk ke *cluster* yang sama seperti berikut.



Gambar 2. *Clustering*

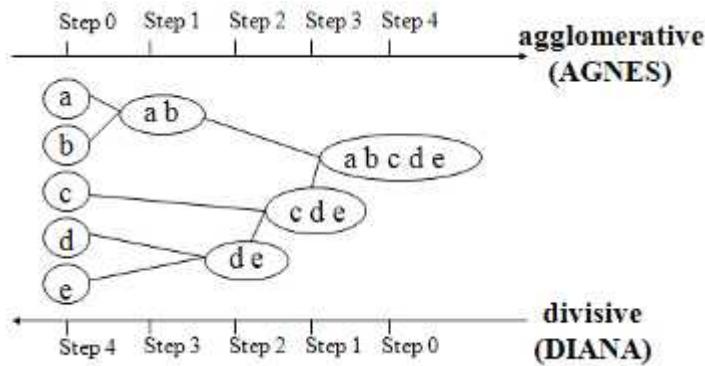
Secara sederhana, langkah-langkah algoritma Agglomerative Nesting dapat dirumuskan sebagai berikut:

- Menginisialisasikan semua titik (*node*) sebagai *cluster*.
- Menggunakan pendekatan *Single-link*. Single-link *clustering* (connectedness atau minimum method): jarak antara satu *cluster* dengan lainnya sama dengan jarak terdekat antara anggota *cluster* yang satu dengan anggota *cluster* lainnya. Jika data memiliki kemiripan, maka tingkat kemiripan satu *cluster* dengan lainnya sama dengan tingkat kemiripan tertinggi antara satu anggota *cluster* dengan anggota dari *cluster* lainnya.

Single-link digunakan untuk mengukur jarak antar *cluster* dengan rumus:

$$d_{\min}(C_i, C_j) = \min_{p \in C_i, q \in C_j} d(p, q)$$

- a. Masing-masing *cluster* diwakili oleh semua objek yang ada dalam *cluster* tersebut.
- b. Kemiripan antar *cluster* dilihat dari kemiripan pasangan titik data yang paling dekat. Dengan kata lain, suatu titik dapat berpindah *cluster* apabila titik tersebut lebih mirip atau “dekat” dengan titik yang ada pada lain.
- c. Mengulangi langkah 1 dan 2 sampai tidak ada lagi objek yang berpindah. Langkah ini tampak jelas pada gambar berikut.



Gambar 3. Ilustrasi Langkah Agglomerative dan Divisive

Kelemahan algoritma Agglomerative Nesting adalah tidak mengukur dengan baik, yaitu memiliki kompleksitas waktu paling tidak $O(n^2)$, dimana n adalah jumlah total obyek. Selain itu dengan algoritma ini tidak dapat mengulang kembali apa yang telah dilakukan diawal.

3.3. Multi Criteria Decision Analysis

Multi Criteria Decision Analysis adalah metode pencarian keputusan yang ditujukan untuk mendukung para pengambil keputusan yang menghadapi masalah kompleks atau berjumlah banyak. Metode ini dapat dikategorikan dalam berbagai cara, seperti berdasarkan modelnya (misal linear, non-linear, stochastic), berdasarkan karakteristik ruang keputusannya (misal infinite atau finite), atau berdasarkan proses pemecahannya (misal prior specification of preferences atau interactive).

Masalah yang dianalisa pada umumnya meliputi sejumlah alternatif untuk mendapatkan hasil keseluruhan berdasarkan kecocokan alternatif-alternatif tersebut terhadap sejumlah kriteria. Kriteria ini umumnya dibobotkan berdasarkan kepentingannya terhadap pembuat keputusan. Setelah dilakukan proses yang tepat, maka akan terbentuk suatu rangking alternatif berdasarkan kecenderungan pengambil keputusan.(Kenevissi, [Farhad](#). 2006)

4. Implementasi

4.1. Form tampilan utama

Program pemilihan ini dibuat dengan menggunakan Teknologi .NET khususnya Microsoft Visual Studio 2005 dengan bahasa pemrograman Visual C# .NET. Form Utama ini merupakan form pada sisi administrator yang menampilkan siapa saja dosen yang sering memberikan nilai baik dan siapa dosen yang paling sering dihadiri kuliahnya oleh mahasiswa dengan nim tertentu. Begitu pula dengan waktu kuliah mana yang paling ideal (mahasiswa sering mendapat nilai baik atau sering hadir). Data ini diperoleh dari hasil clustering yang dilakukan terhadap data pengambilan mahasiswa di masa lampau dengan metode *Agglomerative Nesting*.

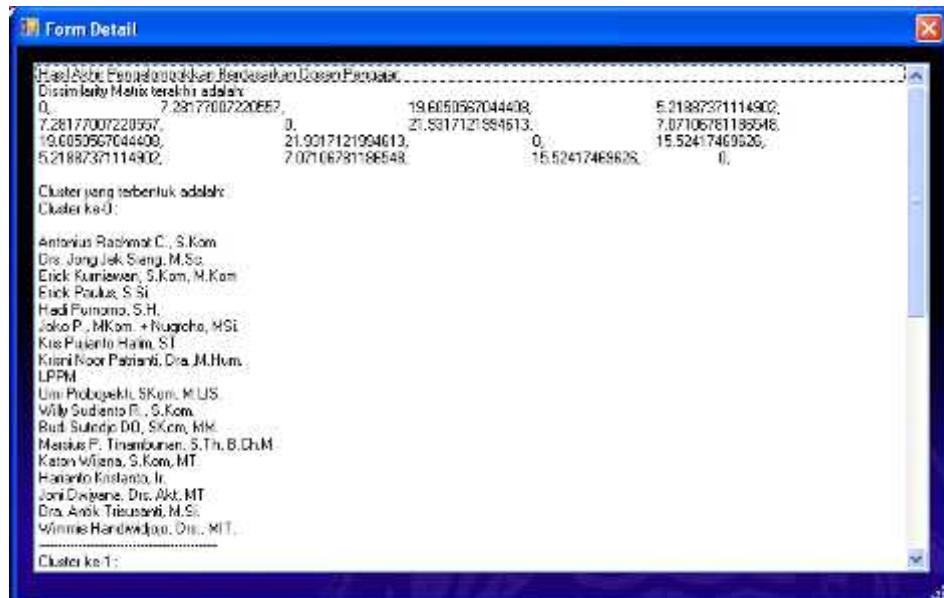


Gambar 4. Form Saran

Hasil ini diperoleh dengan cara membandingkan hasil pengelompokan *Agglomerative Nesting* yang berupa kelompok-kelompok nama dosen dan waktu kuliah dengan data nilai serta presensi masa lalu mahasiswa.

4.2. From Detail Clustering

Detail Clustering merupakan form yang menampilkan secara mendetail mengenai bentuk dan proses pengelompokan yang terjadi dengan metode Agglomerative Nesting. Oleh karena itu form ini menampilkan *dissimilarity matrix* yang terbentuk diakhir proses dan cluster-cluster apa saja yang terbentuk.



Gambar 5. Form Detail

4.3. Form Saran Jadwal Matakuliah

Form ini merupakan form utama pada sisi client, yaitu pihak mahasiswa dimana mahasiswa dapat mengaksesnya melalui perangkat bergeraknya (berbasis Windows Mobile). Pada form inilah mahasiswa akan mendapatkan saran jadwal matakuliah apa yang sebaiknya ia ambil pada saat registrasi. Hasil ini diperoleh dengan memberikan pembobotan terhadap setiap jadwal yang ditawarkan berdasarkan hasil pengelompokan dengan Agglomerative Nesting. Kemudian jadwal dipilihkan menggunakan metode Multi Kriteria.



Gambar 6. Form Hasil

5. Analisis

Penelitian ini melakukan pengujian terhadap 6 mahasiswa Sistem Informasi untuk 3 angkatan (2005-2007), maka jika hasil rekomendasi dibandingkan dengan pengambilan sesungguhnya akan diperoleh data sebagai berikut:

- | | | |
|--------------------|--------|----------------------------------|
| a. Cenderung Tepat | 2 kali | (keduanya angkatan 2005) |
| b. Kurang Tepat | 3 kali | (angkatan 2005, 2006, dan 2007) |
| c. Tidak Tepat | 1 kali | (angkatan 2006) |

Pemilihan jadwal matakuliah merupakan salah satu permasalahan pengambilan keputusan yang kompleks dan dipengaruhi oleh banyak criteria. Oleh karena itu hasil rekomendasi akan sangat sulit untuk dapat tepat 100 %. Dalam penelitian ini ada 2 kali rekomendasi yang cenderung tepat dan keduanya dilakukan terhadap mahasiswa angkatan 2005. Hal ini dimungkinkan karena jumlah kelas yang dibuka bagi mahasiswa Sistem Informasi 2005 masih sedikit dan data masa lalu yang dimiliki oleh mahasiswa 2005 tentu lebih banyak dibandingkan angkatan di bawahnya. Sehingga proses clusterisasi pun menjadi lebih valid.

Pada saat analisis ini dilakukan, data yang digunakan adalah masa registrasi untuk periode semester genap tahun ajaran 2007/2008, dimana pada masa tersebut mahasiswa angkatan 2006 tengah memasukki masa pemilihan bidang minat. Matakuliah yang ditawarkan pun sangat terpengaruh oleh pemilihan bidang minat mahasiswa. Rekomendasi akan menjadi lebih tepat apabila mahasiswa telah yakin dengan pilihan bidang minatnya. Inilah yang menyebabkan hasil rekomendasi banyak tidak atau kurang tepat bagi mahasiswa angkatan 2006. Demikian juga untuk mahasiswa angkatan 2007 yang baru akan memasukki semester ke dua. Data-data masa lampau yang dimiliki masih terlalu sedikit sehingga proses clusterisasi menjadi kurang valid.

6. Kesimpulan

Berdasarkan perancangan dan implementasi program, maka diperoleh kesimpulan sebagai berikut :

- a. Jumlah cluster yang ditetapkan dapat mempengaruhi kelengkapan data. Jika jumlah cluster yang ditetapkan banyak (biasanya jika lebih dari 6 buah), akan ada beberapa kelompok data yang seolah tidak terpakai. Hal ini disebabkan oleh sedikitnya parameter yang diberikan untuk dosen dan waktu, sehingga untuk membedakan mereka kedalam cluster-cluster yang berbeda akan sangat sulit.
- b. Pemilihan jadwal akan menjadi kurang tepat jika dilakukan pada mahasiswa dengan nilai-nilai yang relatif seragam, misalnya selalu dapat A atau selalu dapat D. Karena pada kasus seperti ini, pembobotan berdasarkan minat terhadap dosen dan jam kuliah (dilihat dari nilai dan jumlah kehadiran) akan menjadi sama semua.

7. Daftar Pustaka

- _____, A Comparison in Clustering Validation Techniques, Tanggal akses 27 Juni 2008, <http://grad.uprm.edu/tesis/gonzaleztoledo.pdf>
- _____, AHP-Tutor: an Intellegent Tutoring System for AHP, Tanggal akses 8 September 2008, <http://www.wwz.unibas.ch/wi/projects/AHP/AHP.htm>
- _____, Descriptive Modeling, Tanggal akses 27 Juni 2008, <http://stat.rutgers.edu/~madigan/588/chap9.ppt>
- _____, Knowledge Management Techniques, Module 2: Data Mining, Tanggal akses 27 Juni 2008, www.doc.ic.ac.uk/~yg/course/ida2002/ida-2002-6.ppt

- _____, Machine Learning Introduction, Tanggal akses 26 Juni 2008,
http://webcourse.cs.technion.ac.il/236756/Spring2007/ho/WCFfiles/clustering_handout.pdf
- _____, Multi-Criteria Decision Making, Tanggal akses 14 Januari 2009,
<http://www.edc.ncl.ac.uk/highlight/rhdecember2006.php/>
- Berson, A., Smith S., and Thearling, K., 1999, *Building Data Mining Applications for CRM*. McGraw-Hill.
- Giudici, Paolo., 2003, *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley & Sons: England.
- Han, J. and Kamber, M., 2001, *Data Mining: Concepts and Techniques*. Morgan Kaufmann: New York.
- Marakas, George M., 1999, *Decision Support Systems in the 21st Century*. Prentice Hall.
- Ralph, Kimball, 2002, *The Data Warehouse Toolkit*. John Wiley & Sons Inc: New York.
- Turban, Efraim, Jay E. Aronson, 2001, *Decision Support Systems and Intelligent Systems* 6th edition. Prentice Hall: New Jersey.