

ANALISA PENGGUNAAN NILAI MATA KULIAH UNTUK CLUSTER MAHASISWA PERGURUAN TINGGI DENGAN MENGUNAKAN EM CLUSTERING

Andy Prasetyo Utomo¹

ABSTRAK

Nilai mahasiswa dapat digunakan sebagai acuan dalam menentukan atau menyarankan mata kuliah yang diambil sesuai dengan kemampuan atau potensi mahasiswa dengan melakukan clustering. Mahasiswa dengan nilai baik dapat diarahkan untuk mengambil tipe pendidikan intensif dan sebaliknya. Selain nilai mahasiswa, mata kuliah juga mengalami clustering kedalam kelompok – kelompok jenis keilmuan, sehingga diharapkan mahasiswa mendapatkan kelompok keilmuan yang tepat sesuai dengan kemampuannya masing – masing. Dengan itu perguruan tinggi dapat menciptakan mahasiswa yang berkualitas baik. Dalam makalah ini, teknik clustering yang digunakan adalah EM algoritim dan menghasilkan 6 langkah pembuatan cluster.

Kata kunci : WEKA, clustering, EM algoritim, nilai mahasiswa, kelompok keahlian, prediksi mata kuliah..

ABSTRACT

Grades can be used as a reference in determining or suggest courses taken in accordance with the ability or potential students by performing clustering. Students with good grades can be directed to take the type of intensive education and vice versa. In addition to the students, the courses also have clustering into groups of the kind of science, so that students are expected to get a proper scientific groups according to their abilities. With that college students can create a good quality. In this paper, the clustering technique used was EM algoritim and produces 6 step cluster creation.

Keywords: Weka, clustering, EM algoritim, the student, group skills, prediction of course

I. PENDAHULUAN

Saat ini persaingan dalam mencari kerja telah memaksa perguruan tinggi untuk menciptakan mahasiswa yang berkualitas baik sehingga berbeda dengan para pesaing lainnya dengan memberikan kompetensi tertentu bagi mahasiswanya. Perguruan tinggi memahami bahwa mahasiswa pada dasarnya berasal dari lingkungan yang berbeda-beda, sehingga mereka memerlukan pelayanan yang berbeda agar bisa memberdayakan potensi terbesarnya.

¹ Dosen Fakultas Teknik Universitas Muria Kudus

Dengan makin berkembangnya pemanfaatan teknologi sistem informasi intelijen memberikan peluang bagi perguruan tinggi untuk mempelajari perilaku mahasiswa dan menggunakannya sebagai dasar dalam pengambilan kebijakan.

Clustering digunakan untuk mengelompokkan suatu obyek berdasarkan kemiripan yang dimiliki suatu obyek sehingga setiap cluster diharapkan memiliki perbedaan yang signifikan.

Dalam makalah ini, teknik clustering yang digunakan yaitu EM algoritma yang ada pada WEKA. Data yang digunakan merupakan simulasi data mahasiswa secara acak yang dipunyai oleh program studi Sistem Informasi Universitas Muria Kudus.

Dalam makalah ini diperkenalkan cara mensegmentasi mahasiswa di perguruan tinggi berdasarkan nilai permatakuliahnya. Hasil segmentasi tersebut digunakan untuk melihat potensi mahasiswa, yang akan dijadikan acuan untuk:

1. Memberikan pembelajaran yang berbeda terhadap segmentasi mahasiswa yang bagus pada suatu kelompok ilmu tertentu, namun lemah dalam kelompok ilmu yang lain
2. Memberikan panduan bagi dosen wali untuk membantu mahasiswa menentukan matakuliah pilihan yang cocok dengan mahasiswa tersebut
3. Memberikan panduan bagi dosen pembimbing untuk memandu mahasiswa mengarahkan fokus tugas akhir yang akan dibuat mahasiswa
4. Mendeteksi dugaan adanya kecurangan akademik
5. Evaluasi teknik penilaian dosen

II. LANDASAN TEORI

2.1 WEKA

WEKA dikembangkan oleh universitas Waikato di New Zealand, dan dibuat dengan menggunakan bahasa java serta didistribusikan dalam GNU *general public license*. WEKA dapat dijalankan dalam sistem operasi Linux, Windows dan Macintosh.

WEKA menyediakan implementasi dari algoritma pembelajaran yang mudah diterapkan dalam dataset. WEKA dapat menyelesaikan berbagai metode dalam permasalahan dalam data mining yaitu : *regression*, klasifikasi, *clustering*, *association rule mining* dan pemilihan atribut.

Salah satu cara penggunaan WEKA adalah untuk menerapkan metode pembelajaran sebuah dataset dan menganalisa outputnya untuk mempelajari mengenai data lebih jauh lagi serta dapat digunakan untuk model pembelajaran untuk menghasilkan prediksi pada obyek yang baru. Dalam paper ini WEKA yang digunakan adalah WEKA GUI 3.6.1

2.2 Cluster

Teknik cluster merupakan teknik yang banyak digunakan dalam data mining. Tujuan dari metode cluster adalah pengelompokan sejumlah data/obyek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang memiliki karakteristik yang mirip. Dalam cluster diusahakan menempatkan obyek yang mirip (jaraknya dekat) dalam satu cluster dan membuat jarak antar cluster sejauh mungkin.

Clustering termasuk teknik *unsupervised learning*, dimana dalam hal ini tidak perlu melatih metode tersebut, sebaliknya *supervised learning* adalah metode yang memerlukan training dan testing.

Ada dua pendekatan dalam clustering yaitu partitioning dan hirarki. Dalam partitioning mengelompokkan obyek x_1, x_2, \dots, x_n kedalam k cluster, ini dilakukan dengan menentukan pusat cluster awal lalu dilakukan realokasi obyek berdasarkan kriteria tertentu sampai dicapai pengelompokan yang optimum. Dalam cluster hirarki dimulai dengan membuat m cluster dimana setiap cluster beranggotakan satu obyek dan berakhir dengan satu cluster dimana anggotanya adalah m obyek

2.3 EM Algorithm

EM algorithm adalah pendekatan yang digunakan secara luas untuk mempelajari keadaan dari variabel yang tidak diamati. EM algorithm dapat digunakan untuk variabel yang memiliki nilai yang tak pernah secara langsung diamati, menyediakan bentuk umum dari distribusi probability yang mengstur variabel yang dikenal. EM algorithm juga menjadi landasan untuk algoritma cluster *unsupervised*.

III. PEMBUATAN CLUSTER

Simulasi data nilai mata kuliah yang telah didapatkan secara acak kemudian diatur dan dimasukkan ke dalam format / column yang telah ditentukan.

3.1 Data

Data terdiri dari field sebagai berikut:

Tabel 1. Daftar Format / Kolom yang telah ditentukan

No	Nama	Deskripsi	Tipe data	Keterangan
1	nim	Nomor induk mahasiswa	Varchar	
2	kodematakuliah	Kode matakuliah	Varchar	
3	thajar	Tahun ajar	Integer	
4	ganjil	Semester ganjil / genap	Boolean	
5	nilai	Nilai mahasiswa	Char	A, B, C, D, E
6	kodelokal	Kode lokal matakuliah	Varchar	
7	nama	Nama matakuliah	Varchar	
8	kelompok	Kelompok matakuliah	Varchar	MPK, MKB, MPB, MKK
9	statusmatakuliah	Status Matakuliah pilihan / wajib	Boolean	
10	prasyarat1	Kode matakuliah prasyarat 1	Varchar	
11	prasyarat2	Kode matakuliah prasyarat 2	Varchar	
12	sks	Jumlah sks	Byte	
13	semester	Semester matakuliah ditawarkan	Byte	
14	pengampu	Pengampu matakuliah, Jurusan, atau program studi	Varchar	

3.2 Pengkategorian Mata Kuliah

Sebagaimana telah dijelaskan sebelumnya, matakuliah yang ada harus dikategorikan kepada beberapa jenis matakuliah. Jenis matakuliah ini dititikberatkan pada kedalaman penyampaian materi. Matakuliah yang ada dikategorikan sebanyak 5 kategori yakni

keterampilan, teori, logika, bahasa, dan hitungan. Matakuliah keterampilan adalah matakuliah yang lebih menekankan pada pelaksanaan praktek matakuliah tersebut, baik praktek menggunakan aplikasi tertentu maupun praktek ke lapangan. Matakuliah teori adalah matakuliah yang bersifat hafalan, yang berisi konsep, sejarah, dan jenis hafalan lainnya. Matakuliah logika adalah semua matakuliah yang berhubungan dengan pemrograman, dan pemanfaatan logika dalam pemrograman. Matakuliah bahasa adalah matakuliah yang terkait dengan kemampuan berbahasa, sedangkan matakuliah hitungan adalah matakuliah yang menggunakan model matematis untuk pemecahan masalah. Kategori ini kemudian ditambahkan ke dalam setiap record dari kolom yang telah ditentukan sebelumnya.

3.3 Pembuatan Cluster

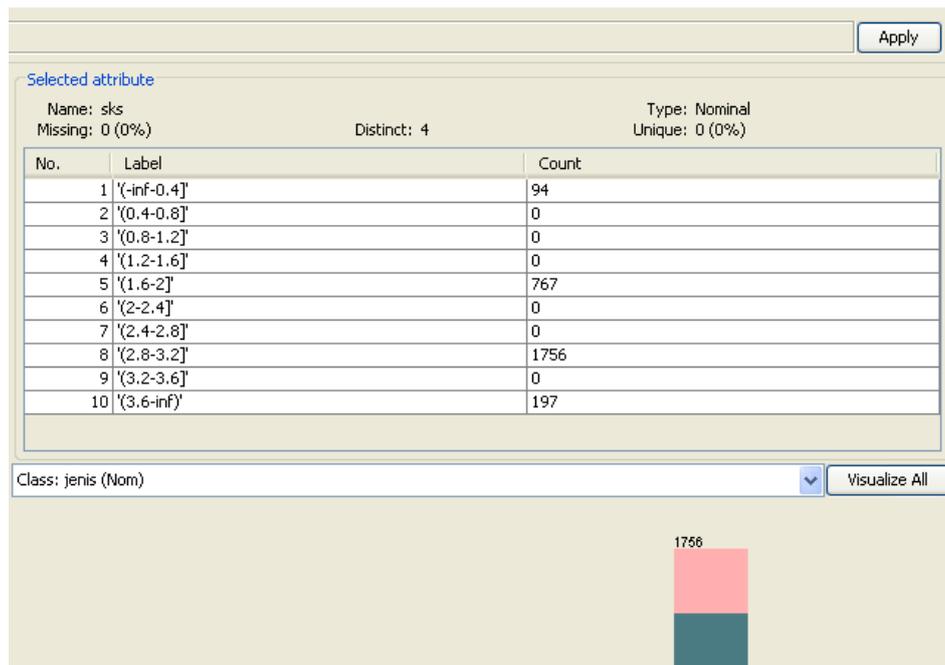
Proses pembuatan cluster dilakukan dengan menggunakan EM-Algorithm. Untuk mendapatkan cluster yang optimal dilakukan percobaan sebagai berikut:

- a. Menggunakan EM Algorithm untuk mencluster 15 field yang ada menggunakan mode tes: *evaluate on training data*.

Pada percobaan ini didapatkan 10 cluster dengan *likelihood* sebesar -14.77986. Meskipun tidak memiliki data pencilan, tetapi cluster yang dihasilkan belum benar, karena data-data yang bersifat angka seperti sks, semester tidak berbentuk diskrit. Sehingga dihitung *mean* dan standar deviasinya. Ini kemungkinan membuat cluster menjadi kacau, oleh karena itu data ini perlu didiskritkan dulu. Bahkan, beberapa tipe data seperti nim, tahun ajaran, ganjil, dan status yang bertipe varchar namun berisi data angka, juga dijadikan tipe data yang bukan diskrit.

- b. Menggunakan EM Algorithm untuk mencluster 15 field yang ada menggunakan mode tes: *evaluate on training data* dengan mendiskritkan tipe data angka.

Di dalam WEKA ada fasilitas untuk memfilter dengan mengubah tipe data angka menjadi diskrit dengan fitur discretize, sayangnya tipe data yang diubah akan dijadikan tipe data range (Gambar 1). Sedangkan diskrit yang dimaksud adalah mengelompokkan data kedalam kelompok yang bukan range, seperti untuk sks, data harus didiskritkan menjadi 2,3, dan 4 Sks. Tahun ajaran harus didiskritkan menjadi 2005, 2006, 2007, dan 2008. Oleh karena itu dilakukan pendiskritan data secara manual dengan menambahkan karakter huruf di depan angka sehingga WEKA bisa mendeteksi tipe data tersebut bertipe string dan didiskritkan.



Gambar 1. *Discretize* menggunakan WEKA

Setelah data selesai didiskritkan, kemudian dilakukan clustering dengan menggunakan EM-Algorithm dan dihasilkan 38 cluster. Dengan *likelihood* = -13.61578. Data diskrit membuat proses *cross validation* yang dilakukan WEKA untuk melakukan EM-Algorithm membuat jumlah cluster semakin banyak. Tentu saja cluster ini tidak juga bisa digunakan.

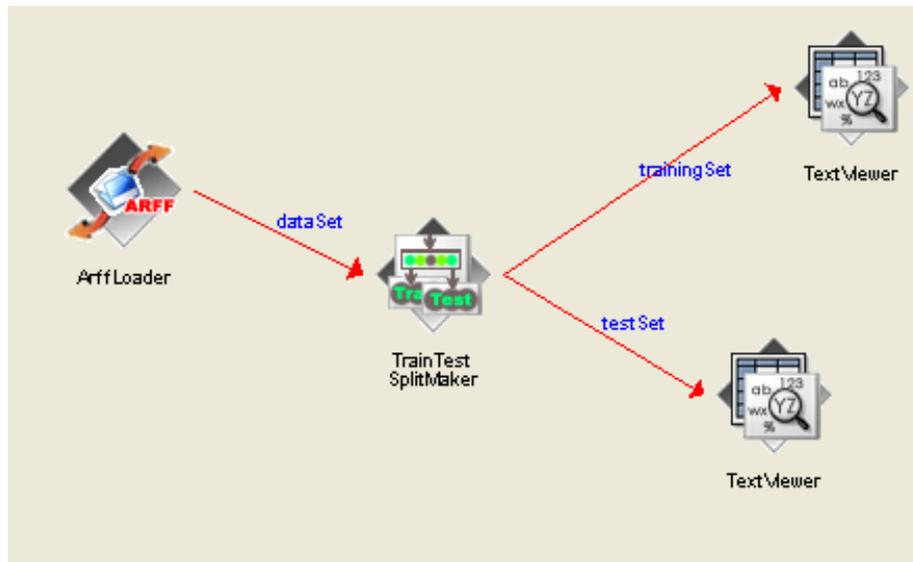
- c. Menggunakan EM Algorithm untuk mencluster 4 field yang ada menggunakan mode tes: *evaluate on training data*.

Percobaan berikut dalam membuat cluster adalah dengan menghilangkan field yang tidak dibutuhkan. Berdasarkan analisa sederhana field seperti nim, nilai, nama, dan jenis adalah field yang paling layak untuk diclusterkan. Field yang lain diabaikan karena dipandang tidak dibutuhkan untuk melakukan analisa.

Dari hasil clustering terhadap ke 4 field ini menghasilkan 9 cluster dengan *likelihood* sebesar -9.60741. Sayangnya terdapat 2 cluster yang kosong. Sehingga model ini belum bisa digunakan.

- d. Menggunakan EM Algorithm untuk mencluster 4 field yang ada menggunakan mode tes: *user supplied test set*.

Pertama-tama data yang sebanyak 2.814 record dipecah menjadi *training set* dan *test set* menggunakan WEKA *KnowledgeFlow Environment* (Gambar 2). Sehingga didapatkan 1.857 training data dengan 957 test data.



Gambar 2. Membagi *testSet* dengan *trainingSet* dengan WEKA *KnowledgeFlow Environment*

Dari langkah terakhir ini didapatkan 8 cluster dengan *likelihood* -9.60741. Dan tidak ada cluster yang persentase kosong, sehingga cluster ini dipandang layak

IV. ANALISIS HASIL DAN PEMBAHASAN

Cluster 1 adalah cluster yang lumayan besar (16%) yang merupakan cluster terbesar ke dua dari keseluruhan cluster. Cluster 1 ini dapat dicirikan dengan dominasi dari jenis mata kuliah logika daripada jenis mata kuliah lainnya. Selain itu distribusi nilai pada cluster ini di dominasi oleh nilai “B” dengan 68% yang kemudian disusul dengan nilai “A” dengan 29% sedangkan nilai lainnya persentasenya sangat kecil. Jumlah mahasiswa yang berada di cluster ini lumayan tinggi yaitu 14%. Mata kuliah sistem basis data dan Algoritma Pemrograman adalah mata kuliah yang mendominasi di cluster ini.

Cluster 2 adalah cluster dengan ukuran yang cukup signifikan (16%) yang merupakan cluster terbesar ke tiga dari keseluruhan cluster. Cluster 2 ini dapat dicirikan dengan dominasi dari jenis mata kuliah bahasa daripada jenis mata kuliah lainnya. Selain itu distribusi nilai pada cluster ini di dominasi oleh nilai “B” dengan 56% yang kemudian disusul dengan nilai

“C” dengan 23% dan nilai “A” dengan 17% sedangkan nilai lainnya persentasenya sangat kecil. Jumlah mahasiswa yang berada di cluster sama dengan cluster 1 yaitu 14%. Mata kuliah Bahasa Inggris adalah mata kuliah yang mendominasi di cluster ini.

Cluster 3 adalah cluster dengan ukuran yang lumayan kecil (7%) yang merupakan cluster terkecil ke dua dari keseluruhan cluster. Cluster 3 ini dapat dicirikan dengan dominasi dari jenis mata kuliah teori daripada jenis mata kuliah lainnya. Selain itu distribusi nilai pada cluster ini di dominasi oleh nilai “C” dengan 72% yang kemudian disusul dengan nilai “D” dengan 10% dan nilai “E” dengan 15% sedangkan nilai lainnya persentasenya sangat kecil. Jumlah mahasiswa yang berada di cluster ini cukup kecil bila dibandingkan dengan cluster yang lain yaitu hanya 9%. Mata kuliah Manajemen Umum, Manajemen Proyek dan Sejarah Teori Organisasi Umum adalah mata kuliah yang mendominasi di cluster ini.

Cluster 4 adalah cluster dengan ukuran kecil (6%) yang merupakan cluster terkecil dari keseluruhan cluster. Cluster 4 ini dapat dicirikan dengan dominasi dari jenis mata kuliah logika daripada jenis mata kuliah lainnya. Selain itu distribusi nilai pada cluster ini di dominasi oleh nilai “C” dengan 84% yang kemudian disusul dengan nilai “D” dengan 10% sedangkan nilai lainnya persentasenya sangat kecil. Jumlah mahasiswa yang berada di cluster ini paling kecil bila dibandingkan dengan cluster yang lain yaitu hanya 8%. Mata kuliah Pemrograman Database 1 dan 2 serta Pemrograman Berorientasi Objek adalah mata kuliah yang mendominasi di cluster ini.

Cluster 5 adalah cluster dengan ukuran sedang yaitu (14%). Cluster 5 ini dapat dicirikan dengan dominasi dari jenis mata kuliah hitungan daripada jenis mata kuliah lainnya. Pada cluster 5 distribusi nilai lebih variatif hanya nilai “E” yang nilainya terlalu kecil sedangkan nilai lainnya mempunyai jumlah yg signifikan dengan nilai terbanyak adalah “C” dengan 38%. Jumlah mahasiswa yang berada di cluster ini adalah 13% yang merupakan jumlah yang sedang bila dibandingkan dengan cluster yang lain. Mata kuliah Matematika Diskrit, Statistik Dasar dan Sistem Informasi Akuntansi adalah mata kuliah yang mendominasi di cluster ini.

Cluster 6 adalah cluster dengan ukuran sedang yaitu (13%). Cluster 6 ini dapat dicirikan dengan dominasi dari jenis mata kuliah teori sama dengan cluster 3 tetapi dengan jumlah lebih besar daripada cluster 3. Pada cluster 6 distribusi nilai di dominasi dengan nilai A yang mencapai 98%. Jumlah mahasiswa yang berada di cluster ini sama dengan cluster 5

yaitu 13%. Matematika Diskrit, Statistik Dasar dan Sistem Informasi Akuntansi adalah mata kuliah yang mendominasi di cluster ini.

Cluster 7 adalah cluster dengan ukuran sedang yaitu (13%) sama dengan cluster 6. Cluster 7 ini dapat dicirikan dengan dominasi dari jenis mata kuliah ketrampilan. Pada cluster 7 distribusi nilai di dominasi dengan nilai B yang mencapai 43% kemudian disusul dengan nilai A (40%) dan nilai C (15%) sedangkan nilai lainnya tidak terlalu signifikan jumlahnya. Jumlah mahasiswa yang berada di cluster ini sama dengan cluster 5 dan cluster 6 yaitu 13%. Mata kuliah Aplikasi Multimedia, Pengantar Jaringan Komputer adalah mata kuliah yang mendominasi di cluster ini.

Cluster 8 adalah cluster dengan ukuran terbesar yaitu (17%). Cluster 8 ini dapat dicirikan dengan dominasi dari jenis mata kuliah teori sama dengan cluster 3 dan 6 tetapi dengan jumlah yang lebih tinggi diantara dua cluster tersebut. Pada cluster 8 distribusi nilai di dominasi dengan nilai B yang mencapai 98%. Jumlah mahasiswa yang berada di cluster ini sama adalah yang paling banyak yaitu mencapai 15%. Mata kuliah Manajemen Umum, Konsep Sistem Informasi dan Sistem Operasi adalah mata kuliah yang mendominasi di cluster ini.

V. Kesimpulan

5.1 Kesimpulan

Dari uraian sebelumnya dapat ditarik kesimpulan sebagai berikut:

- a. Penggunaan WEKA untuk mengcluster nilai mahasiswa permatakuliah dapat dilakukan dengan algoritma EM clustering dapat dijadikan pedoman untuk memberikan pembelajaran yang berbeda kepada mahasiswa.
- b. Mode tes *user supplied test set* untuk EM algorithm lebih cocok diterapkan untuk clustering data nilai mahasiswa dibandingkan dengan mode *tes evaluate on training data*.
- c. Langkah-langkah untuk mengclustering data nilai mahasiswa dengan menggunakan algoritma EM adalah sebagai berikut:
 - 1) Membersihkan data dari data yang hilang / NULL
 - 2) Konvensi penamaan
 - 3) Diskretisasi data yang dibutuhkan
 - 4) Pemilihan atribut yang dibutuhkan

- 5) Buat training set dan tes set
- 6) Lakukan EM Clustering dengan menggunakan WEKA dengan mode tes *user supplied test set*

5.2 Saran

Dari penelitian yang telah dilakukan dapat diberikan saran sebagai berikut:

- a. Menyediakan otomasi langkah clustering data nilai mahasiswa dengan program aplikasi memanfaatkan class-class java yang terdapat di WEKA.
- b. Membuat algoritma clustering untuk otomasi pengkategorian matakuliah yang sebelumnya dilakukan secara manual. Clustering bisa dilakukan berdasarkan persentase materi yang digunakan dari data silabus matakuliah.

DAFTAR PUSTAKA

- Andina Budiarti, Yudho Giri Sucahyo, Yova Ruldeviyani (2006). Studi Karakteristik Kelulusan Peserta Didik dengan Teknik Clustering. Seminar Nasional Sistem dan Informatika 2006
- Andina Budiarti (2006). Aplikasi dan analisis clustering pada data akademik. Skripsi, Universitas Indonesia, Fakultas Ilmu Komputer
- Cai Wu, Jeffrey R. Steinbauer, dan Grace M. Kuo (2005). EM Clustering Analysis of Diabetes Patients Basic Diagnosis Index. AMIA Annu Symp Proc. 2005
- Carlo Vercellis (2009). Business Intelligence Data Mining and Optimization for Decision Making. Indianapolis: Wiley Publishing
- Cooley, R (2000). Web Usage Mining: Discovery and Application of Interesting patterns from Web Data. Ph. D. Thesis, University of Minnesota, Department of Computer Science,
- Ian H. Witten dan Eibe Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publication.
- ITT Telkom (2007). Index of /datamining-tools/WEKA/e-book WEKA. <http://ittelkom.ac.id/dmc/datamining-tools/WEKA/e-book%20WEKA/> diakses pada 26 April 2010

Mitchell, T.M (1997). Machine Learning. Boston: McGraw-Hill.

Remco R. Bouckaert et al (2009). WEKA Manual: for Version 3-6-1. New Zealand:
University of Waikato

Santoso, B (2007). Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis.
Yogyakarta: Graha Ilmu

Zdravko Markov. Using WEKA 3 for clustering.
http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-Ex3.html diakses pada 28
April 2010