

Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means

Ginanjari Abdurrahman¹⁾

¹⁾Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121
Email: ¹⁾abdurrahmanginanjari@unmuhjember.ac.id

ABSTRAK

Ilmu pengetahuan dan teknologi akan mempermudah pekerjaan manusia. Namun, disisi lain akan meningkatkan persaingan. Dalam menghadapi ketatnya persaingan, perlu dipersiapkan sumber daya manusia yang kompeten. Mahasiswa diharapkan siap secara akademik, berupa kesiapan pengetahuan maupun keterampilan untuk menghadapi persaingan yang semakin ketat. Salah satu cara untuk melihat kompetensi mahasiswa adalah dengan melihat hasil belajar yang dapat direpresentasikan dengan nilai ujian yang ditempuh. Ujian tengah semester (UTS) merupakan salah satu bentuk ujian yang menjadi komponen penilaian. Dengan mengetahui nilai UTS, dosen mengetahui sebaran mahasiswa dalam hal kompetensi akademik. Untuk itulah, diperlukan pengelompokan (*clustering*) menggunakan algoritma *k-means* sebagai pertimbangan dosen dalam membentuk kelompok belajar mahasiswa berdasarkan *cluster* nilai UTS.

Kata kunci: UTS, Kompetensi, *Clustering*, *K-means*

1. PENDAHULUAN

Di satu sisi, perkembangan ilmu pengetahuan dan teknologi akan mempermudah pekerjaan manusia. Namun, di lain sisi hal ini akan meningkatkan persaingan baik itu persaingan lokal maupun global. Untuk menghadapi persaingan, perlu dipersiapkan sumber daya manusia. Dalam hal ini mahasiswa yang kompeten dalam bidang akademik berupa pengetahuan dan keterampilan. Salah satu cara untuk mengukur kompetensi akademik adalah dengan nilai mata kuliah. Permendikbud nomor 66 tahun 2013 tentang Standar Penilaian Pendidikan menyebutkan bahwa penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik, dalam hal ini adalah mahasiswa. Sedangkan untuk mengetahui tingkat kompetensi akademik, diperlukan suatu prosedur penilaian yakni

tes/ujian (Gronlund, 1977:1). Ada beberapa komponen penilaian dalam menentukan nilai akhir mata kuliah di tingkat pendidikan tinggi, yang salah satunya adalah nilai Ujian Tengah Semester (UTS). Dengan mengetahui nilai UTS, dosen dapat mengetahui sebaran mahasiswa dalam hal kompetensi akademik.

Dengan pertimbangan tersebut, maka perlu diadakan pengelompokan mahasiswa berdasarkan nilai UTS agar dosen dapat menerapkan metode pembelajaran yang tepat dan menjadi bahan pertimbangan untuk pembentukan kelompok belajar sesuai dengan sebaran kemampuan akademik mahasiswa. Salah satu metode untuk pengelompokan adalah metode *clustering* dengan algoritma *k-means*. Algoritma *k-means* dipilih karena sederhana dan dimulai dengan deskripsi dasar algoritma (Tan, Steinbech, & Kumar, 2006:497). Selain itu, *k-means* juga dapat digunakan untuk

data berukuran besar dan sangat efisien (Tan, Steinbech, & Kumar, 2006:510). Pada Penelitian ini, dataset berupa hasil UTS mata kuliah Data Mining yang akan dikelompokkan dengan metode *clustering* menggunakan algoritma *k-means* dengan bantuan program aplikasi WEKA 3.6.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan proses dalam menemukan hubungan yang berarti, pola dan tren dengan memeriksa data berukuran besar dalam suatu penyimpanan dengan menggunakan teknologi pengenalan pola, misalnya statistik dan matematika. Hal ini sesuai dengan definisi yang dikemukakan oleh Gartner Group dalam Larose (2005:2) yaitu *“Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques”*.

Data mining juga dapat diartikan sebagai proses menemukan pola dalam data. Hal ini dikemukakan oleh Witten, Frank, & Hall (2011:5) yaitu *“Data mining is defined as the process of discovering patterns in data”*. Sedangkan menurut Hand et al dalam (Larose, 2005:2) yaitu *“Data mining is the analysis of (often large) data sets to find unsuspected relationships and to summarize the data that are both understandable and useful”*. Hal ini berarti bahwa data mining merupakan analisis dari data berukuran besar untuk menemukan hubungan tak terduga dan meringkas data agar dapat dipahami dan dapat digunakan. Menurut Pramudiono (Kusrini & Luthfi, 2009: 3) Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang

selama ini tidak diketahui secara manual. Masih menurut Pramudiono (Kusrini & Luthfi, 2009:6) menyatakan bahwa Data Mining diidentikkan dengan *knowledge discovery in database (KDD)* yang mempunyai definisi proses penggalian informasi tersembunyi dalam suatu basis data yang besar.

Dari beberapa pengertian di atas, dapat disimpulkan bahwa Data Mining merupakan serangkaian proses dalam pencarian pola, hubungan, penggalian nilai tambah dari data dan informasi yang berukuran besar berupa pengetahuan dengan tujuan menemukan hubungan dan menyederhanakan data agar diperoleh informasi yang dapat dipahami dan bermanfaat dengan bantuan ilmu statistik dan matematika.

Secara garis besar tahap-tahap Data Mining terdiri dari *selection, preprocessing, transformation, data mining, interpretation/evaluation* (Fayyad dalam Kusrini & Luthfi, 2009: 3). Dalam Data Mining, umumnya metode penelitian berkaitan dengan klasifikasi, clustering, dan Asosiasi. Metode klasifikasi adalah metode yang melibatkan variabel kategori (Larose, 2005:14) yang artinya suatu objek dinyatakan ke salah satu kategori atau kategori yang lain. Metode dalam Data Mining yang merupakan klasifikasi diantaranya adalah *C.4.5, Nearest Neighbor, Naïve bayesian, back propagation*. Metode *clustering* merupakan metode pengelompokan data, observasi, atau kasus menjadi kelas objek-objek yang serupa. Sedangkan *cluster* didefinisikan sebagai kumpulan data yang sama satu sama lain, dan tidak sama dengan data di lain *cluster*. Metode *clustering* mencari segmen keseluruhan data menjadi subgrup-subgrup yang relatif homogen atau biasa disebut sebagai *cluster* (Larose, 2005: 16). Menurut Wu & Kumar (2009:33) algoritma

clustering menempatkan data-data yang sama pada satu kelompok (*cluster*), sedangkan data yang tidak sama pada kelompok (*cluster*) yang lain. Contoh metode clustering ini adalah algoritma *K-means*, algoritma *Agglomerative Hierarchical Clustering*, *Divisive Hierarchical Clustering*. Sedangkan asosiasi sering disebut sebagai analisis keranjang belanja (*market basket analysis*) merupakan metode dalam Data Mining untuk menemukan aturan untuk mengukur hubungan antara dua atau lebih atribut. Aturan asosiasi biasanya dinyatakan dalam bentuk jika anteseden, maka konsekuensi, bersama dengan besarnya nilai *support* dan *confidence* yang berasosiasi dengan aturannya (Larose, 2005:17). Contoh metode ini adalah algoritma *FP-Growth*, algoritma Apriori.

2.2 Algoritma K-means

Algoritma *K-means* merupakan salah satu algoritma *clustering* (pengelompokan). *K-means clustering* merupakan metode *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain, sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi kecil (Agusta, 2007). Sedangkan menurut Wu & Kumar (2009:33) *K-means* adalah algoritma *clustering* yang mempartisi himpunan D menjadi *k cluster* data. Algoritma *K-means* mengkluster semua titik data pada D sedemikian sehingga titik data x_i menjadi satu-satunya *k* partisi. Dengan kata lain, satu titik data hanya masuk ke dalam satu *cluster*.

Adapun langkah-langkah dari algoritma *K-means* adalah sebagai berikut (Tan, Steinbech, & Kumar, 2006:497).

1. Tentukan *K* data sebagai *centroid*, *K* adalah jumlah *cluster* yang diinginkan (ditentukan oleh peneliti).
2. Tiap titik (data) kemudian dicari *centroid* terdekatnya.
3. Setiap himpunan titik (data) yang menjadi *centroid* disebut *cluster*.
4. Hitung kembali *centroid* dari setiap *cluster*.
5. Ulangi langkah 1-4 sampai *centroid* tidak berubah.

Metode *clustering* menggunakan Algoritma *K-means*, ukuran kedekatan data dihitung menggunakan jarak *Euclidean*. Algoritma *K-means* bertujuan untuk meminimumkan jarak total *Euclidean* diantara setiap titik x_i dan *cluster* terdekat yakni c_j (Wu & Kumar, 2009:23). Jarak *Euclidean* ditentukan dengan menggunakan persamaan 2.1 berikut ini (Agusta, 2007).

$$d_{ij} = \sqrt{\sum_{k=1}^n \{x_{ik} - x_{jk}\}^2} \quad \dots (2.1)$$

Keterangan:

d_{ij} = jarak antara data ke-i dan data ke-j

n = dimensi data

x_{ik} = koordinat data ke-i pada dimensi k

x_{jk} = koordinat data ke-j pada dimensi k

Sedangkan menurut Hans & Kamber (2016) algoritma *k-means* membagi data ke dalam *k*-buah *cluster* yang telah ditentukan. Ada beberapa cara yang dapat digunakan untuk menghitung jarak yaitu antara lain dengan menggunakan *Euclidean distance*, *Manhattan distance*, dan *Chebisev distance*. Masing-masing

cara perhitungan jarak tersebut dijelaskan sebagai berikut ini.

1. *Euclidean Distance*

Formula untuk menghitung jarak antar dengan *Euclidean Distance* untuk dua titik dalam satu, dua dan tiga dimensi secara berurutan ditunjukkan pada persamaan (2.2), (2.3), dan (2.4) sebagai berikut.

$$\sqrt{(x-y)^2} = |x-y| \quad \dots (2.2)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \dots (2.3)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2} \dots (2.4)$$

2. *Manhattan Distance (Taxicab distance)*

$$d1(p, q) = \|p - q\|_1 = \sum_{i=1}^n |p_i - q_i| \dots (2.5)$$

3. *Chebisev Distance (Maximum Metric)*

Untuk menentukan jarak dengan menggunakan *Chebisev Distance* dilakukan dengan cara mengambil nilai maksimum dari setiap koordinat dimensinya. Jika dinyatakan dalam persamaan matematika, maka *Chebisev Distance* dapat dilihat pada persamaan (2.6)

$$D_{cheb}(p, q) = \max(|p_i - q_i|) \dots (2.6)$$

Dalam penelitian ini, formula yang akan digunakan untuk menentukan jarak adalah formula *Euclidean Distance* pada persamaan (2.2), (2.3), dan (2.4)

2.3 Ujian/Tes

Tes digunakan untuk mengukur tingkat keberhasilan prosedur pembelajaran (Gronlund, 1977:1). Ebel & Frisbie (1991:30) menyatakan bahwa tes

memberikan informasi yang terbaik bagi guru dan siswa untuk bisa mendapatkan informasi tentang keberhasilan upaya mereka setelah mengajar dan belajar. Ebel & Frisbie (1986: 20-21) menambahkan bahwa fungsi utama dari tes adalah untuk mengukur prestasi siswa dan kontribusinya terhadap evaluasi dan pencapaian proses pendidikan, tes juga membantu guru dalam menentukan nilai yang valid dan level yang reliabel bagi siswa.

Nitko & Brookhart (2011:5) menyatakan "*test is defined as an instrument or systematic procedure for observing and describing one or more characteristics of a student using either a numerical scale or a classification scheme.*" Hal ini berarti tes merupakan suatu instrumen atau prosedur sistematis untuk mengobservasi atau mendeskripsikan satu atau lebih karakteristik siswa dengan menggunakan skala numerik atau skala pengklasifikasian. Menurut Muijs & Reynolds (2005: 232) tes prestasi mengukur kinerja siswa dalam satu mata pelajaran dalam waktu tertentu. Berdasarkan Permendiknas nomor 66 tahun 2013 tentang Standar Penilaian Pendidikan disebutkan bahwa penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa. Prestasi belajar siswa akan diketahui melalui penilaian yang dilakukan guru melalui suatu evaluasi belajar yang biasanya berupa tes tertulis.

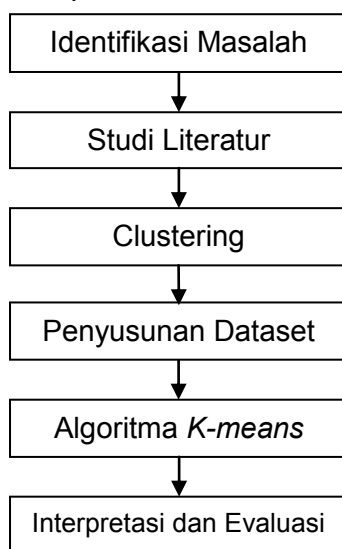
Menurut Miller, Linn, & Gronlund (2009:140-142) pemberian tes dapat dilakukan berdasarkan tujuan pembelajaran dan penilaian yang berbeda-beda, antara lain : (a) sebelum pengujian, diawal suatu pembelajaran bertujuan untuk menentukan apakah siswa memiliki keterampilan prasyarat

yang dibutuhkan untuk pembelajaran, atau untuk mengetahui tingkat kesiapan siswa untuk mencapai tujuan pembelajaran yang direncanakan (untuk menentukan posisi siswa atau modifikasi pembelajaran), (b) Tes formatif, digunakan untuk memonitor kemajuan pembelajaran, mendeteksi miskonsepsi, menyemangati siswa untuk belajar, dan menyediakan umpan balik untuk siswa dan guru, (c) tes sumatif, tes pada akhir pembelajaran dapat digunakan sebagai umpan balik terhadap siswa, menyemangati siswa untuk maju dan mengerjakan yang lebih menantang, menentukan kerja remedial, dan menilai pembelajaran.

Berdasarkan uraian di atas, maka dapat disimpulkan bahwa ujian/tes adalah alat untuk mengukur tingkat keberhasilan prosedur pembelajaran untuk memberikan informasi kepada dosen untuk mengobservasi atau mendeskripsikan satu atau lebih karakteristik siswa dengan menggunakan skala numerik.

3. METODE PENELITIAN

Langkah-langkah dalam penelitian dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Langkah-Langkah Penelitian

a. Identifikasi

Pada tahap ini, dilakukan identifikasi masalah yang ada apa saja serta menawarkan solusi yang dapat digunakan untuk menyelesaikan masalah tersebut.

b. Studi Literatur

Tahap ini merupakan tahap untuk mencari referensi, berupa jurnal penelitian, paper, buku-buku referensi, dan referensi lain terkait penelitian.

c. Penyusunan Dataset

Penelitian ini melibatkan 103 data nilai UTS Mahasiswa semester IV Universitas Muhammadiyah Jember pada Mata Kuliah Data Mining.

d. Clustering

Pada tahap ini, metode *clustering* digunakan untuk mengelompokkan data menjadi tiga kelompok.

e. Algoritma K-Means

Pada tahap ini, algoritma *K-means* diterapkan untuk mengelompokkan dataset yang ada menjadi tiga cluster.

f. Interpretasi dan Evaluasi

Pada tahap akhir ini informasi yang dihasilkan dari proses data mining perlu disajikan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

Sebagai contoh, perhitungan dari algoritma *k-means* berdasarkan dataset pada Tabel 1 dibawah ini.

Tabel 1. Dataset Contoh

Data	Jumlah Transaksi	Total Belanja
1	4	2.255
2	2	1.788
3	2	1.255
4	4	3.322
5	4	3.012
6	3	1.261

Langkah pertama, ditentukan terlebih dahulu jumlah cluster, misalnya $k=3$ dan ditentukan *centroid* (pusat *cluster*) awalnya adalah $C1=(2,1)$; $C2=(4,3)$, $C3=(3,2)$.

Dengan menggunakan Tabel 1, maka dihitung Jarak Euclidean setiap item (Jumlah Transaksi dan Total Belanja) dengan *centroid* awal, maka diperoleh:

Maka data-1, diperoleh:

$$C1 = \sqrt{(4-2)^2 + (2.255-1)^2} = 2.36$$

$$C2 = \sqrt{(4-4)^2 + (2.255-3)^2} = 0.745$$

$$C3 = \sqrt{(4-3)^2 + (2.255-2)^2} = 1.03$$

dan seterusnya, sehingga diperoleh jarak antara data dengan *centroid* awal seperti pada Tabel 2.

Tabel 2. Jarak Euclidean

	C1	C2	C3
1	2.361149085	0.745	1.032000484
2	0.788	2.338577	1.022225024
3	0.255	2.654247	1.247006415
4	3.064585453	0.322	1.657613948
5	2.836925096	0.012	1.422724148
6	1.033499395	2.006021	0.739

Dari Tabel 2 dipilih jarak Euclidean terkecil, sehingga diperoleh matriks penempatan cluster seperti tampak pada Tabel 3 berikut ini.

Tabel 3. Matriks Penempatan Cluster

	C1	C2	C3
1		*	
2	*		
3	*		
4		*	
5		*	
6			*

Ditentukan *centroid* baru yakni $C1=(C11,C12)$, $C2=(C21,C22)$,

$C3=(C31,C32)$ dengan melihat matriks penempatan *cluster* dan menggunakan data pada Tabel 1.

$$C11 = \text{Average}(2,2) = 2$$

$$C12 = \text{Average}(1.788, 1.255) = 1.5215$$

$$C21 = \text{Average}(4,4,4) = 4$$

$$C22 = \text{Average}(2.255, 3.322, 3.012) = 2.863$$

$$C31 = \text{Average}(3) = 3$$

$$C32 = \text{Average}(1.261) = 1.261$$

Sehingga *centroid* baru tersebut adalah:

$$C1 = (C11, C12) = (2, 1.5215)$$

$$C2 = (C21, C22) = (4, 2.863)$$

$$C3 = (C31, C32) = (3, 1.261)$$

Dengan menggunakan *centroid* baru, ditentukan kembali jarak setiap item terhadap *centroid* baru tersebut, sehingga diperoleh data jarak terhadap *centroid* baru seperti terlihat pada Tabel 4.

Tabel 4. Jarak Euclidean

	C1	C2	C3
1	2.130263423	0.608	1.409977305
2	0.2665	2.2706	1.130366755
3	0.2665	2.566255	1.000018
4	2.691059317	0.459	2.290790475
5	2.494311578	0.149	2.016432741
6	1.033373238	1.888493	0

Dari Tabel 4 dipilih jarak Euclidean terkecil sehingga diperoleh matriks penempatan cluster seperti tampak pada Tabel 5.

Tabel 5. Matriks Penempatan Cluster

	C1	C2	C3
1		*	
2	*		
3	*		
4		*	
5		*	
6			*

Dengan membandingkan Tabel 3 dan Tabel 5 ternyata tidak terjadi perpindahan *centroid*, sehingga iterasi *clustering* dihentikan.

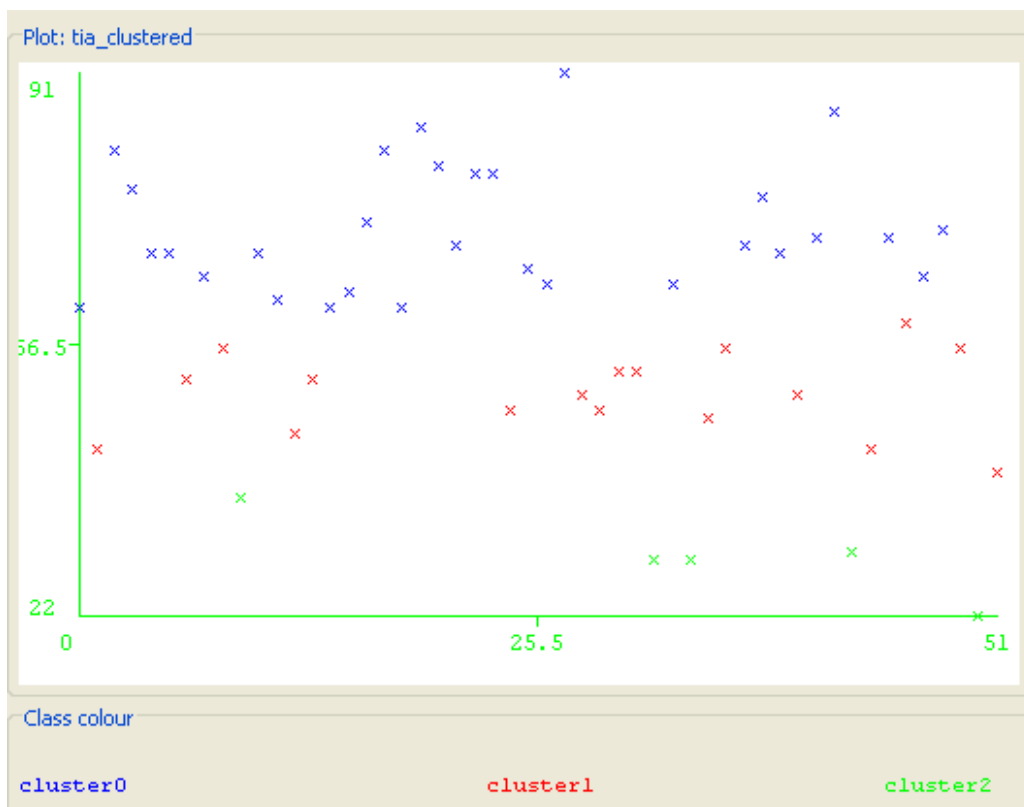
4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan program aplikasi *weka* 3.6 untuk mengelompokkan (*clustering*) nilai UTS mahasiswa dengan algoritma *clustering*, yakni algoritma *k-means*. *Clustering* dilakukan pada nilai Ujian Tengah Semester (UTS) mahasiswa Universitas Muhammadiyah Jember tahun akademik 20152 untuk mata kuliah Data Mining pada kelas Teknik Informatika-A dan Teknik Informatika-B.

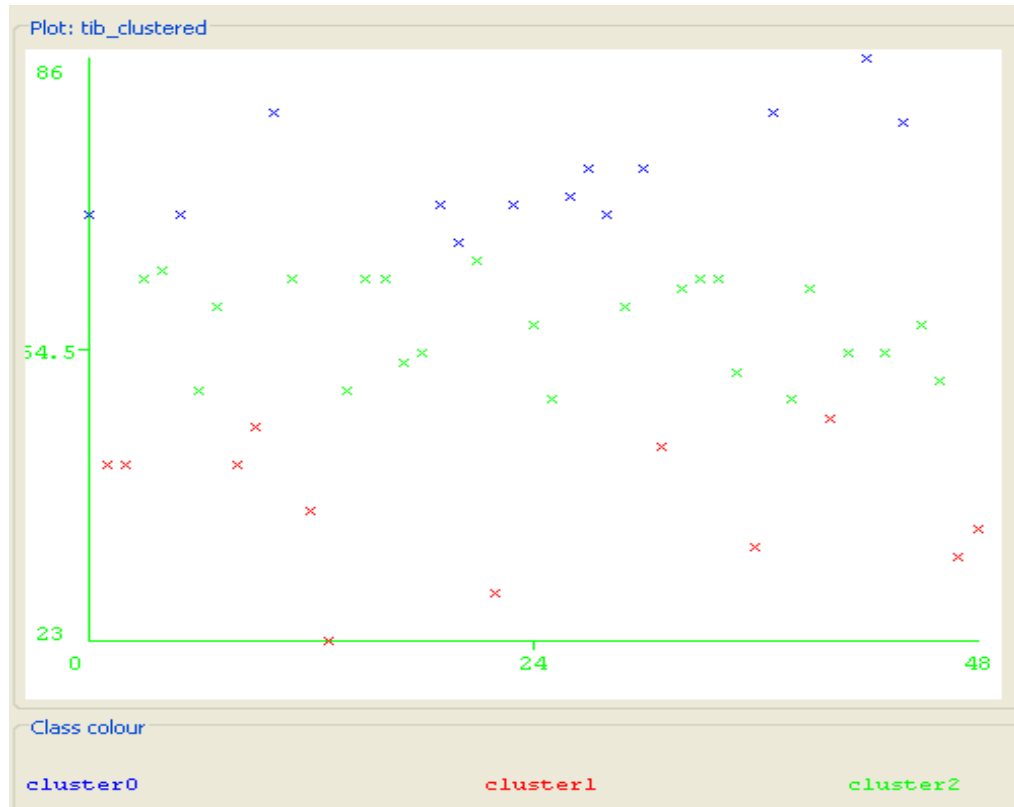
Data dari kelas Teknik Informatika-A dan kelas Teknik Informatika-B, diproses secara terpisah dengan algoritma *k-means* menggunakan aplikasi *WEKA* 3.6 untuk mendapatkan *clustering* mahasiswa berdasarkan nilai UTS. Pada pengujian ini

nilai awal maupun asumsi yang digunakan dalam analisis data yaitu (1) Jumlah *cluster* (*k*) ditentukan sebanyak 3 *cluster*, dan (2) *Cluster-cluster* yang terbentuk yakni *cluster* nilai rendah, *cluster* nilai sedang, dan *cluster* nilai tinggi.

Dari hasil pengolahan terhadap data nilai UTS mahasiswa dengan algoritma *k-means* untuk *clustering* menggunakan aplikasi *WEKA* 3.6, diperoleh dua hasil *clustering*. Hasil *clustering* pertama untuk kelas Teknik Informatika-A dari *output* aplikasi *weka* 3.6 menunjukkan sebanyak 5 (10%) nilai UTS termasuk dalam *cluster* nilai rendah, sebanyak 17 (33%) nilai UTS mahasiswa termasuk dalam *cluster* nilai sedang, dan sebanyak 30 (58%) nilai UTS mahasiswa termasuk dalam *cluster* nilai tinggi. Selengkapnya, hasil *clustering* untuk nilai UTS data mining kelas teknik informatika-A dapat dilihat pada Gambar 2.



Gambar 2. Plot hasil clustering nilai UTS TI-A



Gambar 3. Plot hasil clustering nilai UTS TI-B

Hasil *clustering* kedua untuk kelas Teknik Informatika-B dari *output weka 3.6* menunjukkan sebanyak 12 (24%) nilai UTS termasuk dalam *cluster* nilai rendah, sebanyak 24 (49%) nilai UTS mahasiswa termasuk dalam *cluster* nilai sedang, dan sebanyak 13 (27%) nilai UTS termasuk dalam *cluster* nilai tinggi. Selengkapnya, hasil *clustering* untuk nilai UTS Data Mining kelas Teknik Informatika-B dapat dilihat pada Gambar 3.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan *output* dari aplikasi *weka*, maka dapat diambil dua kesimpulan. Pertama, *clustering* yang terbentuk adalah tiga *cluster*, yakni *cluster* nilai UTS tinggi, *cluster* nilai UTS sedang, dan *cluster* nilai UTS rendah. Kedua, dari *clustering* yang terbentuk dari masing-masing kelas, dapat dijabarkan bahwa untuk kelas Teknik Informatika-A, sebanyak 5 (10%)

nilai UTS termasuk dalam *cluster* nilai rendah, sebanyak 17 (33%) nilai UTS mahasiswa termasuk dalam *cluster* nilai sedang, dan sebanyak 30 (58%) nilai UTS mahasiswa termasuk dalam *cluster* nilai tinggi. Sedangkan untuk Teknik Informatika-B, sebanyak 12 (24%) nilai UTS termasuk dalam *cluster* nilai rendah, sebanyak 24 (49%) nilai UTS mahasiswa termasuk dalam *cluster* nilai sedang, dan sebanyak 13 (27%) nilai UTS termasuk dalam *cluster* nilai tinggi.

Saran pengembangan dapat dilakukan perbandingan dengan algoritma *clustering* yang lain untuk data UTS sebagai perbandingan validitas *clustering* yang terbentuk. Disamping itu, penelitian ini dapat dijadikan sebagai salah satu referensi untuk dosen pengampu dalam pembentukan kelompok belajar mahasiswa, dan sebagai referensi untuk penelitian selanjutnya yang serumpun.

DAFTAR PUSTAKA

- Agusta. (2007). *K-means-penerapan, permasalahan, dan metode terkait stmik stikom bali, denpasar, bali*.
- Ebel, R.L., & Frisbie, D.A., (1986). *Essentials of education measurement (4th ed.)*. Engelwood Cliffs, New Jersey: Prentice Hall.
- Ebel, R.L., & Frisbie, D.A., (1991). *Essentials of educational measurement (5th ed.)*. Engelwood Cliffs, New Jersey: Prentice Hall.
- Gronlund, N.E., (1977). *Constructing achievement test (2nd ed.)*. Engelwood Cliffs, New Jersey: Prentice-Hall.
- Hans, J., & Kamber, M., (2006). *Data mining: concepts and techniques, 2nd ed.* Burlington: Morgan Kaufmann.
- Kemendikbud. (2006). *Peraturan menteri pendidikan dan kebudayaan republik indonesia nomor 66, tahun 2013, tentang standar penilaian pendidikan*.
- Kusrini & Luthfi. (2009). *Algoritma data mining*. Yogyakarta: Andi.
- Larose, D.T., (2005). *Discovering knowledge in data: an introduction to data mining*. New Jersey: John Wiley & Sons.
- Miller, M.D., Linn, R.L., & Gronlund, N.E., (2009). *Measurement and assessment in teaching (10th ed.)*. Upper Saddle River, New Jersey: Pearson Education, Inc.
- Muijs, D., & Reynolds, D., (2005). *Effective teaching: evidence and practice (2nd ed.)*. London: SAGE Publications, Ltd.
- Nitko & Brookhart. (2011). *Educational Assessment of Students*. Boston: Allyn & Bacon..
- Tan, P.N., Steinbech, M., & Kumar, V., (2006). *Introduction to data mining*. Boston: Pearsong Education, Ltd.
- Witten, I.H., Frank, E., Hall, M.A., (2011). *Data mining: practical machine learning tools and techniques third edition*. Burlington: Morgan Kaufmann.
- Wu, X., & Kumar, V., (2009). *The top ten algorithms in data mining*. Boca Raton: CRC Press.