

Rancang Bangun Aplikasi Data Mining Untuk Mendukung Strategi Penentuan Passing Grade Perguruan Tinggi Dengan Metode Clustering

Purwadi^{#1}, M Syaifuddin^{#2}, Hendra Jaya^{#3}

^{#1,2,3} Program Studi Sistem Informasi, STMIK Triguna Dharma

Article Info

Article history:

Received Jan 21th, 2018

Revised Jan 25th, 2018

Accepted Jan 29th, 2018

Keyword:

Clustering

K-Means

Passing Grade

ABSTRACT

Passing grade merupakan suatu standar (grade) dari perhitungan sederhana terhadap hasil ujian SNMPTN yang mengindikasikan kira-kira peserta masuk tidak dalam urutan peringkat kuota penerimaan jurusan yang dipilihnya tersebut. Passing grade umumnya tidak bisa dipatokan secara pasti sebab passing grade sendiri timbul dari nilai-nilai para pesaing yang memperebutkan jurusan tersebut dan selalu berfluktuasi dalam setiap tahunnya. Namun, dalam konteks lain passing grade pun, dapat diprediksi dengan melihat daya tampung dan kuota yang diterima serta tingkat keketatan persaingan tahun sebelumnya. Pengolahan data kemahasiswaan yang dilakukan di perguruan tinggi akan dapat lebih dimanfaatkan secara optimal dengan metode Clustering K-means, mengingat jumlah mahasiswa di perguruan tinggi yang cukup banyak menimbulkan variasi individu yang cukup banyak pula. Data yang akan diolah dengan metode ini adalah data hasil nilai calon mahasiswa dan passing grade sehingga bisa ditemukan visualisasi keterhubungannya. Selanjutnya hasil dari penerapan metode ini diharapkan akan dapat membantu ataupun menjadi acuan dalam pengambilan keputusan penentuan passing grade perguruan tinggi di tahun-tahun berikutnya.

Copyright © 2018 STMIK Triguna Dharma.
All rights reserved.

First Author

Nama : Purwadi, S.Kom, M.Kom

Kantor : STMIK Triguna Dharma

Program Studi : Sistem Informasi

E-Mail : ki_pur@yahoo.co.id

1. PENDAHULUAN

Passing grade merupakan suatu standar (grade) dari perhitungan sederhana terhadap hasil ujian SNMPTN yang mengindikasikan kira-kira peserta masuk tidak dalam urutan peringkat kuota penerimaan jurusan yang dipilihnya tersebut. Passing grade umumnya tidak bisa dipatokan secara pasti sebab passing grade sendiri timbul dari nilai-nilai para pesaing yang memperebutkan jurusan tersebut dan selalu berfluktuasi dalam setiap tahunnya. Namun, dalam konteks lain passing grade pun dapat diprediksi dengan melihat daya tampung dan kuota yang diterima serta tingkat keketatan persaingan tahun sebelumnya.

Data Mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar, yang diharapkan dapat mengatasi kondisi tersebut. Pengolahan data kemahasiswaan yang dilakukan di perguruan tinggi akan dapat lebih dimanfaatkan secara optimal dengan metode Data Mining Clustering, mengingat jumlah mahasiswa di perguruan tinggi yang cukup banyak menimbulkan variasi individu yang cukup banyak pula. Data yang akan diolah dengan metode ini adalah hasil nilai calon mahasiswa dan passing grade sehingga bisa ditemukan visualisasi keterhubungannya. Selanjutnya hasil dari penerapan metode ini diharapkan akan dapat membantu ataupun menjadi acuan dalam pengambilan keputusan penentuan passing grade perguruan tinggi di tahun-tahun berikutnya.

Cluster Analysis adalah sebuah teknik yang digunakan untuk membagi sekumpulan obyek ke dalam k kelompok sehingga nilai dalam setiap kelompok adalah homogen dengan mengacu kepada atribut tertentu berdasarkan kriteria tertentu. Kesamaan nilai (homogeneity) dalam setiap segmen yang diwakili oleh cluster menggambarkan kesamaan pola perilaku perguruan tinggi. Semakin tinggi tingkat akurasi clustering maka

semakin jelas juga kesamaan pola perilaku perguruan tinggi yang dapat digali melalui cluster-cluster yang terbentuk.

2. METODE PENELITIAN

2.1 Data Warehouse

Data warehouse adalah sistem yang mengambil dan menggabungkan data secara periodik dari sistem sumber data ke penyimpanan data bentuk dimensional atau normal (Reinardi, 2008). Data warehouse merupakan penyimpanan data yang berorientasi objek, terintegrasi, mempunyai varian waktu, dan menyimpan data dalam bentuk nonvolatile sebagai pendukung manajemen dalam proses pengambilan keputusan (Han, 2006).

Data warehouse menyatukan dan menggabungkan data dalam bentuk multidimensi. Pembangunan data warehouse meliputi pembersihan data, penyatuan data dan transformasi data dan dapat dilihat sebagai proses yang penting untuk digunakan dalam data mining. Selain itu data warehouse mendukung On-line Analytical Processing (OLAP), sebuah kaskas yang digunakan untuk menganalisis secara interaktif dari bentuk multidimensi yang mempunyai data yang rinci. Banyak metode data mining yang lain seperti asosiasi, klasifikasi, prediksi dan clustering, dapat diintegrasikan dengan operasi OLAP untuk meningkatkan proses mining yang interaktif dari beberapa level abstraksi. Oleh karena data warehouse menjadi platform yang penting untuk data analisis dan OLAP untuk dapat menyediakan platform yang efektif untuk proses data mining. Empat karakteristik dari data warehouse meliputi:

1. Subject Oriented : sebuah data warehouse disusun dalam subjek utama, seperti pelanggan, supplier, produk dan sales. Meskipun data warehouse terkonsentrasi pada operasi harian dan proses transaksi dalam perusahaan, data warehouse rumah fokus pada pemodelan dan analisis data untuk membuat keputusan. Oleh karena itu data warehouse mempunyai karakter menyediakan secara singkat dan sederhana gambaran seputar subjek lebih detail yang dibuat dari data luar yang tidak berguna dalam proses pendukung keputusan.
2. Integrated : data warehouse biasanya dibangun dari macam-macam sumber yang berbeda, seperti database relational, flat files, dan on-line transaction records. Pembersihan dan penyatuan data diterapkan untuk menjamin konsistensi dalam penamaan, struktur kode, ukuran atribut, dan yang lainnya.
3. Time variant : data disimpan untuk menyajikan informasi dari sudut pandang masa lampau (misal 5-10 tahun yang lalu). Setiap struktur kunci dalam data warehouse mempunyai elemen waktu baik secara implisit maupun eksplisit.
4. Nonvolatile : sebuah data warehouse secara fisik selalu disimpan terpisah dari data operasional. Penyimpanan yang terpisah ini, data warehouse tidak memerlukan proses transaksi, recovery dan mekanisme pengendalian konkurensi. Biasanya hanya membutuhkan dua operasi dalam akses data yaitu, initial load of data dan access of data.

2.2 Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencaai pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Data mining juga disebut sebagai serangkaian proses data menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007).

Data mining juga sering disebut sebagai Knowledge Discovery In Database (KDD). KDD adalah kegiatan yang meliputi pengetahuan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data dalam ukuran (Santoso, 2007). Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatian data analysis, image database, signal processing (Han, 2006).

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semi otomatis (Witten, 2005). Karakteristik data mining sebagai berikut:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis terutama dalam strategi (Davies, 2004).

Berdasarkan pengertian tersebut dapat ditarik kesimpulan bahwa data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu lokasi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

2.2.1 Tahapan Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi 7 (enam) tahap, diantara tahap-tahap itu adalah sebagai berikut:

1. Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh dari database suatu perusahaan maupun hasil dari eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Data Integrasi (Data integration)

Integritas data merupakan penggabungan data dari berbagai database ke dalam database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integritas data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integritas perlu dilakukan secara cermat karena kesalahan pada integritas data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan asumsi nantinya.

3. Seleksi Data (data selection)

Data yang ada pada database, sering kali tidak semuanya dipakai, oleh karena itu hanya yang sesuai untuk dianalisis yang diambil dari database. Sebagai contoh sebuah kasus yang meneliti kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan cukup dengan id pelanggan saja.

4. Transformasi data (data transformation)

Data diubah atau digabung ke format yang sesuai untuk diproses dalam data mining. Beberapa metode dalam data mining membutuhkan perangkat data khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar analisis asosiasi dan clustering hanya dapat mengimput data kategorikal. Karena data berupa data numerik yang berlanjut perlu dibagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (pattern evaluation)

Untuk mengidentifikasi pola-pola yang menarik dalam knowledge base yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola khas maupun model prediksi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba data mining lain lagi yang lebih sesuai, atau menerima hasil ini sebagai sebagai hasil yang diluar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan (knowledge presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap akhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari analisis yang didapat. Ada kalanya hal ini harus melibatkan orang yang tidak memahami tentang data mining. Karena presentasi hasil data mining dalam bentuk yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga dapat membantu mengkomunikasikan hasil data mining (Han, 2006).

2.2.2 Clustering

Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. Gagasan mengenai pengelompokan data, atau clustering, memiliki sifat yang sederhana dan dekat dengan cara berpikir manusia kapanpun kepada kita dipresentasikan jumlah data yang besar, kita biasanya cenderung merangkumkan

jumlah data yang besar ini ke dalam sejumlah kecil kelompok-kelompok atau kategori-kategori untuk memfasilitasi analisisnya lebih lanjut.

Bahkan, algoritma clustering berusaha mensegmentasikan seluruh kumpulan data ke dalam subkelompok-subkelompok atau cluster-cluster homogen secara relatif.

Dimana kesamaan record dalam cluster dimaksimalkan dan kesamaan dengan record diluar cluster ini diminimalkan. Clustering sering dilaksanakan sebagai langkah pendahuluan dalam proses pengumpulan data, dengan cluster-cluster yang dihasilkan digunakan sebagai input lebih lanjut ke dalam sebuah teknik yang berbeda, seperti neural network. Karena ukuran yang besar dari banyak database yang dipresentasikan saat ini, dapat membantu untuk menggunakan analisa clustering terlebih dahulu, untuk mengurangi ruang pencarian untuk algoritma-algoritma downstream. Aktivitas clustering pola khusus meliputi langkah-langkah berikut :

- a. Representasi pola (secara opsional termasuk ekstraksi dan/atau seleksi sifat).
- b. Defenisi ukuran kedekatan pola yang tepat untuk domain data.
- c. Clustering pengelompokan.
- d. Penarikan data (jika dibutuhkan).
- e. Pengkajian output (jika dibutuhkan).

Secara garis besar, terdapat beberapa metode clustering data. Pemilihan metode clustering bergantung pada tipe data dan tujuan clustering itu sendiri. Metode-metode beserta algoritma yang termasuk didalamnya adalah sebagai berikut:

1. Partitioning Methods
Metode yang membangun berbagai partisi dan kemudian mengevaluasi partisi tersebut dengan beberapa kriteria. Algoritma yang dipakai pada metode ini adalah K-Means, K-Medoid, PROCLUS, CLARA, CLARANS dan PAM.
2. Hierarchical Methods
Metode yang membuat suatu penguraian secara hierarchial dari himpunan data dengan menggunakan beberapa kriteria. Metode ini terdiri atas dua jenis, yaitu Agglomerative yang menggunakan strategi bottom-up dan Disisive yang menggunakan strategi top-down. Metode ini meliputi algoritma BIRCH, AGNES, DIANA, CURE dan CHAMELEON.
3. Density-Based Methods
Metode ini berdasarkan konektivitas dan fungsi densitas. Metode ini meliputi algoritma DBSCAN, OPTICS dan DENCLU.
4. Grid-Based Methods
Metode ini berdasarkan suatu struktur granularitas multi-level. Metode clustering ini meliputi algoritma STING, WaveCluster dan CLIQUE.
5. Model-Based Methods
Suatu model dihipotesakan untuk masing-masing cluster dan ide untuk mencari best fit dari model tersebut untuk masing-masing yang lain. Metode clustering ini meliputi pendekatan statistik, yaitu algoritma COBWEB dan jaringan syaraf tiruan SOM.

2.3 K-Means Clustering

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristk yang berbeda di kelompokkan ke dalam cluster yang lain. Secara umum algoritma dasar dari K-Means Clustering adalah sebagai berikut :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu Euclidean Distance Space. Euclidean distance space sering digunakan dalam perhitungan jarak, hal ini dikarenakan

hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Adapun persamaannya adalah sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2}$$

dimana :

d_{ij} = Jarak objek antara objek i dan j

P = Dimensi data

x_{ik} = Koordinat dari obyek i pada dimensi k

2.4 Passing Grade

Passing grade adalah suatu yang sifatnya dinamis artinya, setiap periode SNMPTN memiliki standar tersendiri untuk jurusan-jurusan tertentu setiap tahunnya. Ada cara untuk menghitung passing grade dalam menjawab soal-soal SNMPTN tahun lalu atau Try out untuk menghadapi SNMPTN. Contohnya dengan mengerjakan soal SNMPTN tahun lalu dan kemudian mencocokkan jawaban dengan kunci jawaban yang tersedia.

Lalu hitung passing grade dengan rumus

$$\text{Passing Grade: } \frac{(Bx4)-(Sx1)}{JSx4} x 100 = \dots$$

Keterangan :

B = Jumlah jawaban benar

S = Jumlah jawaban salah

JS = Jumlah soal

3. ANALISIS DAN HASIL

Berikut adalah contoh studi kasus penyelesaian masalah dengan aturan cluster k-mens yang diambil dari kampus Unisversitas Sumatera Utara

Dekripsi		
Jumlah soal	Jawaban Benar	Jawaban Salah
100	-	-

- Mahasiswa 1 :

Hari pertama (materi dasar), mahasiswa 1 berhasil menjawab 50 soal SNMPTN dengan benar dan 32 soal yang salah dan 18 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(50x4) - (32x1)}{100x4} x 100 = 42$$

- Mahasiswa 2 :

Hari pertama (materi dasar), mahasiswa 2 berhasil menjawab 37 soal SNMPTN dengan benar dan 49 soal yang salah dan 14 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(37x4) - (49x1)}{100x4} x 100 = 24.75$$

- Mahasiswa 3 :

Hari pertama (materi dasar), mahasiswa 3 berhasil menjawab 38 soal SNMPTN dengan benar dan 48 soal yang salah dan 14 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(50x4) - (48x1)}{100x4} x 100 = 38$$

- Mahasiswa 4 :

Hari pertama (materi dasar), mahasiswa 4 berhasil menjawab 30 soal SNMPTN dengan benar dan 70 soal yang salah.

$$\text{Penyelesaian: } \frac{(30x4) - (70x1)}{100x4} x 100 = 12.5$$

- Mahasiswa 5 :

Hari pertama (materi dasar), mahasiswa 5 berhasil menjawab 34 soal SNMPTN dengan benar dan 32 soal yang salah dan 34 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(34 \times 4) - (32 \times 1)}{100 \times 4} \times 100 = 26$$

- Mahasiswa 6 :

Hari pertama (materi dasar), mahasiswa 6 berhasil menjawab 42 soal SNMPTN dengan benar dan 32 soal yang salah dan 26 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(42 \times 4) - (32 \times 1)}{100 \times 4} \times 100 = 34$$

- Mahasiswa 7 :

Hari pertama (materi dasar), mahasiswa 7 berhasil menjawab 40 soal SNMPTN dengan benar dan 60 soal yang salah.

$$\text{Penyelesaian: } \frac{(40 \times 4) - (60 \times 1)}{100 \times 4} \times 100 = 25$$

- Mahasiswa 8 :

Hari pertama (materi dasar), mahasiswa 8 berhasil menjawab 37 soal SNMPTN dengan benar dan 49 soal yang salah dan 14 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(44 \times 4) - (53 \times 1)}{100 \times 4} \times 100 = 30.75$$

- Mahasiswa 9 :

Hari pertama (materi dasar), mahasiswa 9 berhasil menjawab 50 soal SNMPTN dengan benar dan 48 soal yang salah dan 2 soal tidak dijawabnya.

$$\text{Penyelesaian: } \frac{(50 \times 4) - (48 \times 1)}{100 \times 4} \times 100 = 38$$

- Mahasiswa 10 :

Hari pertama (materi dasar), mahasiswa 10 berhasil menjawab 40 soal SNMPTN dengan benar dan 60 soal yang salah

$$\text{Penyelesaian: } \frac{(40 \times 4) - (60 \times 1)}{100 \times 4} \times 100 = 25$$

Berikut *Passing Grade* USU 2013 yang dijadikan studi kasus:

Kategori IPA

151015	Pendidikan Dokter	40 %
151031	Teknik Mesin	36.3%
151053	Teknik Industri	35.8%
151245	Teknologi Informasi	35.6%
151231	Ilmu Komputer	34.3%
151061	Teknik Kimia	34.2%
151142	Keteknikan Pertanian	33.1%
151045	Teknik Elektro	32.1%
151075	Pendidikan Dokter Gigi	31.2%
151172	Ilmu dan Teknologi Pangan	31.0%
151083	Matematika	30.8%
151261	Teknik Lingkungan	30.5%
151023	Teknik Sipil	30.2%
151112	Fisika	29.2%

151104 Farmasi 28.5%
 151091 Kimia 28.2%
 151134 Agroekoteknologi 27.8%
 151156 Biologi 27.3%
 151194 Arsitektur 26.2%
 151253 Manajemen Sumber Daya Perairan 25.6%
 151215 Kehutanan 24.9%
 151126 Ilmu Kesehatan Masyarakat 24.8%
 151201 Ilmu Keperawatan 24.6%
 151164 Agribisnis 24.5%
 151186 Peternakan 23.8%

Kategori IPS

152011 Ilmu Hukum 35.4%
 152025 Ilmu Sejarah 28.2%
 152033 Antropologi Sosial 27.5%
 152041 Akuntansi 39.2%
 152055 Sosiologi 28.6%
 152063 Ilmu Kesejahteraan Sosial 31.5%
 152071 Ilmu Administrasi Negara 24.2%
 152085 Ilmu Komunikasi 31.5%
 152093 Manajemen 34.3%
 152106 Ekonomi Pembangunan 33.6%
 152114 Ilmu Politik 25.2%
 152122 Sastra Indonesia 26.1%
 152136 Sastra Melayu 19.2%
 152144 Sastra Inggris 28.1%
 152152 Sastra Batak 23.4%
 152166 Sastra Arab 30.6%
 152174 Etno Musikologi 30.2%
 152182 Sastra Jepang 31.1%
 152196 Ilmu Perpustakaan 27.1%
 152203 Sastra Cina 21.4%
 152225 Ilmu Administrasi Niaga/Bisnis 20.2%

Pembangunan perangkat lunak *data mining* dengan metode *clustering* menggunakan algoritma hirarki *divisive* untuk pengelompokan mahasiswa dalam studi kasus ini, fungsi-fungsi yang dipakai adalah fungsi untuk menentukan titik-titik pusat yang berguna sebagai pusat-pusat kelompok mahasiswa. Fungsi - fungsi tersebut adalah sebagai berikut :

Data awal yang digunakan dalam perhitungan ini berjumlah 20. Data mahasiswa sebelum dilakukan perhitungan menggunakan metode *Clustering K-Means*. Dan ketentuan tingkat *passing grade* adalah sebagai berikut :

Tabel 1 Penentuan Titik Pusat

<i>Passing Grade</i> Mahasiswa	Himpunan	Nilai Himpunan
0-20	Rendah	1
20.1 – 40	Cukup	2
> 40	Tinggi	3

Tabel 2 Data Awal

Calon Mahasiswa Ke	Passing Grade (%)	Himpunan
1	42	3
2	24.75	2
3	38	2
4	12.5	1
5	26	2
6	34	2
7	25	2
8	30.75	2
9	38	3
10	25	2

Pada awalnya ditentukan berapa cluster yang akan dibentuk. Sebarang obyek atau elemen pertama dalam cluster dapat dipilih untuk dijadikan sebagai titik tengah (centroid point) cluster. Algoritma metode K-Means selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan (tidak ada obyek yang dapat dipindahkan) :

1. menentukan koordinat titik tengah setiap cluster,
2. menentukan jarak setiap obyek terhadap koordinat titik tengah,
3. mengelompokkan obyek-obyek tersebut berdasarkan pada jarak minimumnya

Langkah-langkah yang dilakukan oleh algoritma metode K-Means adalah sebagai berikut :

1. Pengesetan nilai awal titik tengah.

Langkah pertama, menentukan pusat cluster secara acak pada data awal yang ada di tabel 3.1. Cluster pertama yang akan digunakan dalam perhitungan adalah C1 (10;1) yang artinya passing grade mahasiswa pada pengelompokan pertama adalah 10 dan himpunan ada 1. Cluster kedua C2 (30;2) dan cluster ketiga yang akan digunakan adalah C3(40;3).

2. Menghitung jarak obyek ke centroid dengan menggunakan rumus jarak Euclid Perhitungan jarak mahasiswa pertama dengan pusat cluster pertama adalah :

DAFTAR PUSTAKA

- [1] Eko Prasetyo. 2012. Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta Andi
- [2] Hermawati Fajar Astuti. 2013. Data Mining. Yogyakarta. Andi
- [3] Hendrayudi. 2010. Dasar-dasar Pemrograman Microsoft Visual Basic. Cet.I. Bandung
- [4] Rosa A.S & M. Shalahuddin. (2013). Rekayasa Perangkat Lunak. Bandung. Informatika
- [5] Sani Susanto, Ph. D. & Dedy Suryadi, ST., M.S.(2010). Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data. Yogyakarta : Penerbit ANDI.
- [6] Siallagan Sariadin. 2010. Microsoft Visual Basic 6.0 dan Crystal Report 2008. Yogyakarta, Andi

BIOGRAFI PENULIS

		
Purwadi, S.Kom, M.Kom	M Syaifuddin, S.Kom, M.Kom	Hendra Jaya, S.Kom, M.Kom