

PENGELOMPOKAN MAHASISWA MENGUNAKAN ALGORITMA K-MEANS

Narwati

Dosen Fakultas Teknologi Informasi

Abtrack— Makalah ini membahas pengelompokan mahasiswa berdasarkan data akademik menggunakan teknik clustering dan membuat aplikasinya kemudian menganalisis hasilnya sehingga diharapkan mampu memberikan informasi bagi yang berkepentingan. Algoritma K-Means merupakan salah satu algoritma teknik clustering yang dimulai dengan pemilihan secara acak K , yang merupakan banyaknya cluster yang ingin dibentuk dari data yang akan di kluster, yaitu nilai tes mahasiswa saat masuk dan Indeks Prestasi Kumulatif mahasiswa sampai semester 8. Sistem yang dibuat menampilkan hasil klustering data akademik mahasiswa, yaitu pola dari prestasi mahasiswa yang klusteranya tetap, turun dan naik, dan dapat terlihat dari asal program studi, asal kota dan asal SMA. Dari hasil studi kasus dapat diperoleh informasi mahasiswa yang tetap pada kluster seperti awal masuk sebanyak 422 (45,085%), mahasiswa yang naik kluster sebanyak 284 (30,342%) dan mahasiswa yang turun klusteranya sebanyak 230 (24,573%),

Kata kunci: teknik clustering, Algoritma K-Means.

1. PENDAHULUAN

Seleksi masuknya mahasiswa dalam sebuah perguruan tinggi umumnya dengan memberikan soal-soal test yang harus mereka kerjakan, untuk mengetahui kemampuan dan pengetahuan mereka. Setelah mahasiswa mengalami proses belajar mengajar, maka akan dapat diketahui prestasi mereka setiap akhir semester. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi. Jumlah data yang banyak ini membuka peluang untuk dihasilkan informasi yang berguna bagi pihak universitas.

Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah record dan jumlah field yang cukup banyak) tidak dapat dilakukan dengan mudah. Teknologi data mining merupakan salah satu alat bantu untuk penggalan data pada basis data berukuran besar dan dengan spesifikasi tingkat kerumitan yang telah banyak digunakan pada banyak domain aplikasi seperti perbankan maupun bidang telekomunikasi [1]. Algoritma K-Means merupakan algoritma teknik klustering yang berulang-ulang. Algoritma ini dimulai dengan pemilihan secara acak K , yang merupakan banyaknya kluster yang ingin dibentuk. Kemudian

tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari kluster atau biasa disebut dengan centroid / mean. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus yang sudah disediakan hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasi setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut sampai nilai centroid tidak berubah (stabil).

Data akademik tersebut adalah hasil evaluasi tes masuk penerimaan mahasiswa baru (PMB) berupa nilai tes potensial akademik (TPA) dan nilai tes bahasa inggris, data atribut identitas diri mahasiswa seperti nama, nim, asal sekolah, kota asal, usia, jenis kelamin serta nilai Indeks Prestasi Kumulatif (IPK). Dengan menggunakan data hasil tes masuk, dan pencapaian indeks prestasi kumulatif pada semester 8, maka dapat diketahui minat belajar dari mahasiswa apakah tetap pada nilai test awal masuk atau ada perubahan yang signifikan.

2. DASAR TEORI

2.1 Data Mining

Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan tehnik gambar dari statistik, mesin pembelajaran dan sistem manajemen database [2].

Data mining yang disebut juga dengan *Knowledge-Discovery in Database* (KDD) adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*).

Secara sederhana, data mining dapat diartikan sebagai proses mengekstrak atau “menggali” pengetahuan yang ada pada sekumpulan data. Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge-Discovery in Database* atau yang biasa disebut KDD. Dari sudut pandang yang lain, data mining dianggap sebagai satu langkah yang penting didalam proses KDD. Menurut Han, J. and Kamber, M, 2001, proses KDD ini terdiri dari langkah-langkah sebagai berikut :

1. *Data Cleaning*, proses menghapus data yang tidak konsisten dan kotor
2. *Data Integration*, penggabungan beberapa sumber data
3. *Data Selection*, pengambilan data yang akan dipakai dari sumber data
4. *Data Transformation*, proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining
5. *Data Mining*, suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data
6. *Pattern Evaluation*, proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada didalam data itu sendiri
7. *Knowledge Presentation*, proses visualisasi dan teknik menyajikan knowledge digunakan untuk menampilkan knowledge hasil mining kepada user

2.2. Clustering

Proses pengelompokan sekumpulan obyek kedalam kelas-kelas obyek yang sama disebut clustering / pengelompokan [2]. Pengklasteran merupakan satu dari sekian banyak fungsi proses data mining untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Analisis kluster (Clustering) merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengklasteran adalah menyatakan sekumpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga.

2.3. Algoritma K-means

K-means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster / kelompok. Metode ini mempartisi ke dalam cluster / kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda (*Low inter class similarity*) dikelompokkan pada kelompok yang lain [3]. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n$; $j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k$; $j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat

digunakan formula Euclidean [2] seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots(1)$$

Suatu data akan menjadi anggota dari cluster ke- k apabila jarak data tersebut ke pusat cluster ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap cluster.

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots(2)$$

Nilai pusat cluster yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada cluster tersebut, dengan menggunakan rumus pada persamaan 3:

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p}; \dots\dots\dots(3)$$

Dimana $x_{ij} \in$ cluster ke - k

p = banyaknya anggota cluster ke k

Algoritma dasar dalam k-means adalah

1. Tentukan jumlah cluster (k), tetapkan pusat cluster sembarang.
2. Hitung jarak setiap data ke pusat cluster menggunakan persamaan (2.1).
3. Kelompokkan data ke dalam cluster yang dengan jarak yang paling pendek menggunakan persamaan (2.2).
4. Hitung pusat cluster yang baru menggunakan persamaan (2.3)

Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

3. METODE PENELITIAN

Pembuatan program aplikasi pada penelitian ini mengikuti langkah-langkah pengembangan perangkat lunak model *waterfall* (air terjun). Model ini dipilih dengan alasan untuk membangun sistem ini dibutuhkan beberapa tahap yang berbeda, yang diawali dengan analisa kebutuhan, perancangan, implementasi, pengujian sistem dan perawatan sistem.

3.1 Analisa kebutuhan

Pada analisa kebutuhan sistem membahas beberapa kebutuhan dan atau persyaratan terkait dengan input, proses dan output. Kebutuhan atau persyaratan ini

diperoleh berdasarkan data penerimaan mahasiswa baru dan data akademik yang dapat diperoleh dari server Unisbank. Berdasarkan data tersebut diperoleh hasil analisa kebutuhan sistem berikut:

a. Kebutuhan input

Sistem yang akan dibangun membutuhkan data input, antara lain :

- Data penerimaan mahasiswa baru (NIM, Nama, Progdi, Kategori SMA, Asal SMA, Nilai test)
- Data akademik (NIM, , dan IPK)

b. Kebutuhan Proses

Proses yang digunakan untuk mengolah data input adalah teknik klustering dengan algoritma K-Means.

c. Kebutuhan Output

Output yang diharapkan adalah analisa dari hasil klustering sehingga dapat diketahui informasi / pola dari data input

3.2 Perancangan/Desain

Proses perancangan / desain meliputi penyusunan blok-blok program untuk memudahkan pada saat penulisan program (coding) dan perancangan antar muka (*user interface*) untuk keperluan interaksi sistem dengan pengguna (*user*). Secara umum desain perangkat lunak mempertimbangkan dari segi pengguna, yaitu : tampilan antar muka dan proses perhitungan/komputasi dengan benar.

3.3 Implementasi/Coding

Proses penulisan program dilakukan secara modular sesuai dengan pembagian blok-blok program pada langkah perancangan/desain. Bahasa pemrograman yang digunakan adalah Visual Basic

3.4 Pengujian Sistem/ Testing

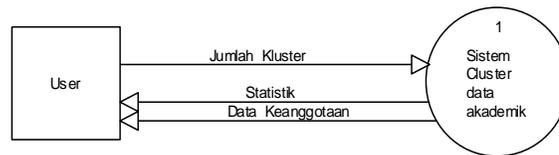
Proses pengujian meliputi ; data akademik mahasiswa beserta data penerimaan mahasiswa baru sebagai masukan (input) sehingga menghasilkan beberapa kluster dan analisisnya sesuai yang rancangan. Untuk tahap kelima yaitu pemeliharaan/*Maintenance*, tidak dilakukan karena keterbatasan waktu.

3.5 Perancangan Proses

Data Flow Diagram (DFD) adalah sebuah teknik grafis untuk menggambarkan bagaimana data diproses melalui sistem dalam bentuk *input* dan *output*. DFD dimulai dengan menggambarkan keseluruhan sistem dan dilanjutkan dengan merinci masing-masing area fungsi.

DFD *level 0* atau dikenal dengan DFD *level konteks* akan menggambarkan aliran data secara umum pada sistem dimana hanya ada satu proses pada level ini.

Perancangan proses sistem dengan menggunakan teknik klustering untuk data nilai test dan IPK mahasiswa pada DFD level konteks digambarkan seperti pada gambar 1.

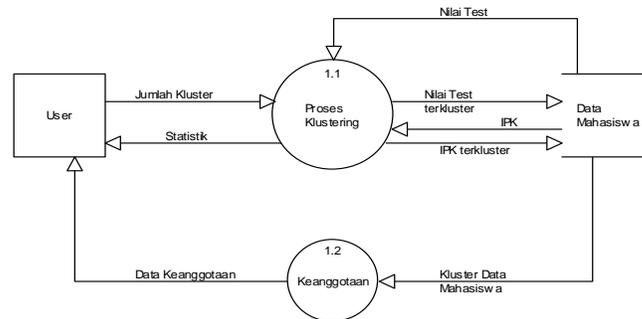


Gambar 1. Diagram konteks clustering

Pada diagram konteks terdapat sebuah entitas luar yaitu *user* yang sekaligus sebagai admin yang berinteraksi dengan sebuah proses yaitu sistem pengklusteran nilai test masuk dan IPK mahasiswa, dalam hal ini bisa kepala program studi dan bagian humas.

DFD level 0 dikembangkan lebih dalam lagi untuk memperjelas dan menambah detail aliran data dan proses pada sistem. Pengembangan tersebut menghasilkan DFD level berikutnya yaitu DFD level 1, seperti pada gambar 2.

Pada DFD level 1 terdapat 2 proses, yaitu proses clustering dan Proses Keanggotaan. Output Keanggotaan adalah jumlah mahasiswa yang tetap, naik dan turun klusternya, beserta dengan rincian program studinya, jenis SMA dan jenis asal kotanya.



Gambar 2. DFD level 1 teknik clustering

Pada proses 1.1 DFD level 1 diperinci lagi menjadi level 2. Pada DFD level 2 terdapat 6 proses yaitu mencari secara random nilaiTes, mencari secara random IPK, hitung kluster nilaiTes, hitung kluster IPK, hitung statistic dan menset jumlah kluster. Pada output statistik akan diperlihatkan jumlah iterasi, nilai centroid tiap kluster.

3.5. Flow Chart

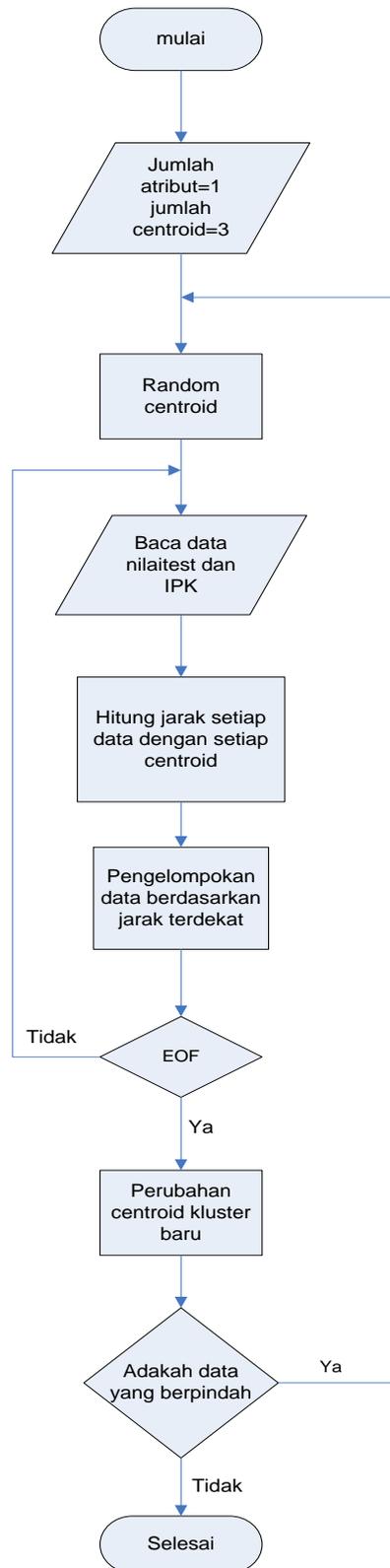
Urutan proses clustering dengan algoritma K-Means adalah sebagai berikut :

- a. Algoritma akan mengambil data yang akan dikluster dari tabel mahasiswa.
- b. Menentukan jumlah atribut, untuk masing-masing kluster terdapat 1atribut, yaitu nilai test dan IPK

mahasiswa, kemudian menentukan jumlah *centroid*/titik pusat sementara sebanyak 3, baik untuk nilai test maupun IPK.

- c. Mengambil nilai *centroid* sementara sebanyak 3 secara random.
- d. Membaca data nilai test dan IPK
- e. Menghitung jarak setiap data dengan *centroid*, jarak terdekat akan menjadi anggota pada kluster tersebut.
- f. Apakah proses perhitungan jarak dan pengelompokan pada langkah e sudah selesai, apabila belum maka akan membaca kembali data nilai test dan IPK. Demikian dilakukan sampai semua data terhitung.
- g. Menghitung rata-rata dari dari semua data yang menjadi anggota kluster tersebut, rata-rata ini akan menjadi titik pusat untuk iterasi selanjutnya. Hal ini dilakukan sampai nilai rata-rata tidak berubah.
- h. Apakah ada perubahan anggota kluster, kalau ada perubahan, mengambil lagi nilai *centroid* secara random dari data mahasiswa

Proses tersebut dapat dilihat pada gambar 3.



Gambar 3. Flow Chart

4. HASIL DAN PEMBAHASAN

4.1 Persiapan Data

Setelah data terkumpul maka tahap selanjutnya adalah mempersiapkan data tersebut agar dapat digunakan untuk proses data mining. Data mentah yang akan digunakan pada aplikasi ini diperoleh dari beberapa sumber penyimpanan data transaksi yaitu data identitas diri pendaftar pmb dan nilai tes masuk serta kemajuan belajar mahasiswa sampai dengan semester delapan(8), sejumlah 577 mahasiswa angkatan 2005 dan 490 mahasiswa angkatan 2006, total sejumlah 1067 record. Pengolahan data awal merupakan bagian dari persiapan data dimana langkah-langkah yang dilakukan antara lain meliputi menghilangkan kerangkapan data dan membersihkan data kotor, menggabungkan data, menentukan atribut yang akan diolah dan mengubah data. Persiapan data dilakukan secara manual menggunakan excel dan hasil proses persiapan data disajikan dalam bentuk tabel. Dari sejumlah data awal 1067 record setelah mengalami proses awal (*pre-processing*) data menjadi 936 record.

4.2 Output Program Aplikasi

Setelah persiapan selesai disusun dan program aplikasi juga telah dibangun, maka tahap selanjutnya adalah menggunakan aplikasi tersebut untuk mengkluster data yang sudah siap diproses.

a. Proses Pengambilan Data

Pada proses ini data yang diambil dari Access hanya satu tabel dengan atribut Progd, NIM, nama, Nilai test, IPK, usia, LP (Jenis Kelamin), SLTA (Nama SMA), Kota (Kota SMA), Jenis SMA. Tabel ini yang akan dipakai untuk proses klustering

b. Proses Klustering

1. Proses mencari random NilaiTes

Pada penelitian klustering data akademik ini peneliti mengelompokkan data nilai tes menjadi 3 cluster, dimana cluster 1 (c1) adalah mewakili data paling bawah/rendah, cluster 2 (c2) adalah mewakili data sedang, dan cluster 3 (c3) mewakili data tertinggi. Pada langkah awal sebelum proses klustering dilakukan, terlebih dahulu sistem akan memilih nilai tes mahasiswa sebanyak 3 angka secara random yang akan digunakan sebagai nilai centroid pertama. Dan pada proses ini apabila ada angka yang sama, maka akan diulang sampai ketiga angka yang terpilih adalah berbeda. Setelah diperoleh 3 angka nilai test, akan dilakukan proses klustering dengan algoritma K-Means.

2. Proses mencari random IPK

Pengelompokkan data ipk juga menjadi 3 cluster, dimana cluster 1 (c1) adalah mewakili data paling bawah/rendah, cluster 2 (c2) adalah mewakili data sedang, dan cluster 3 (c3) mewakili data tertinggi.

Seperti pada NilaiTes langkah awal sebelum proses klustering dilakukan, terlebih dahulu sistem akan memilih ipk mahasiswa sebanyak 3 angka secara random yang akan digunakan sebagai nilai centroid pertama. Dan pada proses ini apabila ada angka yang sama, maka akan diulang sampai ketiga angka yang terpilih adalah berbeda. Setelah diperoleh 3 angka ipk, akan dilakukan proses klustering. Contoh hasil proses mencari angka random dapat dilihat pada gambar 4.



Gambar 4. Hasil pencarian nilai_ tes dan IPK secara random

3. Proses menghitung cluster NilaiTes

Setiap data akan mengelompok pada salah satu cluster dari 3 cluster yang ada, berdasarkan jarak kedekatan yang dihitung dengan rumus *euclidien*. Pada setiap cluster akan dihitung rata-rata dari jumlah data yang menjadi anggotanya, dan nilai rata-rata dari tiap cluster akan menjadi centroid pada perhitungan jarak di iterasi selanjutnya. Jumlah iterasi pada perhitungan jarak dan penentuan anggota cluster tidak tentu, tergantung dari pengambilan secara random untuk nilai_ tes. Pada pengambilan centroid secara random akan menghasilkan rata-rata nilai_ tes yang stabil yaitu cluster 1 pada angka 38,733, cluster 2 pada angka 64,336 dan cluster 3 pada angka 85,825

4. Proses menghitung cluster IPK

Pada pengambilan centroid IPK secara random akan menghasilkan rata-rata IPK yang stabil yaitu cluster 1 pada angka 1,281, cluster 2 pada angka 2,581 dan cluster 3 pada angka 3,221. Contoh hasil rata-rata centroid dapat dilihat pada gambar 5.

STATISTIK	NILAI
Iterasi Tes	14
Iterasi IPK	16
Centroid 1 Nilai Tes	38.733
Centroid 2 Nilai Tes	64.336
Centroid 3 Nilai Tes	85.825
Centroid 1 Nilai IPK	1.281
Centroid 2 Nilai IPK	2.581
Centroid 3 Nilai IPK	3.221

Gambar 5. Centroid rata-rata Nilai_ tes dan IPK

c. Proses Keanggotaan

Proses ini adalah output dari proses setelah muncul

STATISTIK	JUMLAH	PERSEN
+ Tes Cluster	936	100
+ IPK Cluster	936	100
+ Tetap Cluster	422	45.085
+ Naik Cluster	284	30.342
+ Turun Cluster	230	24.573

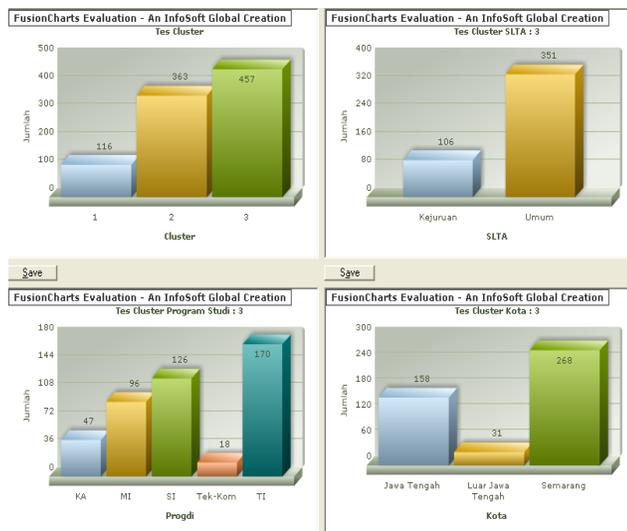
Gambar 6. Output anggota cluster pada tingkat pertama

Pada output menu anggota cluster dibuat dengan diagram pohon, pada cluster nilai_ tes, IPK, cluster tetap, cluster yang naik, cluster yang turun, masing masing mempunyai cabang, seperti terlihat pada gambar 7

STATISTIK	JUMLAH	PERSEN
- Tes Cluster	936	100
+ 1	116	12.393
+ 2	363	38.782
+ 3	457	48.825
+ IPK Cluster	936	100
+ Tetap Cluster	422	45.085
- Naik Cluster	284	30.342
+ Dari 1 ke 2	52	18.310
+ Dari 1 ke 3	42	14.789
+ Dari 2 ke 3	190	66.901
- Turun Cluster	230	24.573
+ Dari 3 ke 2	162	70.435
+ Dari 3 ke 1	41	17.826
+ Dari 2 ke 1	27	11.739

Gambar 7. Output anggota cluster pada tingkat Kedua

Selanjutnya output lainnya dari menu anggota adalah tampilan grafik secara Histogram untuk induk cluster nilaiTes dan cabang-cabangnya. Seperti terlihat pada gambar 8.



Gambar 8. Output anggota cluster pada grafik Histogram induk nilai test dan cabangnya

4.3 Pembahasan

Berdasarkan proses data mining dengan teknik klustering menggunakan algoritma K-Means yang diterapkan pada data akademik mahasiswa, diperoleh informasi dari pengelompokan atau pengklusteran nilaiTes mahasiswa saat masuk dari sejumlah 936 mahasiswa adalah sejumlah 116 mahasiswa atau sebesar 12,393% masuk cluster 1, 363 (38,782%) mahasiswa masuk cluster 2 dan 457 (48,825%) mahasiswa masuk cluster 3. Hal ini berarti hampir sebagian besar kemampuan mahasiswa saat masuk kuliah adalah masuk cluster 3, atau berada pada kemampuan paling atas.

Dari sejumlah 116 mahasiswa yang masuk cluster 1, sebagian besar yaitu sebanyak 100 mahasiswa masuk program studi SI dan sisanya masuk TI 10 mahasiswa, KA dan Tek-Kom 4 dan 2 mahasiswa. Jenis SMA sebanyak 89 mahasiswa dari Umum dan 27 mahasiswa dari Kejuruan. Untuk Asal Kota dari Semarang sebanyak 65 mahasiswa, dari Jawa Tengah 41 mahasiswa dan dari luar Jawa Tengah 10 mahasiswa.

Namun apabila dilihat dari prosentase jumlah mahasiswa setiap program studinya, maka yang menduduki tempat tertinggi adalah MI sebanyak 96 dari 146 mahasiswa di program studi ini berada di cluster 3 dengan prosentase 65%. Dilihat dari Jenis SMA secara keseluruhan jenis SMA Umum yang menempati tempat tertinggi dibanding SMA kejuruan yang masuk cluster ini yaitu sebesar 56,7%. Sedangkan apabila dilihat dari Asal SMA, prosentase rata-rata hampir sama dari setiap wilayah ada di cluster tiga. Untuk lebih jelas dapat dilihat pada tabel 1.

Tabel 1. Rekapitulasi Jumlah Mahasiswa Untuk Program Studi, Jenis SMA dan Jenis Kota Pada Cluster Nilai Test

Atribut	Unit Dari Atribut	Kluster1 (C1)	Kluster2 (C2)	Kluster3 (C3)	Total Unit dari Atribut
Program Studi	KA	4	27	47	78
	MI	2	48	96	146
	SI	100	146	126	372
	Tek-Kom	0	26	18	30
	TI	10	130	170	310
Jenis SMA	Umum	89	265	351	705
	Kejuruan	27	98	106	231
Jenis Kota	Jawa Tengah	41	131	158	330
	Luar Jawa Tengah (LJ)	10	24	31	65
	Semarang	65	208	268	541

Setelah 8 semester berjalan kemampuan mahasiswa dapat dilihat dari IPK yang dicapai. Sebanyak 90(9,615%) mahasiswa masuk cluster 1, 360 (38,462%) mahasiswa masuk cluster 2 dan sebanyak 486(51,922%) mahasiswa masuk cluster 3. Hal ini berarti ada perubahan kemampuan mahasiswa antara

saat masuk kuliah dan setelah 8 semester berjalan. Apabila dilihat dari prosentase jumlah mahasiswa setiap program studi, maka KA menempati urutan pertama dengan jumlah mahasiswa sebanyak 63 dari 78 mahasiswa (80,79%) berada di kluster 3. Hasil ini menunjukkan bahwa kemampuan mahasiswa dari program studi KA adalah paling unggul dibandingkan program studi lainnya. Untuk jenis SLTA dari Umum adalah yang paling unggul dan untuk asal SMA Semarang menempati tempat tertinggi dibanding yang lainnya. Untuk lebih jelasnya dapat dilihat pada tabel 2.

Tabel 2. Rekapitulasi Jumlah Mahasiswa Untuk Setiap Kluster dan Perubahannya

Kluster	1	2	3
Nilai Test	116	363	457
IPK	90	360	486
Tetap Pada	22	146	254
Naik Kluster	1 ke 2	1 ke 3	2 ke 3
	52	42	190
Turun Kluster	3 ke 2	3 ke 1	2 ke 1
	162	41	27

5. KESIMPULAN

Dari hasil pembuatan aplikasi dapat disimpulkan hal-hal sebagai berikut ;

1. Program aplikasi menghasilkan pola dari prestasi mahasiswa yang klusternya tetap, turun dan naik. Pola mahasiswa tersebut dapat terlihat dari asal program studi, asal kota dan asal SMA.
2. Pada hasil kluster untuk nilai tes prosentase terbesar adalah mahasiswa yang masuk kluster 3, dan ini berarti bahwa sebagian mahasiswa yang masuk ke Universitas ini adalah mempunyai kemampuan yang tinggi berdasarkan hasil tes
3. Setelah mahasiswa menempuh kuliah selama 6 semester untuk program diploma-3 dan 8 semester untuk program strata-1, mahasiswa yang berada di kluster 3 tetap dengan jumlah prosentase terbanyak, namun jumlah ini tidak semua berasal dari mahasiswa yang pada saat masuk berada pada kluster ini, karena ada mahasiswa dengan prestasinya menurun dan ada yang naik.
4. Mahasiswa yang tetap pada klusternya, terutama kluster 3 apabila dilihat dari program studinya, Komputerisasi Akuntansi berada di urutan pertama mahasiswa dengan kemampuan tetap di kluster tiga, hal ini bisa disebabkan mereka mempunyai potensi dan kecerdasan yang tinggi. Untuk jenis SMA, prosentase tertinggi adalah mahasiswa yang berasal dari jenis SMA Umum yang tetap di kluster 3. Sedangkan untuk asal SMA masing masing wilayah mempunyai prosentase yang sama.

5. Mahasiswa yang meningkat klusternya dari kluster 1 ke kluster 3 mempunyai prosentase yang cukup kecil, sedangkan kenaikan jumlah prosentase terbanyak adalah dari kluster 2 ke kluster 3. Hal ini berarti bahwa proses belajar mengajar bagi mahasiswa yang kemampuannya berada pada kluster 1 harus lebih ditingkatkan.
6. Mahasiswa yang turun klusternya, paling tinggi prosentasenya adalah dari kluster 3 ke kluster 2. Prosentase tertinggi untuk program studi adalah Tek-Kom yang mengalami penurunan prosentase paling banyak sebesar 40% dari jumlah mahasiswa yang ada..

DAFTAR PUSTAKA

- [1]. Jananto, Arief. 2010. *Memprediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining (Studi kasus data akademik mahasiswa UNISBANK*. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.
- [2]. Han, Jiawei; & Kamber, Micheline. 2001. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kauffman.
- [3]. Giyanto, Heribertus. 2008. *Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva*. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.
- [4]. Tan, P; Steinbach; & Kumar, V. 2006. *Introduction to Data Mining*. Pearson Education.
- [5]. Budiarti, Andina.; Suchyo, Giri, Yudho.; & Ruldeviyani, Yova. 2006. *Studi Karakteristik Kelulusan Peserta Didik Dengan Teknik Clustering*. Seminar Nasional Sistem dan Informatika. Bali.
- [6] Khoo, S.G, Christopher; & Ramaiah, K, Chennupati. 2004. *Profile of LIS Applicants Selecting Different Specialisations*. Libri, vol. 54, pp. 67–81 Germany: Printed.