PENERAPAN CLUSTERING DI BIDANG PENDIDIKAN

DENGAN METODE K-MEANS

Literature ReviewTopik Clustering

****

Mahal Aditya W15.01.55.0032

Fakultas Teknologi Informasi

Universitas Stikubank Semarang

SemarangJuli, 2018

**DAFTAR ISI**

**Daftar Isi**1

**Bab I Pendahuluan**2

* 1. Latar Belakang2
  2. Permasalahan2
  3. Tujuan3

**Bab II Pembahasan**4

2.1. Review Jurnal4

Jurnal 1: Penerapan Data Mining Dalam Meningkatkan Mutu Pembelejaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Program Studi TKJ Akademi Komunitas Solok Selatan)4

Jurnal 2: Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang7

Jurnal 3: Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro 9

2.2. Analisis Jurnal11

**Bab III Penutup**13

3.1. Kesimpulan13

3.2. Saran13

**Daftar Pustaka**14

**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Pendidikan merupakan hal penting, dan akan terus berkembang seiring berjalannya zaman, di Indonesia agak sulit mengimbangi perkembangan itu, karena di Indonesia mutu pendidikannya pun masih tergolong rendah. Peningkatan mutu pembelajaran di perguruan tinggi merupakan urgensi yang mendesak untuk segera dilakukan perbaikan. Peningkatan mutu itu pada dasarnya dapat dilakukan dengan strategi merubah salah satu dari subsistem : manusia, struktur, teknologi, dan proses organisasi.

Bagaimana ekstraksi *Data Mining* yang dihasilkan dapat memberikan sebuah pengetahuan baru terhadap pendidikan di Indonesia dalam meningkatkan mutu pembelajaran*. Data Mining* dilakukan untuk menggali informasi dari sejumalah data. Keluaran dari *Data Mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Salah satu teknik yang dikenal dalam *Data Mining* yaitu *clustering*. Analisis *cluster* berguna dalam meringkas data atau sejumlah variabel untuk menjadi lebih sedikit.

Dengan adanya teknik tersebut, diharapkan mampu meningkatkan kualitas pendidikan di Indonesia khususnya di jenjang perguruan tinggi dalam berbagai hal dan bidang studi. Teknik ini sangat dibutuhkan untuk kemajuan pendidikan.

* 1. **Permasalahan**

Pengolahan data yang tidak baik, ketersediaan data yang tidak dibarengi dengan kemampuan pemanfaatan data yang baik serta ketidaktahuan dalam merubah data menjadi suatu informasi yang bermanfaat merupakan kendala dalam pendidikan yang ada sekarang. Selain itu sumber daya manusia yang belum memadai membuat data yang disajikan menjadi tidak berguna.

* 1. **Tujuan**

Adapun tujuan dari penyusunan laporan ini adalah :

1. Mempelajari serta memberi ulasan terhadap paper yang ditulis oleh para peneliti, khususnya dalam topik *clustering.*
2. Menganalisis jurnal-jurnal tersebut, ditinjau dari segi isi / materi maupun penulisan jurnal.
3. Melakukan komparasi terhadap jurnal-jurnal tersebut, terkait dengan persamaan dan perbedaannya terutama dari segi isi / materi dan penerapan metode maupun teknik-teknik yang digunakan/dilakukan.

**BAB II**

**PEMBAHASAN**

* 1. **Jurnal Review**

**Jurnal 1: Penerapan Data Mining Dalam Meningkatkan Mutu Pembelejaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Program Studi TKJ Akademi Komunitas Solok Selatan)**

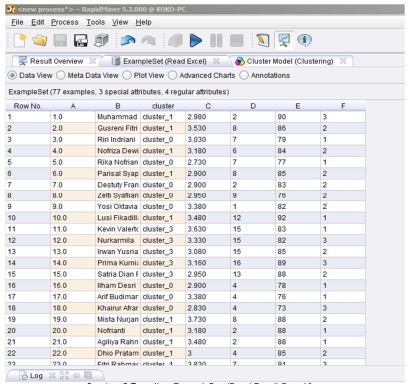
Bagian ini membahas mengenai analisis data dan metode yang digunakan untuk menyelesaikan permasalahan penelitian. Data yang digunakan adalah data mutu pembelajaran Akademi Studi Kasus Di Program Studi TKJ Akademi Komunitas Solok Selatan. Proses clustering dengan menggunakan Algoritma K-Means, sebagai berikut :

1. Menentukan Jumlah ClusterJumlah cluster yang ditentukan untuk mengelompokan data pada penelitian ini sebanyak 4 cluster.
2. Menentukan Centroid Pusat awal cluster (centroid)ditentukan secara random atau acak.
3. Menghitung Jarak dari CentroidMenghitung jarak antara titik centroid dengan titik tiap objek dengan menggunakan Euclidian Distance.
4. Alokasikan masing-masing objek ke centroid terdekat.

Pada tahap pengujian algoritma ini untuk membuktikan kebenaran pada tahap analisa sebelumnya dan pengujian secara manual, maka perlu dilakukan pengujian lagi untuk pengelompokan data meningkatkan mutu pembelajaran menggunakan algoritma K-Means. Selanjutnya setelah database berhasil diproses, maka dapat dilihat hasil pengelompokan data Mahasiswa Program Studi TKJ Akademi Komunitas Solok Selatan. Untuk melihat bukti bahwa database berhasil diproses :

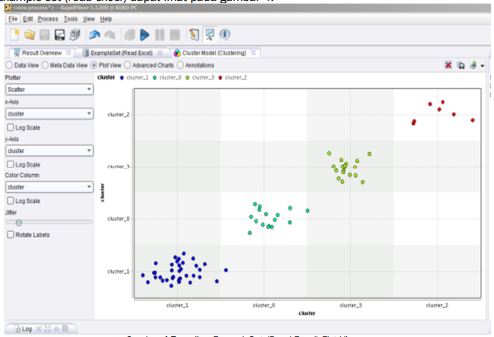
1. DATA VIEW

Dilihat dari example set (read excel) dapat lihat pada gambar 3:



Gambar 3 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya berjumlah 77 data.

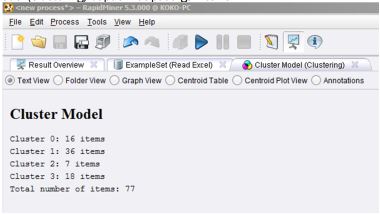
1. FLOT VIEW

Dilihat dari example set (read excel) dapat lihat pada gambar 4.

Merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya. Urutan clusternya yaitu cluster\_1, cluster\_0, cluster\_3, cluster\_2.

1. TEXT VIEW

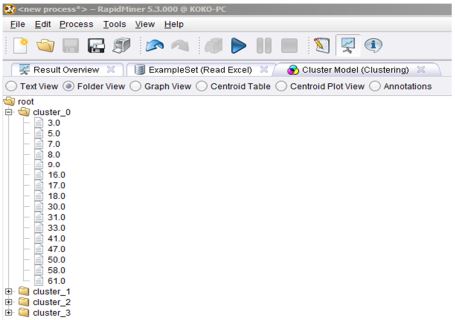
Dilihat dari cluster model (clustering) dapat lihat pada gambar 5



Gambar 5 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya. Hasil dari data 77 dimana C0 memiliki 16 items, C1 memiliki 36 items, C2 memiliki 7 items, C3 memiliki 18 items.

1. FOLDER VIEW

Dilihat dari cluster model (clustering) dapat lihat pada gambar 6:



Pada Gambar 6 menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya.

Jumlah anggota cluster :  
C1 berjumlah 16 = 3,5,7,8,9,16,17,18,30,31,33,41,47,50,58,61  
C2 berjumlah 36 = 1,2,4,6,10,19,20,21,22,23,24,25,29,34,35,36,45,48,49,52,53,54,55,57,60,63,64,65,  
68,69,70,71,72,73,75,77  
C3 berjumlah 7 = 26,38,39,40,66,67,74.  
C4 berjumlah 18 = 11,12,13,14,15,27,28,32,37,42,43,44,46,51,56,59,62,76

Jurnal 2: **Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang**

*Data Pengujian* Data pengujian yang digunakan adalah berupa tabel yang memiliki komponen penyusun sebagai berikut:

1. Memiliki 5 attribute yaitu nim mahasiswa, nilai mata kuliah algoritma dan pemrogaman 1, nilai mata kuliah fisika dasar, nilai kalkulus 1, dan IPK.
2. Jumlah *instance* adalah 124.

*Algoritma K-Means* Oyelade (2010) menyatakan algoritma ini disusun atas dasar ide yang sederhana. Sebaran obyek dan elemen pertama dalam *cluster* dapat dipilih untuk dijadikan sebagai titik tengah (*centroid point*) *cluster.* Algoritma metode KMeans selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan (tidak ada obyek yang dapat dipindahkan):

1. Menentukan koordinat titik tengah setiap *cluster.* Penentuan *cluster* dibuat 4 buah obyek dengan 3 atribut. Metode *Clustering* dengan algoritma K-Means akan  
   menghasilkan 4 *cluster* berdasarkan *class* Indek Prestasi Komulatif (IPK). Pengesetan nilai awal tengah dengan menentukan titik tengah (*centroid*) dari *cluster* seperti pada  
   table 1.
2. Penentuan nilai dari *cluster-cluster* tersebut untuk dijadikan acuan untuk melakukan perhitungan pada setiap baris tabel data pengujian. Contoh penentuan jarak obyek ke *centroid*, pengujian dilakukan pada NIM = 12.0504.0009 yang memiliki IPK = 0.25, dengan mengacu pada rumus Euclid yang telah disederhanakan (*cluster* x-IPK):

Jarak 0 = (0.5 - 0.25) = 0.25

Jarak 1 = (3.4 - 0.25) = 3.15

Jarak 2 = (2.3 - 0.25) = 2.05

Jarak 3 = (2.9 - 0.25) = 2.65

Dari hasil diperoleh jarak 0 = 0.25 akan mendekati nilai pada *cluster* 0, maka proses yang sama akan dilakukan untuk semua data pengujian.

1. Pengelompokan obyek-obyek tersebut berdasarkan pada jarak minimumnya dilakukan dengan menggunakan hasil proses pada langkah 2. Hasil pada perhitungan jarak akan digunakan untuk penentuan *clustering*.

*Pengujian dengan Software Weka* Pengujian data dengan *Software* Weka menghasilkan data berupa:

1. Nilai *cluster centroids* dan *cluster instances*.
2. Grafik *clustering* posisi mahasiswa pada setiap *cluster* masing-masing.

Dari data yang dilatih, didapatkan 4 kelompok dengan hasil sebagai berikut:

1. Mahasiswa dengan IPK = 0.5167 untuk *cluster* 0, sebanyak 9 Mahasiswa dari 124  
   Mahasiswa (7%)
2. Mahasiswa dengan IPK = 3.4143 untuk *cluster* 1, sebanyak 28 Mahasiswa dari 124  
   Mahasiswa (23%)
3. Mahasiswa dengan IPK = 3.3092 untuk *cluster* 2, sebanyak 40 Mahasiswa dari 124  
   Mahasiswa (32%)
4. Mahasiswa dengan IPK = 3.8991 untuk *cluster* 3, sebanyak 47 Mahasiswa dari 124 Mahasiswa (38%)

Maka *cluster* 1 dengan IPK tertinggi bisa digunakan untuk memilih 5 Mahasiswa untuk bisa mewakili lomba.

**Jurnal 3: Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro**

Persiapan data mencakup semua kegiatan untuk membangun *dataset* mahasiswa yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* mahasiswa dan selanjutnya akan melakukan proses *data mining*.

*Tool yang digunakan adalah RapidMiner* versi 5.3:Flowchart algoritma *K-means*:

* 1. Penetapan jumlah *cluster* (K) pada penelitian ini yaitu berjumlah 3 *cluster*.

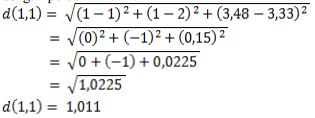
|  |  |  |  |
| --- | --- | --- | --- |
| **Titik Pusat Awal** | **Jurusan** | **Kota Asal** | **IPK** |
| *Cluster* 1 | 1 | 2 | 3,33 |
| *Cluster* 2 | 3 | 1 | 3,53 |
| *Cluster* 3 | 2 | 6 | 2,81 |

* 1. Hitung jarak setiap data ke pusat *cluster* antara objek ke *centroid* dengan perhitungan jarak *Euclidean.* Persamaan yang digunakan adalah :



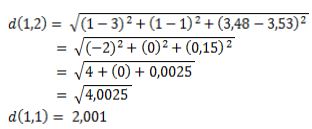
Persamaan 2-3 diatas digunakan karena atribut yang digunakan berjumlah 3.

Sebagai contoh, akan dihitung jarak dari data mahasiswa pertama ke pusat *cluster* pertama dengan persamaan:



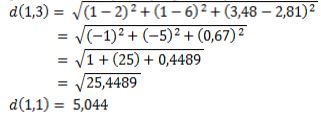
Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 1,011.

Jarak dari data mahasiswa pertama ke pusat *cluster* kedua dengan persamaan:



Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 2,001.

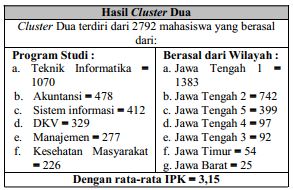
Jarak dari data mahasiswa pertama ke pusat *cluster* ketiga dengan perhitungan:

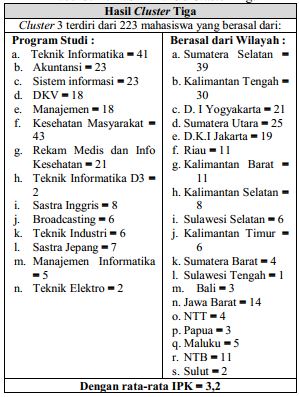


Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 5,044.

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkam bahwa jarak data mahasiswa pertama yang paling dekat adalah *cluster* 1, sehingga mahasiswa pertama dimasukkan ke dalam *cluster* 1.

Hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data mahasiswa pada setiap atribut.





* 1. Analisis Jurnal

**Jurnal 1: Penerapan Data Mining Dalam Meningkatkan Mutu Pembelejaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Program Studi TKJ Akademi Komunitas Solok Selatan)**

Dalam jurnal yang ditulis oleh Koko Handoko ini terdapat beberapa hal yang masih kurang. Hal tersebut adalah metode yang dilakukan masih kurang apabila terdapat data yang lebih banyak.

Jurnal 2: **Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang**

Dalam paper yang ditulis oleh Asroni dan Ronald Adrian ini terdapat beberapa hal yang belum dibahas secara jelas. Hal tersebut adalah hasil yang diperoleh kurang maksimal

**Jurnal 3: Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro**

Dalam paper yang ditulis oleh Rima Dian Ramadhani ini data yang ditampilkan jelas dan memberikan informasi yang baik.

BAB III

PENUTUP

* 1. Kesimpulan

*Clustering* merupakan teknik yang penting dalam pengolahan data pada bidang pendidikan, yang berfungsi untuk meningkatkan mutu pendidikan khususnya di jenjang perguruan tinggi entah itu swasta ataupun negeri. *Clustering*menyajikan 2 metode yaitu K-means dan Hierarchical.

Metode yang sering dan mudah digunakan yaitu K-means. Di dalam metode ini, terdapat algoritma yang cukup mudah untuk mengolah data. Selain itu tools yang digunakan juga bermacam-macam.

Oleh karena itu, algoritma ini sangat mudah diterapkan untuk melakukan penelitian khususnya di bidang pendidikan. Seperti yang dilakukan oleh jurnal-jurnal di atas.

* 1. **Saran**

1. Penjelasan terhadap metode-metode yang digunakan sebaiknya disertai dengan  
   contoh secara konkrit sehingga memudahkan untuk pemahaman serta analisis.
2. Didalam melakukan analisis terhadap hasil eksperimen agar lebih jelas, bukan hanya  
   disebutkan akan tetapi diberi penjelasan secara rinci.
3. Untuk perhitungan nilai, akurasi atau tingkat error dari hasil eksperimen yang  
   dilakukan sebaiknya digambarkan secara rinci, sehingga dapat diketahui performa  
   dari parameter-parameter yang diukur.

**DAFTAR PUSTAKA**

**Handoko, Koko,**  “Penerapan Data Mining Dalam Meningkatkan Mutu Pembelejaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Program Studi TKJ Akademi Komunitas Solok Selatan)”**,** TEKNOSI, Vol. 02, No. 03, Desember 2016

**Asroni, Ronald Adrian,** “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang”, JURNAL ILMIAH SEMESTA TEKNIKA Vol. 18, No. 1, 76-82, Mei 2015

**Dian Ramadhani, Rima**, “Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Di