

Perbandingan Algoritma K-Means dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah

Mardiani

Sistem Informasi STMIK GI MDP

E-mail: mardiani@mdp.ac.id

Abstrak

Dari beberapa fungsionalitas data mining, digunakan clustering untuk mengelompokkan mahasiswa berdasarkan nilai. Cluster dilakukan dengan menggunakan algoritma yang sudah ada yaitu K-Means dan EM (Expectation Maximation). Setelah sebelumnya melakukan proses pembersihan data dengan menggunakan aplikasi SQL Server 2008, kemudian data dalam bentuk tabel diolah dengan aplikasi WEKA (Waikato Environment for Knowledge Analysis) untuk mendapatkan hasilnya. Hasil dari penelitian berupa clustering informasi sekolah mana yang berpotensi menghasilkan lulusan dengan nilai yang baik. Pengelompokan terdiri atas 3 cluster dengan kategori nilai tinggi, sedang dan rendah. Pengelompokan tersebut juga berdasarkan lokasi yang disebut sebagai spatial clustering. Kemudian dilakukan analisis hasil setelah mendapatkan data yang sudah terkelompok. Informasi yang didapat selanjutnya dapat dimanfaatkan untuk pengambilan keputusan di bidang pendidikan bagi mahasiswa dan manajemen STMIK MDP. Bagi pihak manajemen STMIK MDP informasi berguna untuk mengetahui sekolah mana yang memberikan kontribusi mahasiswa dengan nilai tertinggi.

Kata Kunci —Data Mining, Clustering, Spatial Clustering

Abstract

From some of the functionality of data mining, clustering is used to group students based on the value. Clusters is done by using existing algorithms namely K-Means and EM (Expectation Maximation). Having previously done the cleaning process data using SQL Server 2008 applications, then the data in tabular form is processed by the WEKA (Waikato Environment for Knowledge Analysis) to get the result. Results from the study of clustering information which school has the potential to produce graduates with good grades. The grouping consists of three clusters with the category of high value, medium and low. Grouping is also referred to as a location based spatial clustering. Then performed the analysis of results after getting the data is already grouped. The information obtained can then be utilized for decision making in the field of education for students and management STMIK MDP. For the STMIK MDP management information useful to know which schools contribute to student with the highest score.

Keywords —Data Mining, Clustering, Spatial Clustering

1. PENDAHULUAN

Dengan semakin berkembangnya teknologi Informasi, cabang ilmu komputer data mining juga semakin turut ikut ambil bagian dalam perkembangan zaman. Data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru untuk meringkas data agar mudah dipahami serta berguna untuk pemilik data. Perubahan data menjadi Informasi dengan menggunakan data mining membuat ketersediaan informasi menjadi banyak. Hasil dari penggunaan teknologi informasi di hampir semua bidang kehidupan, menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam informasi dari perubahan data tersebut [4].

Data Mining telah dipakai diberbagai bidang seperti ilmu sains, bisnis dan industri, teknik, kesehatan, serta pertahanan dan keamanan. Menurut [2] Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan, menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam limpahan data tersebut, yang kemudian melahirkan data mining. Data mining merupakan proses untuk menemukan pengetahuan (*knowledge discovery*) yang ditambang dari sekumpulan data yang volumenya sangat besar. Aplikasi data mining pada pengelolaan bisnis, pengendalian produksi, dan analisa pasar misalnya, memungkinkan diperolehnya pola dan hubungan yang dapat dimanfaatkan untuk peningkatan penjualan, atau pengelolaan sumber daya dengan lebih baik. Dunia pendidikan memiliki data yang berlimpah dan berkesinambungan mengenai siswa yang dibina dan alumni yang dihasilkannya. Hal ini membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik [2].

Sedangkan di dunia pendidikan, Data Mining juga telah diterapkan untuk berbagai kebutuhan. Perguruan tinggi memiliki berbagai macam data misalya data akademik, pemasaran dan keuangan, meliputi data mahasiswa, dosen, dan karyawan yang beraneka ragam. Menurut [7], pembagian kelas peserta kuliah dapat dilakukan dengan menggunakan *Fuzzy Clustering* dan *Partition Coefficient and Exponential Separation*. Namun hasil yang didapatkan masih belum cukup jelas kesimpulannya. Di sinilah peran data mining dapat dipakai sebagai pembanding untuk mendapatkan hasil dan akhirnya dapat diambil suatu keputusan.

Dengan berbagai pertimbangan tersebut diatas, penerapan teknik data mining sangat diperlukan untuk membantu manajemen STMIK MDP dalam pemanfaatan teknologi informasi tidak hanya dari segi kebutuhan operasional tetapi juga dari segi kebutuhan strategis. Dari berbagai fungsionalitas data mining, dipilih teknik clustering untuk mencari dan melihat berbagai kelompok yang tercipta dari nilai mahasiswa. Pengelompokan data nilai mahasiswa ini diharapkan menghasilkan suatu kesimpulan dalam membantu manajemen untuk keputusan berikutnya. Keputusan tersebut misalnya menentukan kelompok untuk peminatan mata kuliah pilihan, pengelompokan untuk pilihan tema skripsi dan juga mencari kelompok yang tercipta dari nilai-nilai mahasiswa berdasarkan asal sekolah yang dapat memberikan informasi baru kepada bagian pemasaran.

Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma *clustering* merupakan algoritma pengelompokkan sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*). Objek data yang terletak didalam satu *cluster* harus mempunyai kemiripan. Sedangkan yang tidak berada didalam satu *cluster* tidak mempunyai kemiripan.

Setiap *cluster* memiliki *centroid* yang merupakan suatu besaran yang dihitung dari rata-rata nilai tiap *items* dari suatu *cluster* dan juga memiliki *medoid* yang merupakan item yang letaknya paling tengah. Sementara jarak antar *cluster* didefinisikan dengan menggunakan beberapa metode-metode untuk menentukan *cluster* mana yang berdekatan [3]:

1. *Single link*: jarak terkecil antara satu elemen dalam suatu *cluster* dengan elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
2. *Complete link*: jarak terbesar antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.

3. *Average*: jarak rata-rata antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
4. *Centroid*: jarak antara *centroid* dari tiap *cluster* dengan *centroidcluster* lainnya.
5. *Medoid*: jarak antara *medoid* dari tiap *cluster* dengan *medoidcluster* lainnya.

Clustering memiliki banyak algoritma dengan beberapa kategori yaitu *hierarchical* yang menentukan sendiri jumlah *cluster* yang dihasilkan misalnya algoritma SNN, COBWEB, Chandeon dan Rock. *Partitional* yang mengelompokkan data kedalam k *cluster* dimana k adalah banyak *cluster* dari input user misalnya algoritma CLARA, K-Means, EM dan Bond Energy. Serta *clustering large data* yang memiliki volume data paling besar misalnya algoritma BIRCH, DBScan dan Cactus.

1.1. K-Means

Menurut [1] data *clustering* menggunakan metode K-Means secara umum dilakukan dengan algoritma dasar 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

Pengalokasian kembali data ke dalam masing-masing *cluster* dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada. Data dialokasikan ulang secara tegas ke *cluster* yang mempunyai *centroid* terdekat dengan data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut [1]:

$$a_{ik} = \begin{cases} 1; d = \min\{D(x_k, v_i)\} \\ 0; \text{lainnya} \end{cases} \quad (1)$$

dimana:

a_{ik} = Keanggotaan data ke-k ke *cluster* ke-i

v_i = Nilai *centroidcluster* ke-i

1.2. EM

Expectation Maximation (EM) termasuk algoritma *partitional* yang berbasiskan model yang menggunakan perhitungan probabilitas, bukan jarak seperti umumnya algoritma *clustering* yang lainnya, langkah-langkahnya sebagai berikut [3]:

1. Inisialisasi *cluster centers* sebanyak suatu nilai k. Nilai k bisa merupakan input dari user atau merupakan nilai yang didapatkan dengan memanfaatkan penggunaan algoritma lain.
2. Iterasikan 2 langkah yang menjadi dasar penamaan algoritma berikut sampai mencapai suatu titik konvergensi yang telah ditentukan sebelumnya:
 1. langkah *expectation*
 2. langkah *maximization*

Sebagai algoritma yang berdasarkan pada probabilitas, EM memiliki beberapa karakteristik [3] yaitu:

1. Menghasilkan sistem *cluster* yang relatif mudah diinterpretasikan
2. Dapat diberhentikan dan dimulai kembali dengan kelompok data berurutan karena *clusters* memiliki representasi yang berbeda dengan items didalamnya.

1.3. Spatial Clustering

Sebenarnya konsep *spatialclustering* sama dengan *clustering*, hanya saja yang diamati adalah pengelompokan dengan konteks tempat. Menurut [5], *SpatialClustering* adalah proses pengelompokan sekumpulan objek ke dalam kelas atau kelompok sehingga objek dalam sebuah *cluster* memiliki kesamaan yang tinggi dibandingkan satu sama lain, tetapi berbeda jika dibandingkan dengan objek di *cluster* lain.

Pengertian *clustering* secara umum jika dilihat kepada pembagian data ke dalam *cluster* yaitu bahwa poin dalam *cluster* yang sama dipilih yang semirip mungkin, dan poin dalam kelompok yang berbeda dipilih sebagai yang paling berbeda. Dalam topik statistik terdahulu, algoritma-algoritma *clustering* biasanya digunakan untuk pengamatan *cluster*. Sedangkan dalam data spasial, *clustering* digunakan untuk mencari daerah dengan intensitas titik tinggi yang dipisahkan oleh daerah-daerah dengan intensitas rendah [4].

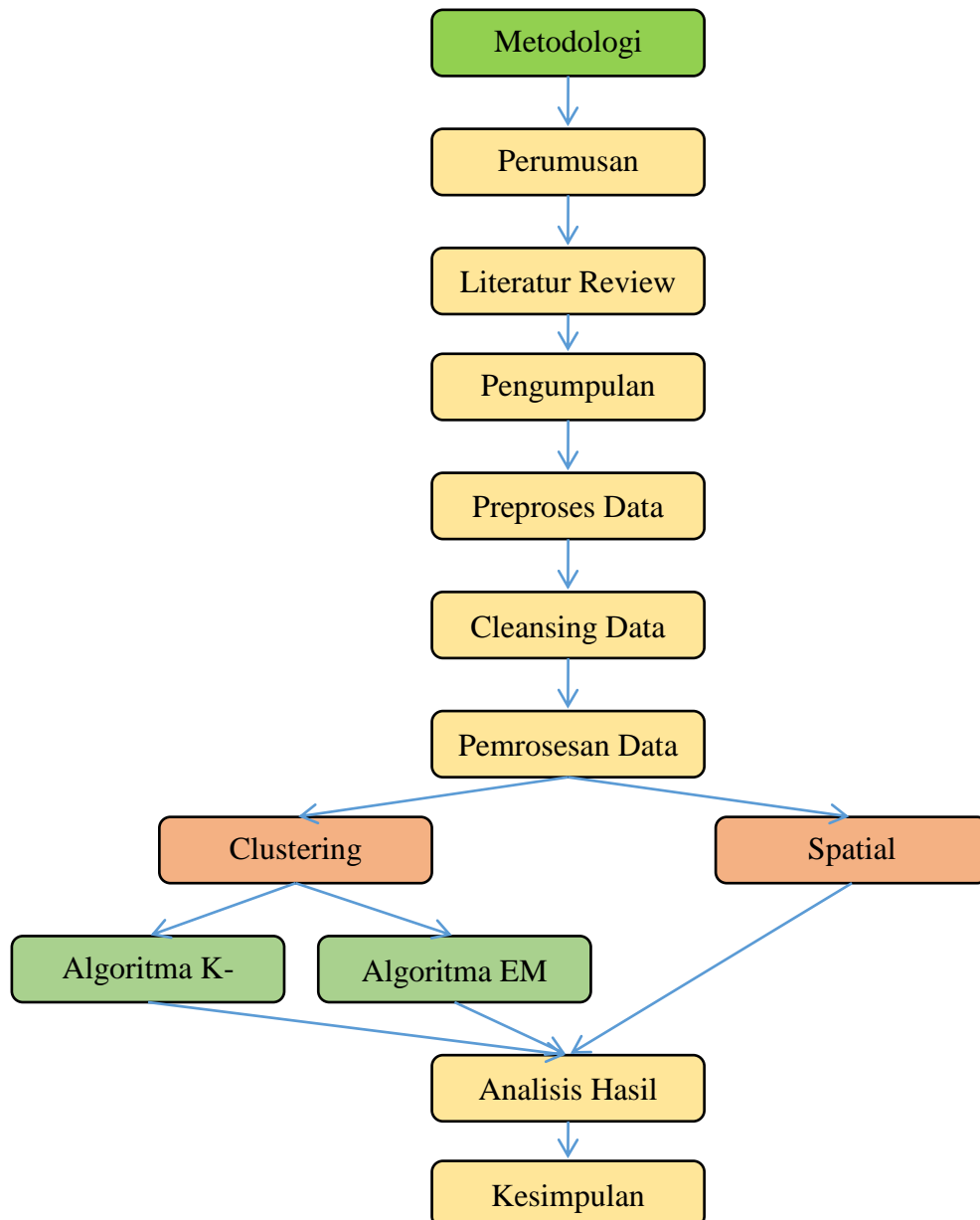
Penelitian mengenai fungsionalitas data mining *clustering* telah banyak dilakukan oleh para peneliti, diantaranya misalnya [3], yang membahas mengenai Aplikasi dan Analisis *Clustering* pada Data Akademik. Sementara untuk untuk *spatialclustering* sendiri pernah dibahas oleh [5] dengan judul penelitian Analisis Spasial dengan Metode *Density-BasedSpatialClusteringof Application with Noise* pada Basis Data Kriminal yang membandingkan Metode *Density-BasedSpatialClusteringof Application with Noise* dengan metode *clustering* lain yaitu *Gaussian Mixture Modeling*, *Fuzzy C-Means*, dan *K-Means*.

1.4. WEKA

WEKA (*Waikato Environment for Knowledge Analysis*) merupakan aplikasi *data miningopen source* berbasis Java [6]. Aplikasi ini dikembangkan pertama kali oleh sebuah Universitas di Selandia Baru yang bernama Universitas Waikato sebelum menjadi bagian di Pentanho. Weka terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling.

2. METODE PENELITIAN

Langkah-langkah yang dilakukan yaitu: perumusan masalah, penentuan teknik *clustering* yang akan dipergunakan, preproses data, transformasi data dengan teknik *clustering*, analisa hasil *clustering*, dan penarikan kesimpulan. Berikut digambarkan diagram tahapan penelitian yang digunakan:



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN


Data yang diambil yaitu data dari alumni STMIK MDP beserta asal sekolah SMA alumni-alumni untuk melihat kecenderungan SMA mana yang paling berpotensi menyalurkan alumninya ke STMIK MDP sekaligus melihat SMA mana yang memiliki alumni dengan IPK terbaik.

2.1. Proses Cleansing


Proses *cleansing* dilakukan karena data yang didapat masih belum bisa langsung digunakan. Proses ini meliputi berbagai tahapan, yaitu diantaranya adalah menyamakan *record* yang seharusnya sama tapi tercetak berbeda dikarenakan proses pada saat *input* data. Misalnya

untuk asal sekolah, dikarenakan diketik dan bukannya dipilih, untuk sekolah yang sama misal SMA Negeri 1 Palembang tercetak berbeda-beda, yaitu SMAN 1 Palembang, SMAN 1 PLB, SMAN 1 PLG dan sebagainya. Proses lain yaitu menormalisasi tabel menjadi beberapa tabel kecil yang saling terhubung, selain itu juga data-data yang recordnya lebih dari satu diperbaiki dan data-data yang tidak berguna dihapus.

Proses *cleansing* akhirnya menghasilkan 3 tabel yaitu Tabel Lulusan (2062 record), Tabel Sekolah (177 record), dan Tabel Alumni (2062 record), *primary key* Tabel Lulusan adalah NPM, *primary key* Tabel Sekolah adalah kodesekolah, *primary key* Tabel Alumni adalah NPM+kodesekolah+ipk dengan dua *foreign key* yaitu NPM ke Tabel Lulusan, dan kodesekolah ke Tabel Sekolah, ditunjukkan berikut ini:

	Column Name	Data Type	Allow Nulls
	npm	nvarchar(50)	<input type="checkbox"/>
	nama	nvarchar(50)	<input checked="" type="checkbox"/>

Gambar 2. Tabel Dimensi Lulusan

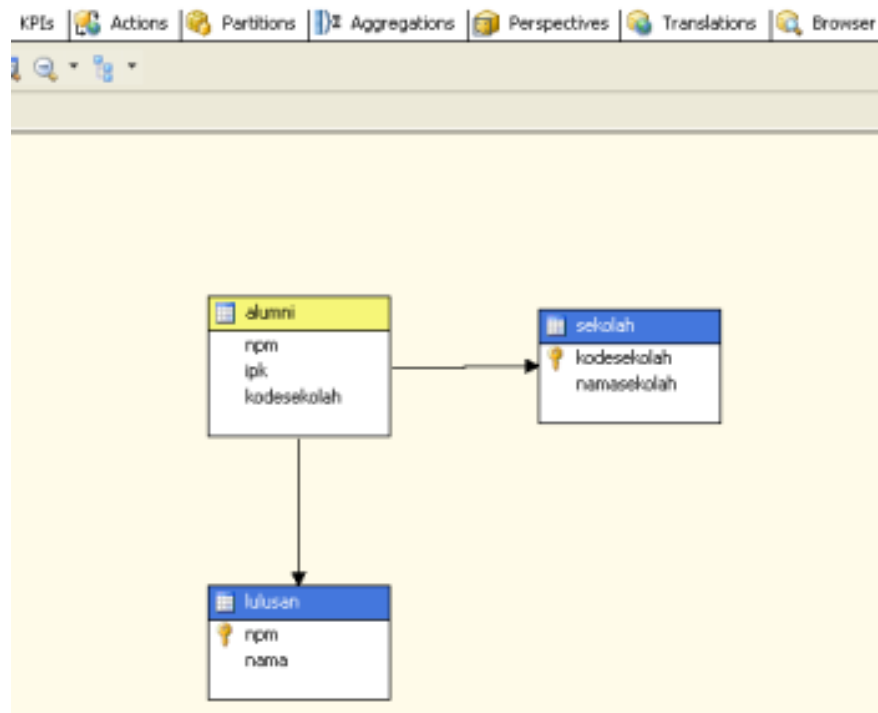
	Column Name	Data Type	Allow Nulls
	kodesekolah	nvarchar(50)	<input type="checkbox"/>
	namasekolah	nvarchar(50)	<input type="checkbox"/>

Gambar 3. Tabel Dimensi Sekolah

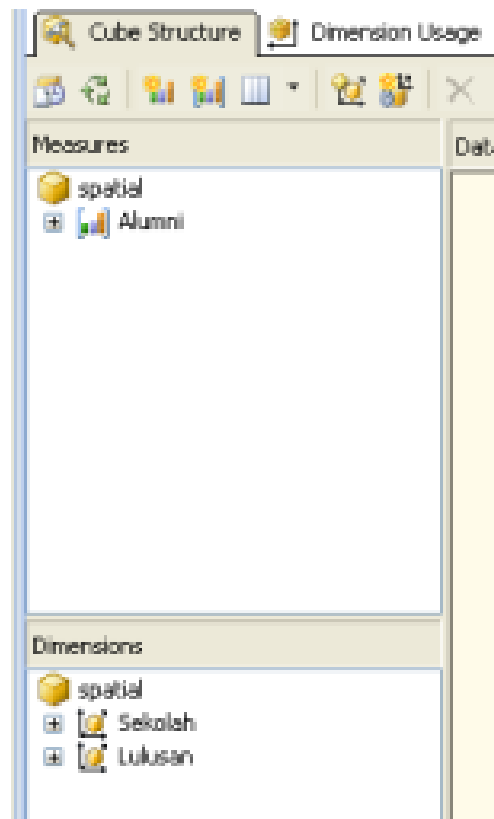
	Column Name	Data Type	Allow Nulls
	npm	nvarchar(50)	<input type="checkbox"/>
	ipk	float	<input checked="" type="checkbox"/>
	kodesekolah	nvarchar(50)	<input checked="" type="checkbox"/>

Gambar 4. Tabel Dimensi Alumni

Untuk mendapatkan tabel yang memuat jumlah alumni dari masing-masing sekolah beserta IPK rata-ratanya, dilakukan pengolahan lebih lanjut dengan menggunakan SQL Server 2008. Tabel alumni dari SQL Server 2008 diambil untuk pembuatan *Data Sources*, *Data Source Views*, *Cubes* dan *Dimensions*. Dari Tabel Alumni diperlihatkan pemetaan IPK dan asal sekolah dari mahasiswa-mahasiswa STMIK MDP yang telah lulus, untuk mendapatkan *cluster* tersebut, terlebih dahulu dibuat kubus untuk data alumni, berikut skema dari kubus yang terbentuk disajikan dalam gambar 5 dan 6:



Gambar 5. Struktur Kubus

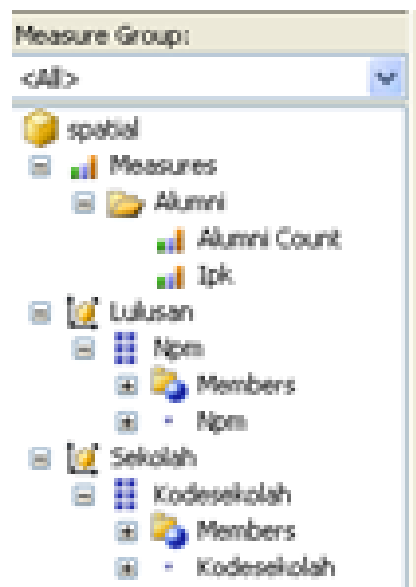


Gambar 6. Struktur Kubus untuk Skema Data Alumni

Dengan browser yang terbentuk dari Kode Sekolah dan IPK disajikan pada gambar 7 dan 8:

Kodesekolah	2008260503		2008260504		2008260505		Grand Total	
	Alumni Count	Ipik	Alumni Count	Ipik	Alumni Count	Ipik	Alumni Count	Ipik
sekolah1							1	2.58620609655172
sekolah10							9	24.8452134100993
sekolah100							9	25.937656641604
sekolah101							1	2.69354838709677
sekolah102							16	49.199951910387
sekolah103							1	2.6551724137931
sekolah104							1	3.15068493150685
sekolah105							2	5.18734335882456
sekolah106							5	15.9572218216775
sekolah107							1	2.92857142857143
sekolah108							3	8.3859649122807
sekolah109							3	8.5046992481203
sekolah11							1	2.81081081081081
sekolah110							1	3.30555555555556
sekolah111							1	3.68918918918919
sekolah112							1	2.66666666666667
sekolah113							7	20.1051999405089
sekolah114							6	15.7290307347506
sekolah115							2	5.19298245614035
sekolah116							1	2.50877192982456
sekolah117							10	30.1440764072343
sekolah118								

Gambar 7. Browser Data Alumni



Gambar 8. Measure Groups Browser Data Alumni

2.2. Analisis Spatial Clustering untuk Cluster Data Alumni

Hasil dari pengelompokan nilai alumni STMIK MDP berdasarkan IPK dan asal sekolah menghasilkan 176 record, yang terdiri atas berbagai sekolah-sekolah dengan mayoritas terletak di Palembang, beberapa di Sumatera Selatan dan sekolah lain di luar Propinsi lain di Indonesia. Pemetaan *spatial clustering* biasanya berbentuk geografis, misalnya dengan aplikasi DIVA, namun karena *softcopy* Propinsi Sumatera Selatan khususnya Palembang cukup sulit didapatkan dan pemetaan sekolah berdasarkan posisinya juga sukar dilakukan maka dilakukan

pengelompokan dengan *filter* sekolah dengan rata-rata nilai IPK paling tinggi sampai ke yang paling rendah saja.

Hasil *clustering* Algoritma K-Means dan EM dibagi menjadi beberapa kelompok. Dari sekolah-sekolah tersebut, nilai IPK dikelompokkan menjadi 3 *cluster* dengan ketentuan tinggi, sedang dan rendah dengan jumlah sekolah (dengan algoritma K-Means) untuk kategori ipk rata-rata rendah 48 sekolah (*Cluster 1*), sedang 94 sekolah (*Cluster 0*) dan tinggi 34 sekolah (*Cluster 2*). Sementara dengan algoritma EM memberikan hasil jumlah ipk mahasiswa yang rendah 49 sekolah (*Cluster 0*), sedang 122 sekolah (*Cluster 2*) dan tinggi 5 sekolah (*Cluster 1*). Diharapkan dengan hasil ini, maka akan memberikan gambaran bagi manajemen untuk lebih melihat lagi strategi pemasaran dan kerjasama dengan sekolah-sekolah terkait berdasarkan hasil *cluster-cluster* tersebut.

Dari hasil ini dapat dilihat bahwa Algoritma K-Means lebih banyak mengelompokkan sekolah-sekolah ke kategori tinggi dibanding EM, sementara untuk algoritma EM lebih banyak mengelompokkan sekolah-sekolah ke kategori sedang, sedangkan untuk kategori rendah, antara algoritma K-Means dan EM cenderung berimbang dalam pengelompokannya.

Jika data sekolah-sekolah tersebut dibagi berdasarkan dengan data tahun keluar dari mahasiswa, maka *cluster-cluster* yang terbentuk akan menjadi sedikit berbeda. Pada prosesnya dengan menggunakan SQL Server 2008 diperlukan satu kolom lagi dari tabel alumni yang menunjukkan tahun keluar dari alumni. Kolom ini bisa didapat dari tabel yang baru yaitu tabel dataijazah. Berikut perintah query dari penggabungan kedua tabel tersebut:

```
SELECT          alumni.npm,      alumni.ipk,      alumni.kodesekolah,
DATAIJAZAH.noijazah, DATAIJAZAH.tglkeluar
INTO            alumnipertahun
FROM            alumni INNERJOIN
DATAIJAZAH ON  alumni.npm = DATAIJAZAH.npm
```

Hasil dari penggabungan tersebut menghasilkan tabel dimensi alumni berdasarkan tahun keluar seperti disajikan pada gambar 8:

Column Name	Data Type	Allow Nulls
npm	nvarchar(50)	<input type="checkbox"/>
ipk	float	<input checked="" type="checkbox"/>
kodesekolah	nvarchar(50)	<input checked="" type="checkbox"/>
Tahun	nvarchar(50)	<input checked="" type="checkbox"/>

Gambar 8 Tabel Dimensi Alumni Berdasarkan Tahun Keluar

Selanjutnya dari tabel tersebut, data di tarik ke aplikasi WEKA untuk mendapatkan *cluster* juga dengan menggunakan dua algoritma yaitu algoritma K-Means dan EM. Kedua algoritma ini sudah tersedia pada WEKA dan bisa langsung digunakan. Hasil *output* dari kedua algoritma ini berlainan dan nantinya akan dibandingkan.

Dari sekolah-sekolah tersebut, nilai IPK dikelompokkan menjadi 3 *cluster* dengan ketentuan tinggi, sedang dan rendah dengan jumlah sekolah (dengan algoritma K-Means) untuk kategori ipk rata-rata rendah 24 sekolah (*Cluster 0*), sedang 51 sekolah (*Cluster 1*) dan tinggi 101 sekolah (*Cluster 2*) sementara dengan algoritma EM memberikan hasil jumlah ipk mahasiswa yang rendah 25 sekolah (*Cluster 0*), sedang 34 sekolah (*Cluster 2*) dan tinggi 117 sekolah (*Cluster 1*).

Dari hasil ini dapat dilihat bahwa Algoritma K-Means dan EM cenderung berimbang dalam pengelompokan masing-masing *cluster*, dibanding jika menggunakan data secara kumulatif.

Mengingat penelitian ini hanya terbatas penyebutan dan pencarian lokasi tanpa pemetaan *spatial* secara geografis seperti disebutkan sebelumnya, penelitian berikutnya dapat

meneruskan dengan menggunakan konsep *Geografis Information System* (GIS) agar lebih jelas terlihat pemetaannya.

4. KESIMPULAN

1. Dari hasil output yang dihasilkan pada tahapan clustering menggunakan algoritma K-Means dan EM, sekolah (SMA) yang memiliki alumni dengan IPK tertinggi adalah SMA Stella Duce 1 dan yang terendah adalah SMA Bhakti Ibu 2 Sekayu. Namun kedua sekolah itu hanya menyumbangkan satu alumninya. Sementara SMA yang memiliki alumni terbanyak yang pernah kuliah di STMIK MDP adalah SMA Xaverius 3 (IPK rata-rata 3.03) dengan 366 alumni dan SMA Xaverius 1 dengan jumlah alumni sebanyak 320 (IPK rata-rata 3,22).
2. Dari analisis hasil, didapatkan algoritma terbaik untuk melakukan clustering untuk kasus-kasus diatas jika dibandingkan antara Algoritma K-Means dan EM adalah Algoritma K-Means, karena dilihat dari nilai-nilai koefisien Silhouette yang telah didapatkan, nilai-nilai koefisien silhouette pada Algoritma K-Means lebih banyak yang mendekati nilai 1 dibanding algoritma EM. Ini berarti pengelompokan cluster pada algoritma K-Means lebih baik dari algoritma EM.

5. SARAN

1. Bagi manajemen STMIK MDP agar dapat lebih memberikan perhatian lebih kepada sekolah-sekolah yang berpotensi bagi STMIK MDP, yang memiliki alumni dengan nilai yang baik dan sekolah mana yang memberikan alumni dengan jumlah terbanyak dan rata-rata IPK yang cukup baik.
2. Mengingat penelitian ini hanya terbatas penyebutan dan pencarian lokasi tanpa pemetaan spasial secara geografis, penelitian berikutnya dapat meneruskan dengan menggunakan konsep *Geografis Information System* (GIS) agar lebih jelas terlihat pemetaannya.

DAFTAR PUSTAKA

- [1] Agusta, Y., 2007, K-Means- Penerapan, Permasalahan dan Metode Terkait, *Jurnal Sistem dan Informatika*, vol 3, hal 47-60.
- [2] Ayub, M., 2007, Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer, *Jurnal Sistem Informasi*, vol 2, no 1, hal 21-30.
- [3] Budiarti, A., 2006, Aplikasi dan Analisis Clustering pada Data Akademik, *Laporan Tugas Akhir*, Fakultas Ilmu Komputer, Universitas Indonesia.
- [4] Han, J., Kamber, M., Tung, A. K. H., 2001, *Spatial Clustering Methods in Data Mining: A Survey*, School of Computing Science Simon Fraser University Burnaby, Canada.
- [5] Lidyawati, P., 2010, Analisis Spasial dengan Metode Density-Based Spatial Clustering of Application with Noise pada Basis Data Kriminal, *Skripsi*, Universitas Indonesia.
- [6] Sulianta, F., Juju, D., 2010, *Data Mining: Meramalkan Bisnis Perusahaan*, Elex Media Komputindo, Jakarta.
- [7] Susanto, S., Ernawati, 2007, Pembagian Kelas Peserta Kuliah Berdasarkan Fuzzy Clustering dan Partition Coefficient and Exponential Separation (PCAES) Index, *Jurnal Teknologi Industri*, Universitas Atmajaya, Yogyakarta.