

ANALISIS ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PRESTASI SISWA STUDI KASUS SMP NEGERI I SUKAHENING

Ai Ilah Warnilah
AMIK BSI Jakarta
aillahwarnilah@gmail.com

Abstract - The high success rate of students and the low level of student failure is a reflection of the quality of education. Education today are required to have the ability to compete by utilizing all available resources. In addition to resource infrastructure, facilities and people, information systems is one of the resources that can be used to improve the ability to compete. Data mining is the process of analyzing the data to find a pattern of the data set. Data mining is able to analyze a large amount of supporting data into information in the form that has meaning for decision support. One of the clustering process of data mining is one of the methods called *k-means*. *K-Means* algorithm is the simplest clustering algorithm than other clustering algorithms. Atribut group student achievement are the Name, extracurricular, value includes the value Skills Knowledge, Attitude value, and the number of absences students. From the results of a case study of 173 students obtained with manhattan distance, chbychep distance euclidian distance the result accuracy 67 %.

Keywords: Clustering, K-means, student achievement

Abstrak - Tingginya tingkat keberhasilan siswa dan rendahnya tingkat kegagalan siswa merupakan cermin kualitas dunia pendidikan. Dunia pendidikan saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana dan manusia, sistem informasi merupakan salah satu sumber daya yang dapat digunakan untuk meningkatkan kemampuan bersaing. *Data mining* merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data. *Data mining* mampu menganalisa jumlah data yang besar menjadi informasi yang mempunyai arti bagi pendukung keputusan. Salah satu proses *data mining* adalah *clustering*. Atribut yang digunakan dalam pengelompokan prestasi siswa adalah nama, ekstrakurikuler, nilai pengetahuan yang meliputi nilai keterampilan, nilai sikap, dan jumlah ketidakhadiran siswa. Studi kasus pada 173 siswa dengan perhitungan jarak menggunakan *manhattan distance*, *chbychep distance* dan *euclidian distance* menghasilkan akurasi sebesar 67%.

Kata Kunci: Clustering, k-means, Prestasi siswa

1. PENDAHULUAN

Dewasa ini pengolahan data elektronik telah menjadi kebutuhan yang sangat utama. Perkembangan pesat dalam teknologi informasi yang menjadikan semua informasi dapat disimpan dalam jaringan komputer telah membuat munculnya sistem basis data yang sangat besar. Dalam hitungan detik, data - data dalam berbagai basis data akan senantiasa diperbarui, baik dikarenakan adanya *update* maupun penambahan data baru. Permasalahan yang kemudian muncul adalah

bagaimana mengetahui informasi yang terdapat dalam basis data yang sangat besar.

Knowledge discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses *knowledge discovery* melibatkan hasil dari proses data mining (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami.

Ada beberapa macam pendekatan berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik and statistik. Beberapa pendekatan memanfaatkan teknik *visualisasi*, pendekatan klasifikasi seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan syaraf tiruan dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada.

Tingginya tingkat keberhasilan siswa dan rendahnya tingkat kegagalan siswa merupakan cermin kualitas dunia pendidikan. Dunia pendidikan saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana dan manusia, sistem informasi merupakan salah satu sumber daya yang dapat digunakan untuk meningkatkan kemampuan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Metode pengambilan keputusan konvensional yang ada, tidak dapat menangani data dalam jumlah yang sangat besar. Hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi yang penting dari kumpulan data, yang disebut dengan data mining.

Secara umum penilaian prestasi siswa yaitu mata pelajaran baik teori dan praktek, penilaian ekstrakurikuler kehadiran dan ketidakhadiran siswa dan pada saat mengikuti ekstrakurikuler, yang dikategorikan dalam nilai pengetahuan, nilai keterampilan dan penilaian sikap. Evaluasi dan penilaian terhadap prestasi siswa dilakukan dengan pemberian nilai oleh pengajar kepada semua siswa yang mengikuti pelajaran yang diajarkan dan ekstrakurikuler yang diikutinya.

Seiring dengan terus bertambahnya jumlah data siswa setiap tahun, maka jumlah data yang siswa yang terus meningkat sehingga penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan

pengetahuan baru melalui pola - pola yang terbentuk dari penumpukan data tersebut. Jumlah data yang terus meningkat ini memerlukan beberapa teknik ataupun metode untuk mengolahnya menjadi sebuah informasi dan pengetahuan yang dapat digunakan sebagai bahan pertimbangan pendidik dalam proses pengambilan kebijakan dan keputusan juga sebagai peringatan dini (*early warning*) bagi siswa tertentu yang berdasarkan hasil pengelompokan prestasi rendah yang berpotensi terhadap ketidak lulusan siswa.

Beberapa penelitian yang telah dilakukan melakukan pemodelan mengenai pemodelan aturan dalam memprediksi akademik mahasiswa, mengevaluasi Kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak, konsep pengclasteran dengan pola data yang sudah diatur, tehnik cluster membuat pengelompokan data iklim tropis di samudera hindia bagian utara.

2. KAJIAN LITERATUR

Berikut ini adalah tinjauan studi terdahulu yang terkait dengan tema K-Means clustering digunakan untuk membantu menyelesaikan tesis ini, meliputi:

1. Zefri Faulanda, 2012 melakukan penelitian mengenai model profil mahasiswa yang potensial Drop Out Menggunakan Teknik *Kernel K-Means Clustering* dan *Decision Tree* dalam penelitian ini membangun model profil mahasiswa yang cenderung Drop Out berdasarkan atribut Minat Belajar, Kepercayaan diri, Prilaku Belajar, Dukungan Orang Tua Dan Predikat. Berdasarkan kelima atribut tersebut yang menyebabkan mahasiswa Drop Out disebabkan karena tidak ada lagi minat belajar mahasiswa, kurangnya faktor dukungan orang tua, kurangnya kepercayaan diri juga perilaku dan waktu belajar mahasiswa.
2. Hikmah Adwin Adam, 2012 melakukan penelitian mengenai pemodelan aturan dalam memprediksi akademik mahasiswa politeknik negeri medan dengan kernel k-means clustering

dalam penelitian ini membangun model aturan dalam memprediksi prestasi akademik mahasiswa, diperoleh suatu model aturan yang dapat digunakan memprediksi prestasi mahasiswa, dalam cluster ini diperoleh mahasiswa yang memiliki predikat dengan pujian apabila nilai rata-rata teori tinggi (>70), nilai rata-rata praktek tinggi (>70) dan kehadiran baik (>93%)

- Mujib Ridwan, Hadi Suyono, dan M. Sarosa 2013, Penelitian ini difokuskan untuk mengevaluasi Kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang dapat lulus tepat waktu atau tidak. Kemudian dari klasifikasi tersebut, sistem akan memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat dengan nilai optimal berdasarkan histori nilai yang telah ditempuh mahasiswa. input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa. Sampel mahasiswa angkatan 2005-2009 yang sudah dinyatakan lulus akan digunakan sebagai data training dan testing. Sedangkan data mahasiswa angkatan 2010-2011 dan belum lulus akan digunakan sebagai data target. Data input akan diproses menggunakan teknik data mining algoritma *Naive Bayes classifier*(NBC) untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi kelulusan mahasiswa. output dari sistem ini berupa klasifikasi kinerja akademik mahasiswa yang diprediksi kelulusannya dan memberikan rekomendasi untuk proses kelulusan tepat waktu atau lulus dalam waktu yang paling tepat dengan nilai optimal
- Capt.Dr.S.Santhosh Baboo,Tajudin, 2013 Penelitian ini adalah membuat konsep pengclasteran dengan pola data yang sudah diatur, tehnik cluster membuat pengelompokan data iklim tropis di samudera hindia bagian utara. Pengumpulan data tropsis mulai dari tahun 2001 sampai dengan 2013. Anggota group dari hasil analisa clustering menghasilkan nilai maksimum dan minimum setiap tahun. Cluster pertama menghasilkan

serangkaian jarak angin yang maksimum yang terjadi setiap tahunnya. Cluster kedua menggambarkan jarak angin minimum dari setiap tahun. Data tersebut tersimpan dalam database dengan menggunakan algoritma cluster

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode kuantitatif untuk mengelompokan prestasi siswa studi kasus SMP N 1 Sukahening untuk menganalisa pengelompokan data siswa berprestasi kategori prestasi paling tinggi, cukup dan menengah. Data yang diolah sebanyak 173 siswa.

A. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah model penelitian eksperimen. Model eksperimen merupakan metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu (Sugiyono, 2010). Jenis penelitian yang dilakukan pada tesis ini yaitu menggunakan metode *k-means* untuk mengelompokan prestasi dari data peserta didik SMP N 1 Sukahening. Jenis ini mengambil algoritma *k-means* sebagai fungsi pengelompokan, yang merupakan ekstraksi pola yang menarik dari data dalam jumlah besar

Kegiatan penelitian ini melalui beberapa tahap dalam pengembangannya yaitu:

1. Studi Pendahuluan

Kegiatan yang dilakukan pada saat studi pendahuluan yaitu mengumpulkan materi-materi kepustakaan yang berhubungan dengan pengambilan judul. Dari penelitian tersebut akan di dapat beberapa kriteria yang digunakan untuk penelitian lebih lanjut

2. Pengumpulan Data

Melakukan pengambilan data dari instansi terkait yang akan digunakan dalam penelitian ini, yang terdiri dari beberapa atribut yaitu data-data masukan berupa data peserta didik kelas 7 yang terdiri dari kelas 7A, 7B, 7C, 7D, 7E dan 7F yang digabungkan menjadi satu data yang berjumlah 173 peserta didik.

3. Pengolahan Data

Setelah mendapatkan data. Maka data diolah menggunakan metode data

mining, yaitu pertama kali menggunakan Microsoft Excel.

4. Laporan

Setelah tahap studi pendahuluan, data, dan pengolahan data tersebut dilakukan maka disusunlah laporan penelitian ini kedalam bentuk jurnal ilmiah

B. Metode Pengumpulan Data

Data yang diolah dalam penelitian ini merupakan data sampel yang diambil dari data raport peserta didik SMP kelas VII tahun ajaran 2011/2012. Dataset mahasiswa terdiri dari atribut data induk dan data raport. Atribut yang dijadikan variable dalam algoritma ini di sederhanakan menjadi atribut yang terdiri dari Nama, ekstrakurikuler, Nilai Pengetahuan, nilai keterampilan, nilai Sikap, dan jumlah ketidakhadiran. Data sampel yang akan diuji cobakan terdiri dari 173 peserta didik.

C. Ruang Lingkup

Sejumlah permasalahan yang dibahas dalam usulan penelitian ini akan dibatasi ruang lingkup pembahasannya, antara lain:

1. Data-data masukan merupakan mahasiswa aktif angkatan 2011/2012 pada sekolah SMP Negeri 1 Sukahening
2. Variabel yang digunakan dalam penelitian ini merupakan atribut nama, Jumlah ekstrakurikuler, Jumlah Absensi ekstrakurikuler, Jumlah Nilai Mata Pelajaran, dan jumlah absensi harian

4. PEMBAHASAN

A. Metode Pemilihan data

Data yang diolah dalam penelitian ini merupakan data sampel yang diambil dari data raport peserta didik SMP kelas VII tahun ajaran 2014-2015. Dataset mahasiswa terdiri dari atribut data induk dan data raport. Atribut yang dijadikan variable dalam algoritma ini di sederhanakan menjadi atribut yang terdiri dari nama, ekstrakurikuler, nilai pengetahuan, nilai keterampilan, nilai sikap, dan jumlah ketidakhadiran. Data sampel yang akan diuji cobakan terdiri dari 173 peserta didik

Data awal tersebut tampak pada Tabel 1.

Tabel 1. Data Peserta didik SMPN1 Sukahening

No	Nama	Ekstrakurikuler						Jumlah			Ketidakhadiran		
		E1	E2	E3	E1	E2	E3	P	K	Sikap	S	I	A
1	AMA	Pramuka	PKS	-	B	A	-	31.94	33.04	33	-	-	-
2	ASN	Pramuka	-	-	A	-	-	36.91	36.91	37.7	-	-	-
3	AMU	Pramuka	-	-	B	-	-	32.16	33.09	33	1	-	-
4	ANN	Pramuka	-	-	B	-	-	32.99	33.4	33.6	-	-	-
5	ASI	Pramuka	PKS	-	B	A	-	33.14	34.62	35.3	1	-	-
6	ASA	Pramuka	-	-	B	-	-	32.97	34.2	34	1	-	1
.
.
.
168	SEN	Pramuka	PKS	-	B	B	-	34.88	37.39	35.9	-	-	-
169	SJY	Pramuka	-	-	B	-	-	31	32.44	32.7	-	-	-
170	SSH	Pramuka	PMR	-	B	B	-	33.57	35.36	33.4	-	-	-
171	THA	Pramuka	PMR	-	B	B	-	34.2	35.87	33.7	-	-	-
172	WWF	Pramuka	PKS	-	B	B	-	36.09	36.79	36.4	-	-	-
173	YHI	Pramuka	-	-	B	-	-	31.01	32.23	32.8	-	-	-

B. Transformasi Data

Data awal ditransformasikan menjadi data nominal untuk variable seperti ekstrakurikuler dan nilai berupa hurup, dari data awal dikelompokan pada tiap-tiap bidang ekstrakurikuler dan

dijumlahkan anggotanya, data tersebut dijadikan sebuah master kemudian diinisialisasikan dengan angka sesuai dengan urutan banyaknya peminat seperti pada Tabel 2.

Tabel 2. Data Master ekstrakurikuler

Nama Ekstrakurikuler	frekuensi	inisialisasi
Pramuka	173	1
Palang Merah Ramaja (PMR)	25	2
PKS	24	3
Paskibra	15	4
Marchingband	5	5
Futsal	4	6
Marawis	4	7
Bola Volly	1	8
Rampak Kembang	1	9

Penilaian ekstrakurikuler diinisialisasikan menjadi data seperti pada Tabel 3.

Tabel 3. Data Master Nilai

Huruf	Nilai
A	4
B	3
C	2

Dari data induk peserta didik dan data kumpulan nilai berupa ilmu pengetahuan Agama, PPKN, B. Indo, Mat, IPA, IPS, Bahasa Inggris, Seni Budaya, Penjas-Orkes, Prakarya, Bahasa Sunda yang mempunyai nilai pengetahuan, keterampilan dan sikap pada setiap bidang, digabungkan menjadi 3 variable yaitu jumlah semua nilai pengetahuan, keterampilan dan sikap, yang kemudian digabungkan menjadi total nilai yang akan dipakai menjadi variable penentuan prestasi peserta didik pada algoritma *k-means*.

C. Metode Analisis dan Pengujian

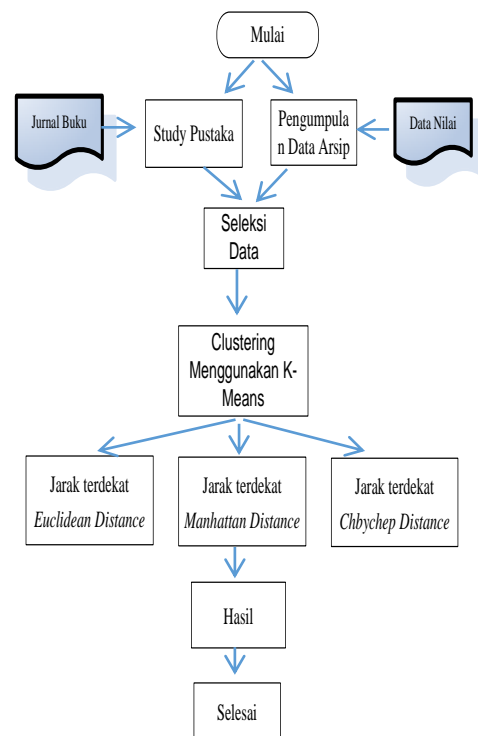
Pengelompokan prestasi peserta didik dari data yang diperoleh dalam penelitian ini menggunakan algoritma *k-means* clustering. Data Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (nonhirarki) data clustering. *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 data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain

Tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster.

D. Langkah-langkah penelitian

Gambar 1. menunjukkan langkah-langkah yang akan diambil dalam penelitian ini.



Gambar 1. Langkah-Langkah Penelitian

Langkah-langkah dalam penelitian ini sebagai berikut:

1. Studi pustaka dan studi penelitian sebelumnya, Studi terhadap penelitian-penelitian yang pernah dilakukan sebelumnya, terutama yang menggunakan metode data mining, yaitu Metode *k-means*. (yang

2. Pengumpulan data, Mengumpulkan data peserta didik dalam beberapa kelas dengan variasi nilai dan kreatifitas.
3. Eksperimen, Pada tahap ini, data training dilatihkan pada algoritma data mining k-means secara manual dan bertahap sampai iterasi terakhir yang memungkinkan data tidak berpindah lagi.
4. Analisa Data, Hasil dari proses eksperimen dilakukan analisa, dengan menganalisa hasil awal cluster pusat yang ditentukan secara random dan pengelompokkan data hasil proses algoritma dengan menghubungkan atribut-atribut yang sebelumnya telah disederhanakan.
5. *Reporting*, menuliskan hasil penelitian dalam bentuk tesis dari pendahuluan sampai dengan kesimpulan

E. Pengolahan Data

Dari data yang diambil sebagai sampel eksperimen, diperoleh data-data yang akan dijadikan variable dalam perhitungan algoritma k-means, dari keseluruhan data induk peserta didik di sekolah, data disederhanakan dalam beberapa variable menjadi nama, ekstrakurikuler, nilai pengetahuan, nilai keterampilan, nilai Sikap, dan jumlah ketidakhadiran. Kemudian data diolah kembali menjadi data yang siap diproses pada Tabel 3.1.

Data tersebut masih berupa karakter seperti pada data ekstrakurikuler, dan nilainya, kemudian data tersebut ditransformasikan menjadi sebuah nilai atau angka yang dapat digunakan dengan menginisialisasikan dengan angka sesuai dengan urutan banyaknya peminat, data transformasi menjadi seperti terlihat pada Tabel 4.

Penilaian ekstrakurikuler diinisialisasikan menjadi data seperti pada Tabel 3.:

Setelah data disederhanakan dan ditransformasikan, data diolah kembali menjadi data *numerical*, terdapat variable ekstrakurikuler dengan macam-macam ekstrakurikuler yang dapat dipilih para peserta didik, masing-masing peserta didik minimal memilih satu dan maksimal tiga jenis ekstrakurikuler, semua ekstrakurikuler dan nilai ekstrakurikuler

diinisialisasikan berdasarkan data di atas, kemudian dijumlahkan menjadi variable ekstrakurikuler dan nilai ekstrakurikuler. Begitupun dengan ketidakhadiran dijumlahkan menjadi satu variable. Hasil akhir data yang didapat seperti pada Tabel 5.

Tabel 4. Data Master Ekstrakurikuler

Ekstrakurikuler		
Nama Ekstrakurikuler	frekuensi	inisialisasi
Futsal	4	5
Paskibra	15	3
Palang Merah Ramaja (PMR)	25	1
Marchingband	5	4
Marawis	4	6
PKS	24	2
Bola Volly	1	7
Rampak Kembang	1	8

Tabel 5. Data yang sudah disederhanakan

id	Eskul	Nilai Eskul	Jumlah IP	Absen Harian
1	4	7	97.98	0
2	1	4	111.5	0
3	1	3	98.28	1
4	1	3	100.02	0
5	4	7	103.09	1
6	1	3	101.17	2
7	1	3	99.73	0
8	1	3	96.26	0
9	1	3	100.91	0
10	1	3	100.38	1
:	:	:	:	:
165	1	6	100.76	0
166	1	6	103.02	0
167	1	3	97.98	0
168	4	6	108.21	0
169	1	3	96.09	0
170	1	6	102.32	0
171	1	6	103.77	0
172	4	6	109.29	0
173	1	3	96.04	0

Data yang sudah diinisialisasikan kemudian ditransformasikan dan diolah

kembali sekarang dapat digunakan dalam algoritma *k-means*.

F. Algoritma K-Means

Langkah pertama algoritma *k-means* adalah menentukan jumlah *cluster*, pada penelitian ini ditetapkan 3 *cluster*, yang dipilih secara random. Dengan variable jumlah ekstrakurikuler, nilai ekstrakurikuler, jumlah ilmu pengetahuan dan absen harian. Data yang dihasilkan tampak pada Tabel 6.

Tabel 6. Cluster Awal

	Jumlah Eskul	Nilai Eskul	Jumlah IP	Absen Harian
Cluster 1	3	6	107.05	0
Cluster 2	8	10	103.77	0
Cluster 3	5	6	98.78	6

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (4.1)$$

Setelah didapatkan data *cluster*, kemudian langkah kedua menghitung jarak setiap data ke pusat *cluster* antara objek ke centroid. Untuk menghitung jarak pada penelitian ini digunakan L_2 (*Euclidean distance space*, jarak antara dua titik dihitung menggunakan rumus sebagai berikut:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + (x_{23} - x_{13})^2 + (x_{24} - x_{14})^2}$$

Dimana p = dimensi data, berdasarkan data p berjumlah 4. Maka dijabarkan rumus menjadi sebagai berikut:

Sebagai contoh perhitungan jarak beberapa data adalah sebagai berikut:

$$D_{L_2}(1,1) = \sqrt{(4-3)^2 + (7-6)^2 + (97,98 - 107,05)^2 + (0-0)^2}$$

- Dari data pertama ke pusat cluster 1

$$D_{L_2}(1,1) = \sqrt{(4-3)^2 + (7-6)^2 + (97,98 - 107,05)^2 + (0-0)^2}$$

$$D_{L_2}(1,1) = \sqrt{(1)^2 + (1)^2 + (-9,07)^2 + (0)^2}$$

$$D_{L_2}(1,1) = \sqrt{1+1+82,26}$$

$$D_{L_2}(1,1) = \sqrt{84,26}$$

$$D_{L_2}(1,1) = 9,18$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data pertama dengan *cluster* pertama adalah 9,18.

- Dari data pertama ke pusat cluster 2

$$D_{L_2}(1,2) = \sqrt{(4-8)^2 + (7-10)^2 + (97,98 - 103,77)^2 + (0-0)^2}$$

$$D_{L_2}(1,2) = \sqrt{(-4)^2 + (-3)^2 + (-5,79)^2 + (0)^2}$$

$$D_{L_2}(1,2) = \sqrt{16+9+33,52}$$

$$D_{L_2}(1,2) = \sqrt{58,52}$$

$$D_{L_2}(1,2) = 7,65$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data pertama dengan *cluster* kedua adalah 7,65.

- Dari data pertama ke pusat cluster 3

$$D_{L_2}(1,3) = \sqrt{(4-5)^2 + (7-6)^2 + (97,98 - 98,78)^2 + (0-6)^2}$$

$$D_{L_2}(1,3) = \sqrt{(-1)^2 + (-1)^2 + (-0,8)^2 + (-6)^2}$$

$$D_{L_2}(1,3) = \sqrt{1+1+0,64+36}$$

$$D_{L_2}(1,3) = \sqrt{38,64}$$

$$D_{L_2}(1,3) = 6,21$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data pertama dengan *cluster* ketiga adalah 6,21.

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data pertama yang paling dekat adalah *cluster* 3, sehingga peserta didik pertama dimasukkan ke dalam *cluster* 3. Hasil perhitungan selengkapnya dapat dilihat pada Tabel 7.

Tabel 7. Jarak Data pertama dengan Pusat Cluster pada iterasi 1

N0	Es kul	Nilai Eskul	Jumlah IP	Absen Harian	cluste r 1	cluste r 2	cluste r 3	cluste r 1	cluste r 2	cluste r 3
1	4	7	97.98	0	9.18	7.65	6.22			*
2	1	4	111.5	0	5.27	12.03	14.76	*		
3	1	3	98.28	1	9.53	11.36	7.09			*
4	1	3	100.02	0	7.90	10.59	7.91	*		
5	4	7	103.09	1	4.32	5.14	6.75	*		
6	1	3	101.17	2	7.18	10.43	6.83			*
7	1	3	99.73	0	8.16	10.69	7.87			*
8	1	3	96.26	0	11.38	12.43	8.21			*

9	1	3	100.91	0	7.12	10.30	8.10	*		
10	1	3	100.38	1	7.65	10.51	7.25			*
11	1	3	98.77	1	9.09	11.14	7.07			*
12	1	3	104.65	0	4.33	9.94	9.77	*		
13	1	3	99.21	0	8.63	10.90	7.82			*
14	1	3	98.51	5	10.53	12.27	5.11			*
15	1	3	97.88	1	9.90	11.56	7.13			*
:	:	:	:	:	:	:	:	:	:	:
158	1	3	105.64	0	3.87	10.07	10.40	*		
159	1	3	104.87	0	4.21	9.96	9.90	*		
160	1	6	101.66	0	5.75	8.33	7.76	*		
161	1	3	99.23	0	8.61	10.89	7.82			*
162	1	3	101.56	0	6.57	10.14	8.29	*		
163	1	3	100.35	0	7.61	10.47	7.97	*		
164	1	6	103.76	0	3.85	8.06	8.76	*		
165	1	6	100.76	0	6.60	8.61	7.48	*		
166	1	6	103.02	0	4.50	8.10	8.37	*		
167	1	3	97.98	0	9.76	11.47	7.85			*
168	4	6	108.21	0	1.53	7.19	11.22	*		
169	1	3	96.09	0	11.54	12.53	8.26			*
170	1	6	102.32	0	5.14	8.19	8.03	*		
171	1	6	103.77	0	3.84	8.06	8.77	*		
172	4	6	109.29	0	2.45	7.90	12.14	*		
173	1	3	96.04	0	11.59	12.56	8.28			*

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut. Contoh perhitungannya adalah sebagai berikut :
Rata-rata ekstrakurikuler yang ada pada cluster 1:

$$= \frac{1+1+4+1+1+4+1+1+1+1+1+1+7+4+6+3+4+\dots+4+1+1+4}{66}$$

$$= \frac{170}{77} = 2,21$$

Rata-rata ekstrakurikuler yang ada pada cluster 2:

$$= \frac{8+7+6+8+8+6+1+1+1+0+8+6+5+8+6}{12}$$

$$= \frac{145}{20} = 7,25$$

Rata-rata ekstrakurikuler yang ada pada cluster 3:

$$= \frac{4+1+1+1+1+1+1+1+1+\dots+3+1+1+1+1+1+1+1+1+1+1+1+1}{66}$$

$$= \frac{163}{76} = 2,14$$

Rata-rata Nilai ekstrakurikuler yang ada pada cluster 1:

$$= \frac{4+3+7+3+3+6+3+3+\dots+3+3+6+6+6+6+6+6+6}{66}$$

$$= \frac{361}{77} = 4,69$$

Rata-rata nilai ekstrakurikuler yang ada pada cluster 2:

$$= \frac{10+9+6+10+10+9+9+9+6+9+6+6+10+9}{12}$$

$$= \frac{176}{20} = 8,80$$

Rata-rata nilai ekstrakurikuler yang ada pada cluster 3:

$$= \frac{7+3+3+3+3+3+3+3+\dots+3+6+3+3+3+3+3+3+3+3}{66}$$

$$= \frac{291}{76} = 3,83$$

Rata-rata Nilai Ilmu pengetahuan yang ada pada cluster 1:

$$= \frac{111.50+100.02+103.09+100.91+104.65+\dots+108.21+102.32+103.77+109.29}{66}$$

$$= \frac{8.049,59}{77} = 104,54$$

Rata-rata nilai ilmu pengetahuan ada pada cluster 2:

$$= \frac{104.28+101.91+102.31+103.77+101.83+\dots+101.34+102.10+103.18+106.05}{12}$$

$$= \frac{2.058,91}{20} = 102,95$$

Rata-rata nilai ilmu pengetahuan yang ada pada cluster 3:

$$= \frac{97.98+98.28+101.17+99.73+96.26+\dots+98.37+99.23+97.98+96.09+96.04}{66}$$

$$= \frac{7.342,85}{76} = 96,62$$

Rata-rata Absensi yang ada pada cluster 1:

$$= \frac{1+1+1+1+3+1+2}{66} = \frac{10}{77} = 0,13$$

Rata-rata absensi ada pada cluster 2:

$$= \frac{3+3+1+2+2+4}{12} = \frac{16}{20} = 0,80$$

Rata-rata absensi yang ada pada cluster 3:

$$= \frac{1+2+1+1+5+1+1+6+9+7+17+\dots+10+1+6+2+5+1+1+2}{66} = \frac{169}{70} = 2,41$$

Jadi, hasil centroid baru *cluster 1*, *cluster 2* dan *cluster 3* adalah Centroid Baru

Setelah didapatkan titik pusat baru dari tiap cluster, hitung kembali data dengan pusat cluster yang baru, ulangi sampai didapatkan pola terakhir yang sudah tidak berpindah, dalam penelitian ini, data dihitung ulang sampai iterasi ke 6, dimana setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain. Hasil dan pola terakhir dapat dilihat pada Tabel 8.

Tabel 8. Centroid Baru

cluster 1	2.21	4.69	104.54	0.13
cluster 2	7.25	8.80	102.95	0.80
cluster 3	2.14	3.83	96.62	2.41

Tabel 9. Hasil dan Pola terakhir jarak antra centroid dan pusat cluster

No	Nama	Eskul	Nilai Eskul	Jumlah IP	Absen Harian	C 1	C 2	C 3	C 1	C 2	C 3
1	AMA	4	7	97.98	0	7.39	5.68	5.11			*
2	ASN	1	4	111.5	0	6.95	11.84	15.31	*		
3	AMA	1	3	98.28	1	6.67	8.63	2.1			*
4	ASN	1	3	100.02	0	4.96	8.23	4.05			*
5	ASI	4	7	103.09	1	3.66	3.34	8.21		*	
6	ASA	1	3	101.17	2	4.33	7.63	4.92	*		
7	AHT	1	3	99.73	0	5.23	8.31	3.79			*
8	CCI	1	3	96.26	0	8.57	9.92	1.59			*
9	DFN	1	3	100.91	0	4.15	8.04	4.89	*		
10	DSI	1	3	100.38	1	4.71	7.9	4.15			*
11	HYI	1	3	98.77	1	6.2	8.42	2.57			*
12	IMH	1	3	104.65	0	1.81	8.29	8.51	*		
13	ISI	1	3	99.21	0	5.73	8.48	3.33			*
14	MMR	1	3	98.51	5	8.03	8.78	4.14			*
15	PNH	1	3	97.88	1	7.05	8.82	1.73			*
....											
158	ISA	1	3	105.64	0	2.07	8.64	9.48	*		
159	ISU	1	3	104.87	0	1.82	8.36	8.73	*		
160	IKA	1	6	101.66	0	3.46	6.54	6.22	*		
161	KOK	1	3	99.23	0	5.71	8.47	3.34			*
162	MRI	1	3	101.56	0	3.57	7.96	5.5	*		
163	MFA	1	3	100.35	0	4.66	8.14	4.36			*

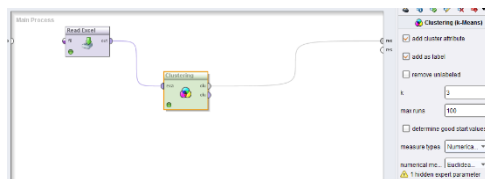
164	NAI	1	6	103.76	0	1.97	6.7	8.1	*		
165	PNH	1	6	100.76	0	4.26	6.68	5.46	*		
166	RAI	1	6	103.02	0	2.39	6.57	7.42	*		
167	RSA	1	3	97.98	0	6.9	8.99	2.32			*
168	SEN	4	6	108.21	0	4.37	7.28	12.6 1	*		
169	SJY	1	3	96.09	0	8.74	10.02	1.61			*
170	SSH	1	6	102.32	0	2.91	6.52	6.8	*		
171	THA	1	6	103.77	0	1.96	6.7	8.11	*		
172	WW F	4	6	109.29	0	5.29	8.19	13.6 3	*		
173	YHI	1	3	96.04	0	8.79	10.05	1.61			*

Data yang dikelompokkan pada cluster 1 berjumlah 68 orang, di cluster 2 berjumlah 38 orang dan di cluster 3 berjumlah 67 orang. Dengan criteria data bahwa pada cluster 2 berisi peserta didik yang memiliki prestasi paling tinggi, cluster 1 memiliki criteria peserta didik yang berprestasi menengah, dan cluster 3 memiliki criteria peserta didik dengan prestasi cukup. Prestasi tersebut dilihat dari aktifnya peserta didik pada kegiatan ekstrakurikuler, nilai-nilai pada ilmu pengetahuan, dan absensi peserta didik harian.

G. Evaluasi dan Validasi Hasil

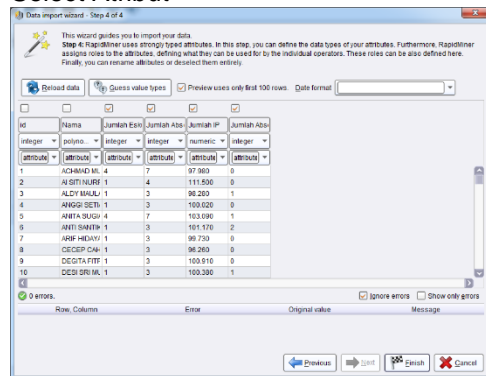
pengujian jarak kedekatan dengan *euclidean distance*, *Mahattan Distance*, *Chebifhap Distance*

1. Cluster model Dengan Kedekatan Jarak *Euclidean Distance*



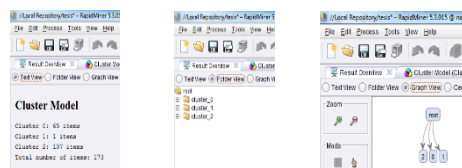
Gambar 2. penentuan jarak terdekat dengan *Euclidean distance*

Select Atribut



Gambar 3. Seleksi atribut

Cluster model yang diperoleh dari hasil pengujian terhadap data menggunakan k-mean clustering terlihat pada gambar 4.

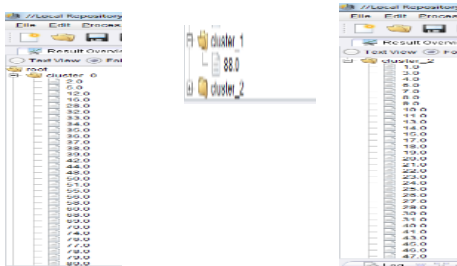


Gambar 4. Cluster Model

Dapat dilihat cluster model yang dihasilkan terdiri dari cluster 0 65 item, cluster 1 1 item, cluster 2, 107 item. Dari total jumlah 173 item. Dalam bentuk folder root tampak bahwa root memiliki 3 cluster yaitu folder cluster 0, cluster 1, cluster 2. Cluster model juga dapat digambarkan dalam tree. Induk adalah root yang memiliki child berupa 2,0,1 .

Untuk memudahkan melihat anggota yang dimiliki oleh setiap folder cluster,

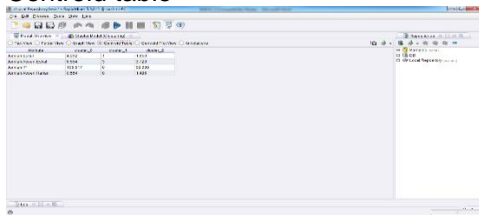
gambar 4.2 menampilkan membership dari masing-masing cluster.



Gambar 5. Tampilan Member masing-masing cluster

Pada gambar 5 ditampilkan membership masing-masing cluster dimana setiap anggota diwakili oleh nomor barisnya. Cluster 0 memiliki anggota nomor baris 88.0, cluster 1 memiliki anggota nomor baris 2.0 dan cluster 2 memiliki anggota nomor baris 1.0

Centroid table

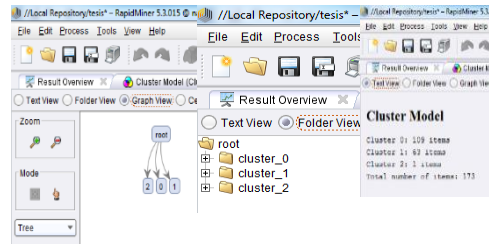


Gambar 6. Tampilan Member masing-masing cluster

Hasil Analisa k-means clustering pada nilai pretasi siswa hasilnya Valid 116 siswa dan tidak valid 57 siswa dari data training 157 siswa didapat 67,05% valid dan 32,95% tidak valid

2. Pengujian dengan jarak kedekatan Mahattan Distanc

Cluster model yang diperoleh dari hasil pengujian terhadap data menggunakan k-mean clustering terlihat pada Gambar 7.

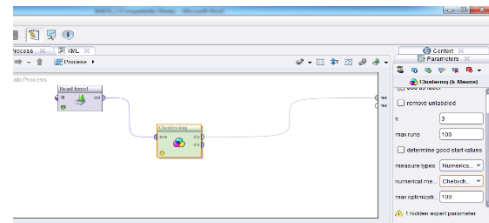


Gambar 7. Cluster Model

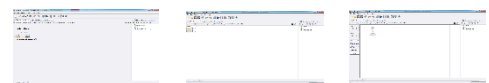
Dapat dilihat cluster model yang dihasilkan terdiri dari cluster 0 109 item, cluster 1 109 item, cluster 2, 67 item. Dari total jumlah 173 item. Dalam bentuk folder root tampak bahwa root memiliki 3 cluster yaitu folder cluster 0, cluster 1, cluster 2. Cluster model juga dapat digambarkan dalam tree. Induk adalah root yang memiliki child berupa 2,0,1.

Untuk memudahkan melihat anggota yang dimiliki oleh setiap folder cluster, gambar 4.8 menampilkan membership dari masing-masing cluster

3. Cluster model Dengan Kedekatan Jarak Chebifhap Distance



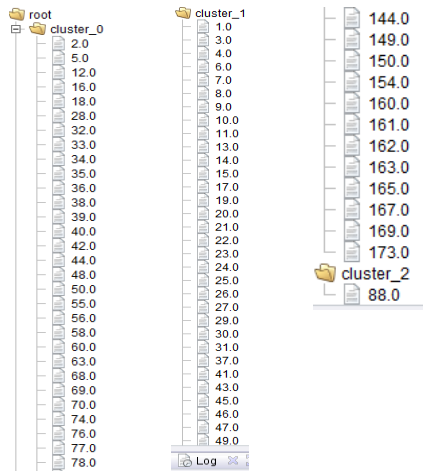
Gambar 8. penentuan jarak terdekat dengan Mahattan Distance



Gambar 9. Cluster Model

Dapat dilihat cluster model yang dihasilkan terdiri dari cluster 0 75 item, cluster 1 97 item, cluster 2, 1 item. Dari total jumlah 173 item. Dalam bentuk folder root tampak bahwa root memiliki 3 cluster yaitu folder cluster 0, cluster 1, cluster 2. Cluster mode l juga dapat digambarkan dalam tree. Induk adalah root yang memiliki child berupa 2,0,1 .

Untuk memudahkan melihat anggota yang dimiliki oleh setiap folder cluster, gambar 4.12 menampilkan membership dari masing-masing cluster Membership



Gambar 10. Tampilan Member masing-masing cluster

Attribute	cluster_0	cluster_1	cluster_2
Jumlah Eskul	3.840	1.948	1
Jumlah Absen Eskul	6.067	3.814	3
Jumlah IP	105.112	98.639	0
Jumlah Absen Harian	0.427	1.680	0

Gambar 11. Tampilan Centroid Table

Tabel 11. Hasil Optimal Jarak Terdekat

Cara Perhitungan Jarak	Euclidean Distance			Mahattan Distance			Cheby Distance chev		
	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2
Jumlah Ekstra	4.292	1	1.85	1.119	5.635	1	3.84	1.1948	1
Jumlah Absen Ekstra	6.554	3	3.72	3.376	7.254	3	6.067	3.814	3
Jumlah IP	105.517	0	98.989	100.222	103.605	0	105.112	98.639	0
Jumlah Absen harian	0.554	0	1.486	0.862	1.603	0	0.427	1.68	0
Jumlah Data Testing	173			173			173		
Valid	116			34			106		
Tidak Valid	57			139			67		
Akurasi Valid	67%			19.65%			61.27%		
Akurasi Tidak Valid	32%			80.35%			38.73%		

Hasil data valid berdasarkan kedekatan jarak *Euclidean Distance*

$$= \frac{116}{173} * \frac{100}{100} = 67.05 \%$$

Hasil data tidak valid berdasarkan kedekatan jarak *Euclidean Distance*

$$= \frac{57}{173} * \frac{100}{100} = 32.95 \%$$

Hasil data valid berdasarkan kedekatan jarak *Manhattan Distance*

$$= \frac{34}{173} * \frac{100}{100} = 19.65 \%$$

Hasil data tidak valid berdasarkan kedekatan jarak *Manhattan Distance*

$$= \frac{139}{173} * \frac{100}{100} = 80.35 \%$$

Hasil data valid berdasarkan kedekatan jarak *Cheby Chev Distance*

$$= \frac{106}{173} * \frac{100}{100} = 61.27 \%$$

Hasil data tidak valid berdasarkan kedekatan jarak *Cheby Chev Distance*

$$= \frac{67}{173} * \frac{100}{100} = 38.73 \%$$

5. PENUTUP

Penelitian ini dibuat untuk mengkaji penerapan model data mining dalam mengelompokkan prestasi siswa, pada penelitian ini diimplementasikan algoritma data mining *k-means*. Dari hasil analisis diketahui bahwa:

- a. Dengan Analisis Algoritma *k-means* dengan penentuan jarak terdekat menggunakan *euclidian distance*
- b. lebih optimal dibandingkan dengan menggunakan *mahattan distance* dan *chbychep distance* dalam mengelompokkan prestasi siswa
- c. Penentuan centroid (titik pusat) pada tahap awal algoritma *k-means* sangat berpengaruh pada hasil cluster seperti pada hasil pengujian yang dilakukan dengan menggunakan 100 record dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda juga.
- d. Diperoleh suatu model *clustering* yang dapat digunakan di SMP Negeri 1 Sukahening dalam pengambilan keputusan pengelompokan prestasi siswa tinggi, menengah, cukup.

DAFTAR PUSTAKA

- [1] Agusta, Y. (2007). *K-means - Penerapan*, Garcia, E.P.I. dan P.M. Mora. 2011. *Model Prediction of Academic Performance for First Year Students*. *IEEE Computer Society*. 10 Oktober 2014
- [2] Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika* Vol. 3 (Februari 2007): 47-60.
- [3] Berkhin, Pavel. *Survey on clustering data mining techniques*, http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf, 20 Oktober 2014.
- [4] Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer
- [5] Kabakchieva, D. 2012. *Student Performance Prediction by Using data Mining Classification Algorithms*. *IJCSMR*. Vol 1 Issue 4: 86-690, 15 Oktober 2014.
- [6] Kusriani, & Luthfi, T. E. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi
- [7] Kusumadewi, Sri., dan Purnomo, Hari. 2010. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan Jakarta, Graha Ilmu*
- [8] Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- [9] Pudjo Widodo, P.Trias Handayanto, R.Herlawatni. (2013). *Penerapan Data Mining Dengan Matlab*. Bandung Rekayasa Sains
- [10] Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [11] Santoso, S. (2010). *Statistik Multivariat*. Jakarta: Elex Media Komputindo.
- [12] Shovon, Md.H.I. dan M. Haque. (2012). *Prediction of Student Academic Performance by an Application of K-Means clustering Algorithm*. *IJARCSSE*. Volume 2 Issue 7: 353-355, 20 Juli 2014
- [13] Slameto, (2010). *Belajar & Faktor-Faktor yang Mempengaruhi*. Jakarta: Rineka
- [14] William, Graham, *Data Mining Cluster*, 20 Oktober 2014 http://datamining.anu.edu.au/student/math3346_2005/050809-maths3346-clusters-2x2.pdf
- [15] Wahono, R. S. (2012, 06 18). *Kiat Menyusun Alur Latar Belakang Masalah Penelitian*. Dipetik 02 13, 2013 dari <http://romisatriawahono.net/>: <http://romisatriawahono.net/2012/06/18/kiat-menyusun-alur-latar-belakang-masalah-penelitian/>
- [16] Wahono, R. S. (2012, 08 07). *Kiat Menyusun Kerangka Pemikiran Penelitian*. Dipetik 05 27, 2013, dari <http://romisatriawahono.net/>: <http://romisatriawahono.net/2012/08/07/kiat-menyusun-kerangka-pemikiran-penelitian/>