

# KLASIFIKASI SISWA UNTUK MENINGKATKAN NILAI RATA-RATA KELAS MENGGUNAKAN METODE DATA MINING

Langgeng Listiyoko<sup>\*1</sup>, Rosalia Wati<sup>2</sup>, Achmad Fahrudin<sup>3</sup>  
<sup>1,2,3</sup>STMIK Muhammadiyah Banten

Kontak Person :

Langgeng Listiyoko, Rosalia Wati, Achmad Fahrudin  
e-mail: langgeng.listiyoko@stmikmbanten.ac.id<sup>\*1</sup>, rosaliawati@stmikmbanten.ac.id<sup>2</sup>,  
achmad.fahrudin@stmikmbanten.ac.id<sup>3</sup>

## Abstrak

*Heterogenitas pada sekelompok siswa dapat mempengaruhi tingkat pencapaian prestasi individu dan penilaian sekolah secara umum. Salah satu usaha yang dapat ditempuh untuk meningkatkan nilai rata-rata kelas/ sekolah adalah dengan mengelompokkan siswa dengan kesamaan tertentu. Kelompok yang terbentuk diharapkan tidak menimbulkan kesenjangan yang cukup jauh sehingga anggota kelompok memiliki potensi yang relatif sama. Dengan bantuan data mining clustering maka akan didapat paket kelompok siswa dengan kemampuan yang setara sehingga dapat diberikan penanganan yang terarah. Hasil penelitian secara detil akan memberikan informasi setiap individu mengenai clusternya berdasarkan hasil ujian terhadap sekelompok mata pelajaran utama dan kompetensi yang dilakukan pada satu waktu.*

**Kata kunci:** data mining, clustering, klasifikasi, cluster.

## 1. Pendahuluan

Persaingan bisnis dalam dunia pendidikan salah satunya dapat ditunjukkan pada pencapaian nilai rata-rata mata pelajaran utama (sebagai item Ujian Nasional). Beberapa usaha khususnya berkaitan dengan teknis proses belajar mengajar dilakukan setiap institusi pelaksana pendidikan demi meningkatkan prestasinya. Sebab bukan tidak mungkin hasil pencapaian yang sempurna akan meningkatkan nilai jual institusi yang bersangkutan.

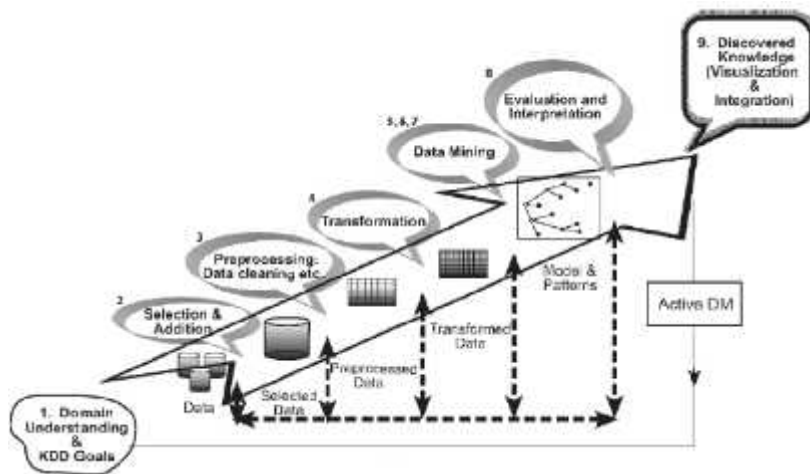
Sistem informasi sebagai tools modern dapat diaplikasikan demi mendukung usaha peningkatan prestasi institusi pendidikan, dalam hal ini perolehan nilai rata-rata Ujian Nasional. Selain teknis operasional langsung dengan pemberian waktu belajar tambahan, materi latihan dan segala bentuk *try out*, dipandang perlu juga memperhatikan kondisi dan kemampuan setiap siswa. Sejumlah siswa yang terjebak dalam lingkungan yang kurang sehat dapat mempengaruhi daya belajarnya. Hal ini dapat dipahami jika siswa merasakan secara psikis kesenjangan yang cukup berarti di lingkungannya. Klasifikasi terhadap sekelompok siswa dianggap dapat memberikan kontribusi peningkatan pencapaian prestasi sekolah, yaitu dengan mengelompokkan siswa dengan kemiripan potensi dan kemampuannya sehingga muncul rasa percaya diri di setiap individu [1].

Data *mining* digunakan untuk melakukan pekerjaan klasifikasi dimaksud [12]. Dengan metode *clustering* maka akan didapat *cluster* yang mengelompokkan siswa dengan memperhatikan data histori perolehan nilai di tahun-tahun sebelumnya, test khusus berkaitan dengan minat dan potensi, *attitude* sehari-hari dan kegiatan penunjang yang diikuti (ekstra kurikuler). Perlu dikesampingkan variabel status sosial sebab hal tersebut dinilai subyektif dan tidak berpengaruh pada peningkatan prestasi siswa secara individu.

Dalam penelitian ini siswa SD tahun ke-5 pada akhir semester kedua akan dilakukan peninjauan tren prestasi sejak tahun pertama. Hasil peninjauan ini selanjutnya dilakukan pengelompokan untuk menentukan sub kelas di tahun ke-6. Hal ini dinilai mampu memberikan proyeksi tentang langkah-langkah strategis pembelajaran demi mencapai hasil maksimal yang dapat dirasakan seluruh siswa dan sekolah pada umumnya. Pada sistem konvensional, penentuan sub kelas dapat dengan mudah dilakukan dengan pemeringkatan pada hasil ujian di tahun ke-5. Hal ini dipandang kurang representatif sebab adakalanya orang tua dan wali kelas hanya akan fokus di masa tersebut, jadi tidak merefleksikan prestasi siswa sesungguhnya. Tentu hal ini kurang baik sebab akan menimbulkan kesan siswa "instant".

## 2. Metode Penelitian

Langkah-langkah penelitian dalam menentukan sub kelas tahun ke-6 siswa SD sebagaimana disadur dari Maimon [7] diilustrasikan dalam gambar berikut :



Gambar 1 Tahapan Penelitian Data Mining

### 2.1 Domain Understanding & KDD (Knowledge Discovery in Database) Goal

Langkah pertama dalam penelitian data mining adalah menentukan secara pasti tujuan utama penelitian. Selain itu secara bersama-sama juga harus membatasi ruang lingkup penelitian, agar dapat dilaksanakan dengan terarah (fokus). Dalam hal penentuan sub kelas siswa kelas 6 (dilakukan di akhir tahun ke -5) dengan jelas disepakati bahwa tujuan utama adalah mendapatkan paket kelompok yang memiliki kemiripan agar dapat diberikan langkah strategis untuk mencapai nilai rata-rata kelulusan yang optimal. Ruang lingkup penelitian dibatasi hanya pada siswa kelas 5 SD yang memiliki 5 sub kelas.

Saat ini pengelompokan sub kelas dilakukan pada tahun pertama awal masuk, tanpa ada kriteria tertentu. Selama 6 tahun kemudian belajar di sekolah praktis tidak ada penyegaran mengenai lingkungan belajar di dalam kelas, sehingga pola pikir yang terbentuk dinilai relatif monoton dan lambat berkembang. Satu-satunya penyegaran yang dialami adalah pergantian wali kelas.

### 2.2 Selection & Addition

Pada dasarnya ada banyak sekali data atau atribut yang melekat di setiap entiti (siswa) yang dapat memberikan informasi. Namun kurangnya pengetahuan tentang potensi informasi menjadikan rekaman data siswa hanya tersimpan sebagai arsip statis. Kondisi ini sangat merugikan sebab bertambahnya jumlah file yang berakibat pada meningkatnya kebutuhan ruang penyimpanan (baik *hard copy* maupun *soft copy*) tidak diimbangi dengan *outcome* yang memadai. Dalam penelitian ini akan memanfaatkan lebih banyak data-data siswa yang dihimpun sejak pertama para siswa bergabung di sekolah.

Atribut selain data pribadi siswa yang *dideliver* saat pendaftaran bersifat dinamis. Diperlukan pemilihan atribut yang sesuai untuk penelitian dan mungkin juga penambahan atribut (*variable*) untuk mendukung pengolahan *data mining* menggunakan metodologi *clustering* [2]. Kompleksitas atribut pada objek penelitian menuntut pemrosesan data menggunakan *data mining*. Hal ini berkaitan dengan keterbatasan kemampuan manusia dalam merumuskan sistem pendukung keputusan.

Oleh karena metodologi yang dipilih adalah *clustering*, maka atribut yang diperlukan adalah atribut dengan *value* numerik. Jika nantinya harus melibatkan *value non* numerik maka akan dilakukan transformasi terlebih dahulu. Atribut yang dimaksud adalah Nilai Total, Nilai Bahasa Indonesia, Nilai Matematika, Nilai IPA, Nilai IPS pada masing-masing semester. Selain itu diperlukan juga atribut mengenai keaktifan di luar kelas. Apakah pernah mengikuti perlombaan, apakah mengikuti ekstrakurikuler, apakah memiliki kepribadian yang baik dan seterusnya yang sifatnya individual.

### 2.3 Pre Processing

*Preprocessing* adalah kegiatan persiapan data sebelum dapat benar-benar digunakan dalam penelitian [2]. Umumnya preprocessing berkaitan dengan penyeragaman format. Seperti diketahui ada

beberapa format yang dapat dipakai untuk merepresentasikan nilai numerik, seperti penentuan jumlah digit berarti, penulisan lambang desimal, atau pembulatan. Data yang akan diproses harus dalam format yang sama, termasuk format hasil konversi dari data non numerik. Kesalahan-kesalahan kecil mungkin timbul dalam proses input data manual ke dalam *soft copy*.

## 2.4 Transformation

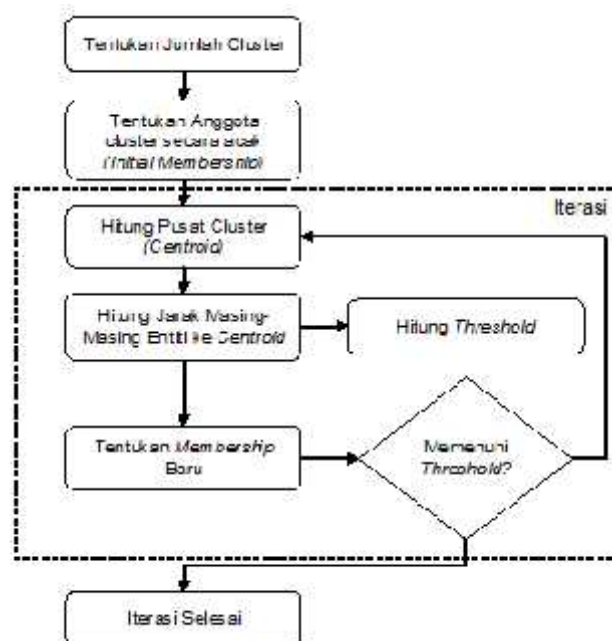
Pada tahap ini yang paling penting adalah konversi data non numerik menjadi numerik. Teknik yang dapat dilakukan dapat berupa binerisasi, klasifikasi, atau pemeringkatan. Binerisasi bisa dilakukan pada data dengan 2 variasi *value*, seperti jenis kelamin. Klasifikasi bisa mengatasi atribut dengan *value* yang lebih kompleks, misalnya tentang kedisiplinan atau *attitude*. Sedangkan pemeringkatan adalah teknik umum yang dapat dilakukan pada data yang berkenaan dengan prestasi atau pencapaian target. Semua usaha transformasi yang dilakukan akan menghasilkan data numerik.

## 2.5 Data Mining

Ini adalah inti dari penelitian yang dapat menentukan keberhasilan pengambilan kesimpulan dalam konteks sistem pendukung keputusan. Pemilihan metodologi juga penting dengan memperhatikan kompleksitas data, tujuan yang ingin diraih, koleksi data yang dimiliki, maupun kesiapan kompetensi peneliti. *Clustering* dipilih dalam menentukan kelompok kelas oleh karena prosesnya melibatkan data numerik dengan jumlah atribut maupun entiti yang cukup banyak. *Clustering* bekerja berdasarkan kemiripan hasil perhitungan masing-masing entiti yang dengan sendirinya akan membentuk kelompok-kelompok [1][2][12][14]. Entiti yang tergabung dalam satu kelompok dapat dianalogikan seperti sekumpulan material metal yang tertarik oleh magnet yang berada dalam jangkauan medan magnetnya.

Pada teknik *clustering* jumlah *cluster* sudah ditentukan di muka [6], dalam hal ini adalah jumlah sub kelas, yakni 5 cluster. Dengan begitu dapat dipandang tugas utama *clustering* adalah mengelompokkan sejumlah siswa yang ada ke dalam 6 kelompok. Potensi masalah yang akan timbul dari proses ini adalah jika hasil anggota dalam setiap kelompok tidak sama. Maka akan dilakukan teknik lanjutan untuk memenuhi kuota setiap kelompok.

Adapun algoritma data mining yang akan diterapkan adalah sebagai berikut :



Gambar 2 Algoritma Clustering

## 2.6 Evaluating & Interpretation

Tahapan selanjutnya adalah evaluasi dan interpretasi. Evaluasi berkenaan dengan pengukuran hasil penelitian dibandingkan dengan kondisi sebelum implementasi. Penelitian ini dilakukan sebatas

perancangan proses pengolahan data menjadi informasi, oleh karenanya fokus evaluasi tidak akan merujuk pada hasil implementasi melainkan pengujian konsistensi proses *clustering*. Adapun interpretasi di sini adalah bagaimana menerjemahkan hasil pengolahan data ke dalam laporan yang dapat dipahami para *stakeholder*. *Delivery* informasi ini sangat penting dalam komunikasi peneliti untuk menghindari kesalahpahaman.

## 2.7 Visualization & Integration

Visualisasi pada tahap akhir adalah bentuk laporan komprehensif yang diajukan kepada pengambil keputusan sebagai bentuk laporan dan dasar pertimbangan atau alternatif solusi. Bentuk visualisasi berupa grafik dapat diintegrasikan ke dalam sistem *dashboard* yang lebih baik sehingga pemahaman tentang penelitian dapat mudah dipahami dengan interpretasi yang tepat.

## 3. Hasil Penelitian dan Pembahasan

Hasil penelitian dalam menentukan sub kelas pada sekelompok siswa SD tahun ke-5 dijelaskan per sub bab sesuai tahapan penelitian yang dianut.

### 3.1. Seleksi Data

Dari sekumpulan data siswa yang diteliti, dipilih data siswa kelas 5 tahun berjalan sebagai populasi, dengan jumlah sample 33 siswa (1 kelas). Kemudian untuk keperluan pengolahan *clustering*, dipilih atribut atribut sebagai berikut :

**Tabel 1** Seleksi Atribut Sebagai Dimensi Data

No	Atribut ( <i>Variable</i> )	Sumber Data
1	Nilai Total	Raport Siswa
2	Nilai Matematika	Raport Siswa
3	Nilai Bahasa Indonesia	Raport Siswa
4	Nilai IPA	Raport Siswa
5	Nilai IPS	Raport Siswa
6	Jenis Kelamin	Akademik
7	Juara Eksternal	Akademik
8	Member Ekskul	Kesiswaan

Dari kebutuhan variabel di atas tampak bahwa sudah melibatkan 3 sumber data. Secara alami, tidak mudah bagi pengambil keputusan menilai kemiripan sifat entiti secara konvensional. Oleh karenanya diterapkanlah data mining yang memiliki beberapa teknik untuk menyelesaikan permasalahan ini. Namun begitu sistem hanya akan menampilkan alternatif solusi, adapun pengambil keputusan tetap dikendalikan manusia (*manager*).

Dimensi atau atribut data yang dibutuhkan bukanlah data mentah seperti koleksi data dari kelas 1 hingga kelas 5 (**Gambar 3**), melainkan hanya berupa nilai rata-rata /*average* (**Gambar 4**). Atribut tambahan ini adalah atribut yang hanya dihasilkan pada tahap *pre processing*.

No	Nama	Nilai Total				
		1	2	3	4	5
1	Abdulloh Azzam Izzuddin	78	82	80	77	83
2	Adelia Putri Ramadhani	68	68	66	70	68
3	Adrian Dwi Purwanto	68	68	72	78	80
4	Affan Dwi Deswanto	80	88	78	72	82
5	Anggite Helfi Al Karim	70	70	85	76	78
6	Ania Dina Rahma	66	80	78	88	74
7	Azzah Jazilatur Rahma	72	68	72	78	90
8	Bagus Okka Amrullah	92	90	86	88	74
9	Daffa Afabian	88	88	84	90	91
10	Daffa Serya Nurmaruf	90	91	90	92	92

**Gambar 3** Koleksi Data Nilai Total Semester 2 Kelas 1-5

No	Nama	Avg
1	Abdulloh Azzam Izzuddin	80
2	Adelia Putri Ramadhani	68
3	Adrian Dwi Purwanto	75.2
4	Affan Dwi Deswanto	80
5	Anggito Helfi Al Karim	75.8
6	Aulia Dina Rahma	77.2
7	Azzah Jazilatur Rahma	76
8	Bagus Okka Amrullah	86
9	Daffa Afabian	88.2
10	Daffa Setya Nurma'ruf	91

**Gambar 4** Penyederhanaan Data Nilai Menjadi *average*

**Gambar 4** adalah hasil rata-rata dari data seperti tampak pada **Gambar 3**. Namun perlu diperhatikan bahwa keluaran pada nilai rata-rata tidak menunjukkan adanya keseragaman struktur data numerik (*constraint*). Hal ini terlihat jelas pada penulisan digit berarti, yang tidak diatur sebelumnya. Oleh karenanya diperlukan tahap *pre processing*.

### 3.2. Pre Processing

Hasil dari *pre processing* khususnya data nilai adalah sebagai berikut :

Data ke i	Av. NT	Av. MTK	Av. BI	Av. IPA	Av. IPS
1	80.00	8.18	9.00	9.16	6.26
2	68.00	8.26	9.16	9.00	9.16
3	75.20	7.32	6.72	9.00	9.00
4	80.00	8.90	6.26	7.74	6.72
5	75.80	8.14	8.62	6.72	6.72
6	77.20	9.00	9.26	6.26	9.00
7	76.00	9.16	9.02	9.00	9.16
8	86.00	7.71	8.96	9.16	9.16
9	88.20	6.72	8.96	9.00	9.00
10	91.00	6.26	9.16	9.00	6.72

**Gambar 5** Hasil *Pre Processing*

Data *average* nilai yang telah diolah dengan *constraint* yang ditentukan kini dapat digunakan sebagai inputan pemroses *clustering* selanjutnya.

### 3.3. Transformasi

Transformasi menghasilkan data sebagai berikut :

Data ke i	Ext	JK	Ekskul
Abdulloh Azzam Izzuddin	3	1	1
Adelia Putri Ramadhani	0	0	0
Adrian Dwi Purwanto	2	1	1
Affan Dwi Deswanto	2	1	1
Anggito Helfi Al Karim	1	1	1
Aulia Dina Rahma	0	0	0
Azzah Jazilatur Rahma	0	0	0
Bagus Okka Amrullah	1	1	1
Daffa Afabian	0	1	0
Daffa Setya Nurma'ruf	0	1	1

**Gambar 6** Hasil Transformasi Pada Data Prestasi Eksternal, Jenis Kelamin, dan Keaktifan Ekstra Kurikuler

Transformasi yang dimaksud adalah mengubah struktur dan atau *type* data *text* menjadi numerik. Pada dimensi Ext (prestasi eksternal) diisi nilai sesuai hasil pencapaian prestasinya. Siswa yang tidak

memiliki prestasi dianggap memiliki nilai nol. Sementara Jenis kelamin dan keikutsertaan ekstra kulikuler yang merupakan data biner lebih mudah diubah. Untuk jenis kelamin nilai 1 diberikan untuk laki-laki dan 0 untuk perempuan, sedangkan keikutsertaan ekstra kurikuler diberikan nilai 1 untuk yang mengikuti kegiatan.

### 3.4. Data Mining

Metodologi *clustering* yang dikerjakan menganut algoritma sebagai berikut [4][6] :

#### a. Menentukan jumlah *cluster*

Jumlah cluster yang diinginkan adalah sejumlah sub kelas yang ada, yakni 5. Prototype model pada penelitian ini akan membagi 33 sample siswa ke dalam 5 sub kelas.

#### b. Inisialisasi

Seperti yang berlaku saat ini dalam pengelompokan kelas, secara alami siswa dikelompokkan berdasar nomor urut pendaftaran. Dalam penelitian ini akan dilakukan inisialisasi anggota kelas juga berdasarkan nomor urut.

**Tabel 2** Penentuan Anggota *Cluster* Awal (Inisialisasi)

Sub Kelas	Nomor Urut
<b>A</b>	1-7
<b>B</b>	8-14
<b>C</b>	15-21
<b>D</b>	22-28
<b>E</b>	29-33

#### c. Hitung *centroid*

Pada masing-masing cluster awal yang telah berisi sekumpulan entiti (siswa) dihitung pusat cluster atau *centroid*. *Centroid* dihitung untuk setiap dimensi atau variable yang entiti, per *cluster* sehingga pada akhirnya akan membentuk matriks. Adapun perhitungan *centroid* dilakukan menggunakan rumus (1) berikut [3]:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

Untuk cluster 1 (C1) didapat perhitungan sebagai berikut :

Hitung centroid									
C1									
Data ke-i	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	
1	80.00	8.18	9.00	9.16	6.26	3.00	1.00	1.00	
2	68.00	8.26	9.16	9.00	9.16	0.00	0.00	0.00	
3	73.70	7.32	8.72	9.00	9.00	2.00	1.00	1.00	
4	80.00	8.90	6.26	7.74	6.72	2.00	1.00	1.00	
5	75.80	8.14	9.52	6.72	6.26	1.00	1.00	1.00	
6	77.30	9.00	9.26	6.26	9.00	0.00	0.00	0.00	
7	76.00	9.16	9.02	9.00	9.16	0.00	0.00	0.00	
	75.71286	8.42	8.79	8.13	7.94	1.14	0.57	0.57	Centroid

**Gambar 7** Perhitungan *Centroid* Pada C1 Inisialisasi

#### d. Menghitung jarak pusat *cluster*

Pada setiap value entiti dihitung jarak ke pusat cluster masing-masing dimensi. Adapun formula perhitungannya digunakan formula *Euclidean Distance*, seperti dijelaskan dalam rumus (2).

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

Matriks yang akan terbentuk dari hasil perhitungan jarak *Euclidean* diilustrasikan dalam gambar berikut :

	No Cluster					Nilai Min
	1	2	3	4	5	
1	5.1085	7.0185	4.5571	1.6575	3.3067	1.6575
2	8.0584	18.5052	15.1784	12.0431	12.7142	8.0584
3	3.7917	13.1192	9.4637	7.1965	7.1901	3.7917
4	5.2564	7.0084	4.1059	2.4565	3.2114	2.4565
5	2.5062	10.8021	8.2230	4.1710	6.0833	2.5062
6	2.7539	9.6137	7.4654	4.7087	4.0630	2.7539
7	2.1246	10.6592	7.8655	5.1224	5.0918	2.1246
8	10.3189	1.7676	4.5391	7.4405	5.8110	1.7676

**Gambar 8** Jarak *Euclidean* Masing-Masing Entiti

Nilai minimum menunjukkan keanggotaan entiti ke dalam satu cluster. Pada gambar di atas untuk entiti nomor 1 maka akan tergabung dalam *cluster* 4, mengingat pada *cluster* tersebut jarak *Euclidean* entiti yang bersangkutan memuat nilai minimum. Begitu seterusnya untuk entiti nomor 2 tergabung di *cluster* 1, hingga seluruh entiti memiliki *clusternya* masing-masing.

e. Alokasi item set ke dalam *centroid*

Setelah seluruh entiti diketahui nilai minimum jarak *Euclidean*-nya maka selanjutnya adalah mengalokasikan entiti ke dalam *cluster* baru. Perubahan anggota dalam *cluster* yang baru mengindikasikan seberapa dekat prediksi inisialisasi dengan perhitungan *data mining*. Jika anggota cluster tidak berubah sama sekali, maka data mining hanya berlangsung 1 putaran saja. Pada entiti dengan jumlah besar, sungguh berat memeriksa konsistensi keanggotaan cluster. Untuk itu digunakanlah pendekatan matematis untuk memeriksa apakah proses clustering sudah optimal atau belum.

Nilai minimum setiap entiti membantu penelitian menentukan *threshold* (nilai kepercayaan). Nilai rata-rata dari total jarak minimum *Euclidean* setiap dimensi kemudian disebut Fungsi Obyektif. *Threshold* adalah selisih antara Fungsi Obyektif (FO) baru dan FO lama [12] seperti dirumusan (3). Idealnya *threshold* adalah nol, namun pada penelitian dengan kondisi tertentu bisa jadi lebih besar, namun tetap kurang dari 1. Dalam penelitian ini ditetapkan *threshold* adalah nol.

$$T = FO_{\text{baru}} - FO_{\text{lama}} \quad (3)$$

f. Iterasi

Iterasi akan dilakukan ketika nilai *threshold* belum terpenuhi. Proses iterasi akan dimulai dengan menghitung kembali *centroid* untuk kemudian menghitung jarak *Euclidean*. Iterasi akan berhenti ketika nilai *threshold* tercapai, umumnya ditandai pula dengan tidak berubahnya keanggotaan *cluster* dari iterasi sebelumnya.

### 3.5. Evaluasi dan Interpretasi

Hasil pengolahan data dalam sampling 33 siswa kelas 5 adalah sebagai berikut :

ID	Nama	Cluster
1	Abdulloh Azram Irzuddin	4
2	Adelia Putri Ramadhani	1
3	Adrian Dwi Purwanto	1
4	Affan Dwi Deswanto	4
5	Anggito Heli Al Karim	1
6	Aulia Dina Kalana	1
7	Azzah Jaenkatun Rahma	1
8	Bagus Olka Anzullah	2
9	Daffa Alabian	2
10	Daffa Saqya Nurma'rif	?
32	Vina Regita Adhita Pramesri	1
34	Wahyu Hidayat	?

**Gambar 9** Hasil Pemetaan *Cluster* Berdasar Perhitungan Jarak *Euclidean*

### 3.6. Visualisasi

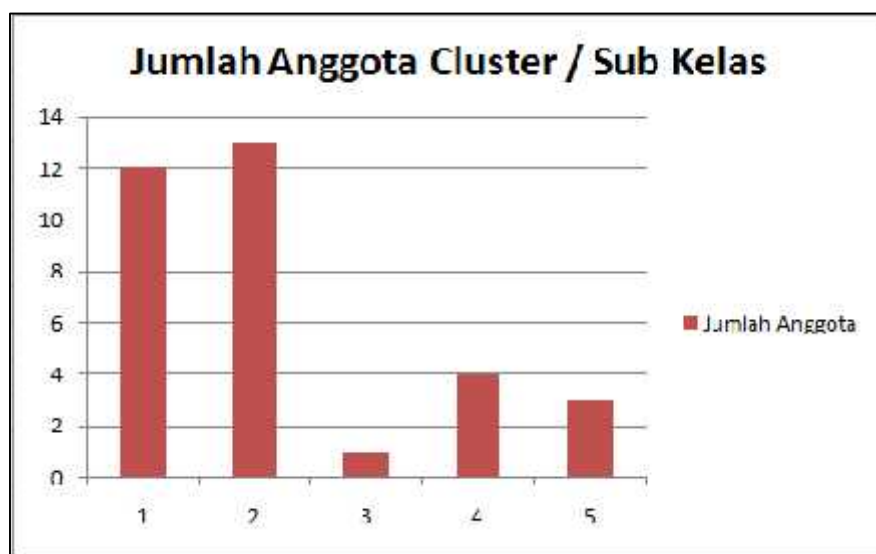
Dari hasil perhitungan yang dilakukan dengan teknik clustering, dapat disajikan data seperti dalam tabel di bawah ini. Sebanyak 12 siswa memiliki kemiripan dan bergabung di cluster 1, 13 siswa tergabung di cluster 2, 1 siswa di cluster 3, 4 siswa di cluster 4 dan 3 siswa lainnya tergabung di cluster 5. Di cluster 3 yang hanya beranggotakan 1 siswa menandakan adanya kesulitan yang serius sehingga yang bersangkutan sama sekali berbeda dengan teman lainnya. Tentu saja konseling dapat dilakukan dengan lebih terarah setelah melihat fakta ini.

No Cluster	Jumlah Anggota	%
1	12	36%
2	13	39%
3	1	3%
4	4	12%
5	3	9%

**Gambar 10** Visualisasi Laporan Dalam Prosesentase Jumlah Anggota Tiap Cluster

**Grafik 1** memperjelas pemahaman tentang kondisi sample siswa yang diteliti, bahwa lebih dari 50% siswa berada dalam cluster yang sama dengan rincian 36% masuk cluster 1 dan 39% masuk cluster 2. Sisanya sejumlah 8 siswa (24%) tersebar dalam 3 cluster, yang mengindikasikan mereka tidak dapat mengikuti prestasi kelas secara umum. Dengan perspektif yang lain bisa dilihat bahwa kelas tersebut termasuk kelas unggulan meskipun masih kurang ideal.

Konsistensi keanggotaan cluster mengindikasikan bahwa komposisi siswa sudah optimal dan bisa dikatakan memiliki kemiripan atau tren prestasi yang seimbang [14]. Hasil penelitian semacam ini sangat dibutuhkan bagi manajemen untuk menentukan langkah-langkah strategis terkait dengan penanganan siswa yang berada di luar cluster.



**Grafik 1** Visualisasi Laporan Dalam Bentuk Grafik

### 4. Kesimpulan

Dari 33 sample siswa yang diteliti dapat dilaporkan 24% siswa berpotensi tidak dapat mengikuti kegiatan belajar mengajar yang baik sehingga dapat mempengaruhi rata-rata pencapaian nilai Ujian Nasional. Deteksi awal yang lebih akurat pada siswa yang relatif lemah seperti laporan penelitian ini dapat membantu menentukan langkah strategis untuk meningkatkan pencapaian Sekolah secara umum.

Saran untuk penelitian selanjutnya adalah dapat diimplementasikan untuk penjaringan masuk perguruan tinggi unggulan atau sistem pemberian beasiswa.



## 5. Daftar Notasi

- M : jumlah item set yang tergabung dalam cluster  $i$   
j : atribut ke- $j$   
x : nilai dari atribut  $j$   
D : jarak  $x_2$   
 $x_1, x_{2j}$  : nilai variable  $j$  untuk data ke- $i$  sedangkan  
 $x_{1j}$  : nilai *centroid* cluster pada variable  $j$   
P : jumlah variable yang ada pada dataset.  
T : nilai *threshold*

## Referensi

- [1] Gorunescu, F., 2011, "Data Mining Concepts, Models and Techniques", Springer, Berlin.
- [2] Han, J., & Kamber, M., 2006, "Data Mining : Concepts and Techniques", Elsevier, San Fransisco.
- [3] Hand, D., Mannila, H., & Smyth, P., 2001, "Principles of Data mining", MIT Press, London.
- [4] Hermawati, F. A., 2013, "Data Mining", Andi Publisher, Yogyakarta.
- [5] Larose, D. T., 2005, "Discovering Knowledge in Data : An Introduction to Data Mining", John Wiley & Sons, Inc., New jersey.
- [6] Kusriani, Luthfi, E.T., 2009, "Algoritma Data Mining", Andi Publisher.
- [7] Maimon, O., & Rokach, L., 2010, "Data Mining And Knowledge Discovery Handbook (Second Edition)", Springer, New York.
- [8] Maurina, D., Fanani, A.Z., 2015, "Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SMA Muhammadiyah Gubug Menggunakan Algoritma C4.5", Jurnal Udinus.
- [9] Meilani, B.D., Slamet, A.F., 2013, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree", Jurnal ITATS.
- [10] Pratiwi, L.D.B., Wibowo, W., Zain, I., 2015, "Klasifikasi Nilai Peminat SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri) ITS dengan Pendekatan CART", Jurnal Sains dan Seni ITS Vol.4 No. 2 D193-198.
- [11] Susanti, 2013, "Klasifikasi Kredit Menggunakan Metode Decision Tree Pada Nasabah PD BPR BKK Gabus", Jurnal Udinus.
- [12] Susanto, S., & Suryadi, D., 2010, "Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data", Andi Publisher, Yogyakarta.
- [13] Suyanto, 2014, "Artificial Intelligence", Informatika, Bandung.
- [14] Vercellis, C., 2009, "Business Intelligence : Data Mining And Optimization For Decision Making", John Wiley & Son Ltd, Padstow.