

## **PENERAPAN DATA MINING DALAM MENENTUKAN PILIHAN JURUSAN BIDANG STUDI SMA MENGGUNAKAN METODE CLUSTERING DENGAN TEKNIK SINGLE LINKAGE**

**Trinanda Syahputra<sup>1</sup>, Jufri Halim<sup>1</sup>, Ery Promo Sintho<sup>1</sup>**

<sup>1</sup>Sistem Informasi, STMIK Triguna Dharma

*email:* trinandasyahputra@gmail.com

**Abstract:** One of the ways to improve the quality of high school students is to form a favorite class in which the class will be a flagship class compared to other classes. Students will be selected based on 3 variables, that is the attendance index, semester average score and also the ethics of the students themselves. Data mining is the mining or discovery of new information by searching for a particular pattern or rule of large amounts of data that is expected to overcome the condition, utilizing data obtained from SMA for grouping students using clustering method using single linkage technique. By using clustering method is expected to make it easier for grouping students - students who go to class XI. So the results of this study will also be used by in determining the students-students who will enter the majors class.

**Keywords:** Data Mining, Clustering, Single Linkage, Selection of Departments.

**Abstrak:** Salah satu cara untuk meningkatkan kualitas siswa sekolah menengah adalah dengan membentuk kelas favorit di mana kelas akan menjadi kelas unggulan dibandingkan dengan kelas lain. Siswa akan dipilih berdasarkan 3 variabel, yaitu indeks kehadiran, nilai rata-rata semester dan juga etika siswa itu sendiri. Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola tertentu atau aturan sejumlah besar data yang diharapkan untuk mengatasi kondisi tersebut, memanfaatkan data yang diperoleh dari SMA untuk mengelompokkan siswa menggunakan metode pengelompokan menggunakan teknik hubungan tunggal. Dengan menggunakan metode clustering diharapkan akan mempermudah pengelompokan siswa - siswa yang masuk ke kelas XI. Jadi hasil penelitian ini juga akan digunakan dalam menentukan siswa - siswa yang akan masuk kelas jurusan.

**Kata kunci:** Data Mining, *clustering*, *single linkage*, pemilihan departemen

### **PENDAHULUAN**

Lembaga pendidikan seperti pendidikan SMA membutuhkan suatu bentuk keputusan dalam memilih jurusan yang sesuai untuk siswa-siswi SMA. Keputusan yang diambil dalam memilih jurusan mungkin hampir benar sesuai

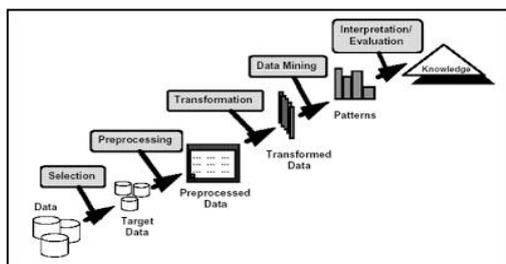
dengan kemampuan, bakat dan minat siswa atau mungkin juga salah. Pembuatan keputusan harus benar-benar mempertimbangkan pilihan yang sesuai untuk penjurusan tersebut, sehingga dibutuhkan sistem yang dapat mengklasifikasikan pola penjurusan kelas XI SMA dengan mempertimbangkan

kemampuan bakat juga minat siswa terhadap suatu jurusan. Jenjang pendidikan di Indonesia terutama tingkat menengah pada pendidikan formal yang ada dan kurikulumnya sudah ada suatu mata pelajaran di mulai dari kenaikan kelas X, ke kelas XI untuk mata pelajaran IPA dan IPS.

*Clustering* (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. Gagasan mengenai pengelompokan data, atau *clustering*, memiliki sifat yang sederhana dan dekat dengan cara berpikir manusia kapanpun di presentasikan jumlah data yang besar, kita biasanya cenderung merangkumkan jumlah data yang besar ini ke dalam sejumlah kecil kelompok-kelompok atau kategori-kategori untuk memfasilitasi analisisnya lebih lanjut. Bahkan, algoritma *clustering* berusaha mensegmentasi seluruh kumpulan data ke dalam subkelompok-subkelompok atau *cluster-cluster* homogen secara relatif. Dimana kesamaan *record* dalam *cluster* di maksimalkan dalam kesamaan dengan *record* diluar *cluster* ini diminimalkan.

## METODOLOGI

Proses persiapan data mining ini dapat mencapai 60% dari keseluruhan proses dalam *Data Mining*. Proses KKD secara garis besar terdiri dari 5 tahapan yaitu *Data selection*, *pre-processing/cleaning*, *transformation*, *data mining* dan *interpretation/evaluation*.



Gambar 1. Proses KKD

Berikut ini merupakan tahapan dari proses KKD:

- a. *Data selection* (Seleksi data)  
Data yang ada dalam *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk analisis yang akan diambil dari *database*.
- b. *Data Integration* (Integrasi data)  
Integrasi data merupakan penggabungan data dari berbagai *database* kedalam satu *database* yang baru.
- c. *Data selection* (Seleksi data)  
Data yang ada dalam *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk analisis yang akan diambil dari *database*.
- d. *Data Transformation* (Transformasi data)  
Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining.
- e. *Proses Mining*  
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan data mining.
- f. *Pattern evaluation* (Evaluasi Pola yang ditemukan)  
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.
- g. *Knowledge Presentation* (Presentasi Pola yang untuk menghasilkan aksi)  
Merupakan visualisi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

## HASIL DAN PEMBAHASAN

Data yang digunakan untuk menghasilkan suatu informasi yang baru serta pengetahuan yang baru. Terutama

Tabel 1. Tabel Siswa dan Variabelnya

No	Nama	Indeks Absen	Nilai UAS	Nilai Kerajinan
1	Adelia Syahitri	4	85	80
2	Andre Agasi	3	87	86
3	Aninditha Cahyadi R.	3	88	80
4	Ari Surya Pratama	3	84	75
5	Dede Syahbana Lubis	4	87	75
6	Febri	3	87	75
7	Heri Surya	4	87	75
8	Heru Wanda	3	88	78
9	Joshua Siahaan	3	89	80
10	Joko Barus	3	80	89
11	Prima Yusati	4	82	70
12	Riski Sembiring	4	86	70
13	Rizaldi	4	81	78
14	Septian Dwiki H.	4	83	70
15	Siska Tripani Br Sitepu	4	85	73
16	Muhamad Irvan Pranata	3	85	70
17	Muhamad Rizki Fadli	4	83	76
18	Muhamad Ridwan	4	86	72
19	Wahyu Ramadhan	4	88	80
20	Wiranata	3	86	87

Keterangan: Indeks Absen (4 = Baik Sekali, 3 = Baik, 2= Kurang Baik, 1 = Tidak Baik)

di sekolah PAB 9 Patumbak memiliki dua jurusan yaitu jurusan IPA dan IPS. Untuk menentukan siswa untuk masuk di jurusan IPA dan IPS adalah dapat kita gunakan metode clustering. Metode clustering adalah metode pengelompokan data yang dimana data nilai siswa akan dikumpulkan berdasarkan 3 Variabel utama. Variabel itu mencakup Indeks Absen, nilai UAS dan Nilai kerajinan. Salah satu teknik penyelesaiannya digunakan dengan teknik *single linkage*. Perhitungan algoritma ataupun teknik *single linkage* ini dilakukan dengan cara mencari nilai kedekatan antar objek satu dengan objek yang lain. Sehingga akan mendapatkan hasil akhir yaitu berupa cluster 0 dan 1. Dalam melakukan perhitungan *Data Mining* dengan Metode *Clustering* menggunakan teknik *single linkage* untuk mengelompokkan data nilai kelas X. Dalam penelitian ini, fungsi-fungsi yang digunakan merupakan fungsi untuk menentukan nilai cluster yang

berguna untuk pembentukan nilai-nilai yang akan diperhitungkan.

Langkah-langkah dalam melakukan pengelompokan siswa

- Melakukan standarisasi data untuk tiap variabel pada masing-masing siswa SMA sebanyak siswa 20 orang. Untuk siswa A atas nama Adelia Syahfitri dengan Variabel 1 maka mencari standarisasi nya adalah sebagai berikut ini :
  - Mencari rata-rata data variabel 1 dengan persamaan (1)

$$X = \frac{\sum_{i=1}^n Xi}{n}$$

$$VI = \frac{V1_1 + V1_2 + V1_3 + V1_4 + V1_5 + V1_6 + V1_7 + V1_8 + V1_9 + V1_{10} + V1_{11} + V1_{12} + V1_{13} + V1_{14} + V1_{15} + V1_{16} + V1_{17} + V1_{18} + V1_{19} + V1_{20}}{20}$$

$$VI = \frac{4 + 3 + 3 + 3 + 4 + 3 + 4 + 3 + 3 + 3 + 4 + 4 + 4 + 4 + 3 + 4 + 4 + 4 + 3}{20} = 3,55$$

- Mencari standarisasi deviasi data V1 dengan persamaan (2)

$$Std(V1) = \sqrt{\frac{(V1_1 - \bar{V1})^2 + (V1_2 - \bar{V1})^2 + (V1_3 - \bar{V1})^2 + (V1_4 - \bar{V1})^2 + (V1_5 - \bar{V1})^2 + (V1_6 - \bar{V1})^2 + (V1_7 - \bar{V1})^2 + (V1_8 - \bar{V1})^2 + (V1_9 - \bar{V1})^2 + (V1_{10} - \bar{V1})^2 + (V1_{11} - \bar{V1})^2 + (V1_{12} - \bar{V1})^2 + (V1_{13} - \bar{V1})^2 + (V1_{14} - \bar{V1})^2 + (V1_{15} - \bar{V1})^2 + (V1_{16} - \bar{V1})^2 + (V1_{17} - \bar{V1})^2 + (V1_{18} - \bar{V1})^2 + (V1_{19} - \bar{V1})^2 + (V1_{20} - \bar{V1})^2}{N - 1}}$$

$$Std(V1) = \sqrt{\frac{(4 - 3,55)^2 + (3 - 3,55)^2 + (3 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (3 - 3,55)^2 + (3 - 3,55)^2 + (3 - 3,55)^2 + (3 - 3,55)^2 + (3 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (3 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (4 - 3,55)^2 + (3 - 3,55)^2}{20 - 1}}$$

Berdasarkan data siswa diatas, dapat disimpulkan bahwa dari *cluster* 0 adalah kelompok siswa yang memiliki nilai perhitungan yang tinggi dan *cluster* 1 adalah kelompok siswa yang memiliki nilai perhitungan yang rendah. Sehingga *cluster* 0 berhak masuk dikelas IPA

Tabel 2. Hasil *Cluster* Pengelompokan Siswa

Kode Huruf	Nama Siswa	No Cluster
E	Dede Syahbana Lbs	Cluster 0
G	Heri Surya	
S	Wahyu Ramadhan	
L	Riski Sembiring	
R	M uhamad ridwan	
O	Siska T. Br. Sitepu	
M	Rizaldi	
Q	Muhamad R. Fadli	
K	Prima Yusati	
N	Septian D.Herdi	
A	Adelia Syahfitri	Cluster 1
C	Aninditha C.Rahma	
H	Heru Wanda	
I	Joshua Siahaan	
F	Febri	
B	Andre Agasi	
T	Wiranata	
D	Ari Surya Pratama	
P	Muhamad I. Pranata	
J	Joko Barus	

## SIMPULAN

Dengan menerapkan metode pemecahan masalah dengan teknik *single linkage* maka SMA dapat mengetahui seberapa besar nilai pengelompokan jurusan, dan dapat melakukan pengujian dari hasil perhitungan data siswa di SMA untuk menentukan jurusan bidang *study* lebih cepat dan akurat.

## DAFTAR PUSTAKA

- Haviludin. (2011). *Memahami Penggunaan UML (Unified Modelling Language)*.
- Hermawati, Astuti, Fajar. (2013). *Data mining*. (Edisi 1). Yogyakarta: Andi.
- Susanto, S. & Suryadi, D.(2010). *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi
- Rossa, A. S., & Salahuddin, M (2014), *Rekayasa Perangkat Lunak*. Bandung: Informatika.