

IMPLEMENTASI DATA MINING DALAM PENGELOMPOKAN DATA PESERTA DIDIK DI SEKOLAH UNTUK MEMPREDIKSI CALON PENERIMA BEASISWA DENGAN MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS SMAN 16 BEKASI)

Anggoro Eko Wicaksono

*Jurusan Teknik Informatika, Universitas Gunadarma
anggoroeko17@gmail.com*

Abstrak

Data Mining merupakan suatu konsep yang digunakan untuk mencari nilai tambah yang tersembunyi dalam database dengan tujuan untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna. Pada penelitian ini dikembangkan suatu konsep data mining untuk membantu memprediksi peserta didik yang mendapatkan program beasiswa dengan menggunakan algoritma K-Means Clustering. Algoritma ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam cluster yang lain. Aplikasi data mining ini menampilkan hasil berupa masing-masing data yang dikelompokkan berdasarkan nilai dan gaji orang tua sebagai pertimbangan untuk mendapatkan beasiswa. Pembuatan aplikasi data mining ini menggunakan bahasa pemrograman Java dan database MySQL, dan telah diuji coba dengan menggunakan metode black box testing. Aplikasi data mining ini dikembangkan untuk memudahkan dalam memprediksi peserta didik yang lulus beasiswa untuk sekolah tingginya karena adanya perbandingan antara nilai akademik dengan gaji orang tua yang dijadikan acuan.

Kata kunci : *Data Mining, Pengelompokan Data, Beasiswa, Algoritma K-Means.*

DATA MINING IMPLEMENTATION IN STUDENTS CLASSIFICATION DATA AT SCHOOL TO PREDICT FUTURE SCHOLARSHIP GRANTEE BY USING K-MEANS ALGORITHM (CASE STUDY: SMAN 16 BEKASI)

Abstract

Data mining is a concept to search hidden added-value in the database for extraction and identify potential, useful information purposes. In this research, the data-mining concept developed to help predict potential students to be future scholarship grantee by using K-Means Clustering algorithm. The algorithm partitioning the data into cluster based on characteristics, so the data that have similar characteristics placed in the same group, while the data that has different characteristics grouped in another group. The data-mining application displays the result in the form of individual data grouped based on grades and parents income as consideration for gaining scholarship. The data mining application is build using Java programming and MySQL database. The application tested using the black box testing method. This data mining application developed to predict students to be scholarship grantee in their future studies based on the comparisons between their academic grades and parent's income as references.

Keywords: *Data Mining, Data Grouping, Scholarships, K-Means algorithm.*

PENDAHULUAN

Beasiswa adalah pemberian bantuan berupa keuangan yang diberikan kepada perorangan, pelajar atau mahasiswa yang digunakan demi keberlangsungan pendidikan yang ditempuh. Menurut Murniasih (2009) beasiswa diartikan sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi [12]. Penghargaan itu dapat berupa akses tertentu pada suatu institusi atau penghargaan berupa bantuan keuangan. Umumnya pemberian beasiswa diberikan berdasarkan besarnya gaji orang tua atau kepada para peserta didik yang memiliki nilai akademiknya baik di sekolah.

Sekolah merupakan media yang dijadikan oleh para peserta didik untuk menempuh pendidikannya. Setiap sekolah seharusnya dapat memantau para peserta didiknya agar dapat mengarahkan peserta didiknya untuk menjadi lebih baik, seperti untuk mendapatkan program beasiswa yang diberikan kepada peserta didik yang memiliki nilai akademiknya baik atau kepada peserta didik yang orang tuanya berpenghasilan di bawah rata-rata. Salah satu sekolah yang harus menerapkan ini adalah SMAN 16 Bekasi.

SMAN 16 Bekasi merupakan sebuah sekolah negeri yang terletak di Jalan Arteri Toll JORR Jatimelati, Kota Bekasi. Dalam melaksanakan kegiatan belajar mengajarnya, saat ini SMAN 16 Bekasi dibantu dengan 36 guru, dan memiliki kelas yang berjumlah 27 kelas (<http://sekolah.data.kemdikbud.go.id/>). Berdasarkan informasi yang diperoleh dari (<http://www.sman16bekasi.sch.id>) pada tahun ajaran 2015/2016 SMAN 16 Bekasi memiliki peserta didik yang berjumlah 1000 peserta didik untuk kelas 10, 11, dan 12.

Sampai dengan saat ini, SMAN 16 Bekasi sudah meluluskan 6 angkatan sejak tahun 2008, dan sudah menunjukkan kreatifitasnya dalam pengembangan bi-

dang pendidikan dengan memanfaatkan infrastruktur yang dimilikinya yang dibuktikan dengan meningkatnya jumlah peserta didik pada tahun 2011 – 2013, sehingga dapat meluluskan para peserta didiknya hingga 100 persen. Selama 6 tahun terakhir terdapat sekitar 1155 peserta didik kelas 12 yang lulus dari SMAN 16 Bekasi, namun selama ini proses pemantauan peserta didik dilakukan secara manual.

Dengan banyaknya jumlah peserta didik, tentu pemantauan yang dilakukan secara manual tidak akan efektif, sehingga peserta didik yang memiliki nilai akademiknya baik atau yang orang tuanya berpenghasilan kurang dari cukup tidak semuanya terpantau dan sulit diprediksi untuk mendapatkan beasiswa setelah lulus dari sekolah. Oleh karena itu dibutuhkan suatu metode untuk mengelompokkan peserta didik tersebut apakah layak mendapatkan beasiswa berdasarkan nilai akademiknya atau berdasarkan gaji orang tuanya. Metode yang digunakan adalah *data mining*.

Data mining adalah bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil [10]. Pemanfaatan *datamining* berguna untuk menentukan kelompok dari nilai – nilai akademik peserta didik beserta gaji orang tuanya, sedangkan algoritma yang digunakan untuk proses pengelompokannya menggunakan algoritma *k-means*.

Algoritma *k-means* merupakan bentuk *clustering* non hirarki yang berusaha mempartisi data ke dalam bentuk satu atau lebih *cluster*. Dengan algoritma *k-means* data nilai akademik dan gaji orang tua peserta didik dikelompokkan berdasarkan titik tengah suatu *cluster*.

Dalam implementasinya menggunakan algoritma *k-means clustering*. Penelitian dilakukan untuk mencari kelompok

nilai akademik dan gaji orang tua dengan cara menghitung jarak antara data dengan titik tengah suatu *cluster*. Sehingga pengelompokan tersebut menghasilkan suatu kelompok yang paling berhak mendapatkan program beasiswa setelah lulus dari SMAN 16 Bekasi.

METODE PENELITIAN

Dalam bab ini dibahas metode-metode yang digunakan dalam penelitian seperti data mining, clustering, euclidean distance dan algoritma k-means.

Data Mining

Data mining adalah suatu konsep yang digunakan untuk mencari nilai tambah yang tersembunyi dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk ekstraksi dan mengidentifikasi informasi pengetahuan potensial dan bermanfaat yang tersimpan di dalam *database* besar.

Data mining adalah bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil [10]. Teknik *data mining* secara garis besar dibagi dalam dua kelompok, yaitu verifikasi dan *discovery*. Metode verifikasi meliputi teknik-teknik statistik seperti *goodness of fit*, dan analisis variansi. Sedangkan, metode *discovery* dibagi lagi menjadi model prediktif dan model deskriptif. Model prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda. Model prediktif dapat dibuat berdasarkan penggunaan data historis lain. Sementara itu, model deskriptif bertujuan mengidentifikasi pola-pola atau hubungan antar data dan memberikan cara untuk mengeksplorasi karakteristik data yang diselidiki [4].

Clustering

Clustering adalah proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelompok-kelompoknya berdasarkan tingkat kesamaan. Sehingga dalam suatu *cluster* merupakan data dengan karakteristik yang hampir sama. Analisis *clustering* adalah sebuah teknik dari analisis *multivariable* yang digunakan untuk mengelompokkan obyek-obyek (*variable* atau data) sehingga dapat menghasilkan suatu informasi untuk membantu pelaksanaan pengujian terhadap obyek dan pada akhirnya dapat menyajikan suatu hipotesis berdasarkan relasi yang terjadi. Tujuan dari teknik ini adalah untuk melakukan pengelompokan berdasarkan kriteria tertentu sehingga obyek-obyek tersebut mempunyai variasi di dalam *cluster* relatif kecil dibandingkan variasi antar *cluster*. Untuk mengukur tingkat kesamaan tersebut digunakan suatu rumus untuk perhitungan jarak, rumus yang sering digunakan adalah rumus jarak *euclidean*.

Euclidean Distance

Metode yang umumnya digunakan dalam pengukuran jarak antara *data point* salah satunya adalah *Euclidean Distance*. *Euclidean distance* merupakan metode yang dikembangkan berdasarkan teorema *pythagoras*. Rumus *Euclidean Distance*:

$$d(x_j - c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

x_j = *Data Point*, dan c_j = *Centroid*

Algoritma K-Means

K-Means merupakan salah satu metode *clustering* non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Algoritma ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu

cluster yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain. Algoritma ini sederhana untuk diterapkan dan dijalankan, relatif cepat, mudah diadaptasi, dan umum digunakan dalam praktek. Berikut ini adalah tahap-tahap algoritma *K-Means clustering* :

1. *Input* data yang digunakan dalam *clustering*. Data ini digunakan untuk menentukan nilai rata-rata *data point* yang berada dalam satu *cluster* dan menentukan jarak dari setiap *data point* ke *centroid*.
2. Alokasi ke *Cluster* secara acak, dalam tahap ini pertama kalinya *data point* dialokasikan ke *cluster* secara acak tanpa ada kriteria tertentu.
3. Hitung *centroid data point* yang ada pada setiap *cluster*. Nilai *centroid* pada *k-means* digunakan sebagai pusat *cluster*. Dengan menentukan anggota *cluster* secara acak pada tahap sebelumnya, maka terbentuk iterasi awal sebagai pusat *cluster* acak.
4. Alokasi ke *centroid* terdekat, pada tahap ini hasil *centroid* dari setiap *cluster* sudah diketahui, kemudian *datapoint* dialokasikan pada *centroid* terdekat berdasarkan nilai jarak *similarity data point* terhadap *centroid*. Jarak *similarity* dari *data point* ke *centroid* pada masing-masing *cluster* diperoleh dari perhitungan *Euclidean Distance*. Kemudian nilai jarak setiap *data point* ke *centroid cluster* dibandingkan, dan *data point* menjadi anggota dari *cluster* berdasarkan jarak *data point* ke *centroid* terdekat.
5. Konvergen, mengalokasikan *data point* ke *centroid* dengan nilai jarak terdekat, dengan menguji apakah *cluster* yang terbentuk telah membentuk *cluster* yang konvergen atau tidak. *Cluster* dinyatakan konvergen jika anggota dari masing-

masing *cluster* yang terbentuk tidak mengalami perubahan anggota. Tetapi jika masih terjadi perubahan anggota *cluster*, maka akan kembali dilakukan tahapan menghitung *centroid* dari masing-masing *cluster* yang terbentuk dan diikuti dengan perhitungan nilai *similarity* ke *centroid* yang baru terbentuk. Proses tersebut terus berulang sampai hasil *cluster* konvergen.

PEMBAHASAN

Pada pembahasan akan dibahas hasil analisis kebutuhan fungsional, analisis sistem, dan analisis dengan algoritma K-means.

Analisis Kebutuhan Fungsional

Kebutuhan fungsional memiliki cakupan aspek bagaimana sistem dapat berelasi dengan perangkat lunak, dan bagaimana perilaku sistem pada situasi tertentu. Kebutuhan fungsional tersebut antara lain :

1. Memasukan data peserta didik ke dalam *database*.
2. Mengambil data nilai akademik peserta didik dan gaji orang tua yang sudah ada untuk dihitung dengan metode *k-means*.
3. Menampilkan hasil pengelompokan dalam bentuk tabel.

Analisis Sistem Berjalan pada SMAN 16 Bekasi

Pada saat melakukan observasi langsung ke SMAN 16 Bekasi yang beralamatkan di jalan Tol Jor, Jatimelati, Pondok Melati, Bekasi. Diamati bahwa dalam memprediksi calon penerima beasiswa SMAN 16 Bekasi belum memiliki sebuah sistem yang khusus ditujukan untuk melakukan pengelompokan berdasarkan data-data peserta didik. Proses pengelompokan data peserta didik di SMAN 16 Bekasi terdiri dari beberapa tahapan. Tahap yang pertama

adalah mengumpulkan data peserta didik berdasarkan nilai akademik. Tahapan berikutnya adalah data nilai akademik tersebut diseleksi secara manual, lalu apakah sesuai kriteria atau tidak, jika ya maka lulus.

Analisis dengan Algoritma K-Means

Dalam algoritma *k-means* banyaknya *cluster* ditentukan terlebih dahulu, dengan syarat banyaknya *cluster* harus lebih kecil dari banyaknya data ($k < n$). Banyaknya *cluster* pada aplikasi *data mining* ini adalah 2 ($k=2$). Selanjutnya menentukan *centroid* awal dengan metode mengambil data dari *data point* secara acak. *Centroid* merupakan titik tengah suatu *cluster* yang berupa nilai, dalam algoritma *k-means* untuk menentukan *centroid* awal ditentukan dengan mengambil *data point* secara acak (Tabel 1).

Sebanyak lima *data point* diambil sebagai *sample* dengan nilai dan gaji orang tua yang dikelompokkan diidentifikasi sebagai kelompok nilai dan kelompok gaji. Digunakan kelompok nilai dan kelompok gaji agar memudahkan dalam

perhitungan jarak data dengan *centroid* dengan menggunakan rumus *euclidean distance*. Kelompok nilai dan kelompok gaji harus sama jumlahnya, dan untuk memudahkan perhitungan diberikan kelompok nilai dan kelompok gaji sebanyak 1 – 4 kelompok yang masing-masing dibandingkan (Tabel 2).

C1 merupakan *centroid* pada *cluster* pertama, dan C2 merupakan *centroid* pada *cluster* dua yang ditentukan secara acak. Setelah *centroid* awal ditentukan, menghitung jarak data dengan *centroid* menggunakan rumus *euclidean distance*. Menghitung jarak data dengan *centroid* awal :

$$d(x_1, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_2, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_3, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_4, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_5, c_1) = \sqrt{(3 - 3)^2 + (2 - 1)^2} = 1$$

$$d(x_1, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_2, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_3, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_4, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_5, c_2) = \sqrt{(3 - 3)^2 + (2 - 2)^2} = 0$$

Tabel 1. *Data Point*

No	Kel. Nilai	Kel. Gaji
1	3	1
2	3	1
3	3	1
4	3	1
5	3	2

Tabel 2. *Centroid Awal*

	Kel. Nilai	Kel. Gaji
C1	3	1
C2	3	2

Tabel 3. Hasil Perhitungan Jarak Awal

No	Kel. Nilai	Kel. Gaji	DC1	DC2	C1	C2
1	3	1	0	1	V	
2	3	1	0	1	V	
3	3	1	0	1	V	
4	3	1	0	1	V	
5	3	2	1	0		V

Tabel 4. Centroid Pengulangan ke-1

	Kel. Nilai	Kel. Gaji
C1	3	1
C2	3	2

Tabel 5. Hasil Perhitungan Jarak Pada Pengulangan ke-1

No	Kel. Nilai	Kel. Gaji	DC1	DC2	C1	C2
1	3	1	0	1	V	
2	3	1	0	1	V	
3	3	1	0	1	V	
4	3	1	0	1	V	
5	3	2	1	0		V

Hasil perhitungan jarak antara *datapoint* dengan *centroid* awal dengan menggunakan rumus *euclidean distance*, dan untuk mencari *cluster* pertama atau dua dapat dilihat pada data yang memiliki nilai terkecil, hasil perhitungan jarak didefinisikan sebagai DC1 dan DC2 (Tabel 3).

Hasil perhitungan jarak *data point* dengan *centroid* awal dihasilkan, jika $DC1 < DC2$ maka masuk ke *cluster* pertama (C1). Namun jika $DC2 < DC1$ maka masuk ke *cluster* dua (C2). Untuk *centroid* selanjutnya ditentukan berdasarkan data yang diambil dari rata-rata masing-masing *cluster* (Tabel 4).

Menghitung jarak data dengan *centroid* pada pengulangan ke-1 :

$$d(x_1, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_2, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_3, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_4, c_1) = \sqrt{(3 - 3)^2 + (1 - 1)^2} = 0$$

$$d(x_5, c_1) = \sqrt{(3 - 3)^2 + (2 - 1)^2} = 1$$

$$d(x_1, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_2, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_3, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_4, c_2) = \sqrt{(3 - 3)^2 + (1 - 2)^2} = 1$$

$$d(x_5, c_2) = \sqrt{(3 - 3)^2 + (2 - 2)^2} = 0$$

Implementasi

a. Tampilan Login

Pada saat aplikasi dijalankan, proses *login* adalah proses pertama yang dijalankan untuk memastikan yang menggunakan aplikasi adalah guru dari SMAN 16 Bekasi. Tampilan login dapat dilihat pada Gambar 1. Halaman ini memiliki dua *button* yaitu, *login* dan *sign up*. Jika seorang guru belum terdaftar diharuskan melakukan proses *sign up* terlebih dahulu untuk bisa menggunakan aplikasi.



Gambar 1. Tampilan Proses Login



Gambar 2. Tampilan Proses Sign Up

No	No Peserta Didik	Nama	Jurusan	No1	No2	No3	Total No1,2,3	No4	No5	No6	No7	No8	No9	No10
2	10002	AGUS PRASEPTO	IPA	807	811	813	400000	3	1	V	V	V	V	
3	10003	ABIMANUJUSALDA	IPA	809	814	821	410000	3	1	V	V	V	V	
4	10004	ABIMANUJUSALDA	IPA	815	812	813	400000	3	1	V	V	V	V	
5	10005	ADHARULHAQ	IPA	812	817	812	400000	3	2	V	V	V	V	
6	10006	ADHARULHAQ	IPA	755	802	804	400000	3	1	V	V	V	V	
7	10007	ADHARULHAQ	IPA	757	802	802	400000	3	1	V	V	V	V	
8	10008	ADHARULHAQ	IPA	758	802	802	400000	3	1	V	V	V	V	
9	10009	ADHARULHAQ	IPA	800	842	822	400000	3	1	V	V	V	V	
10	10010	ADHARULHAQ	IPA	759	804	822	400000	3	1	V	V	V	V	
11	10011	ADHARULHAQ	IPA	759	802	802	400000	3	1	V	V	V	V	
12	10012	ADHARULHAQ	IPA	809	811	811	400000	3	1	V	V	V	V	
13	10013	ADHARULHAQ	IPA	810	811	811	400000	3	1	V	V	V	V	
14	10014	ADHARULHAQ	IPA	802	841	847	400000	3	1	V	V	V	V	
15	10015	ADHARULHAQ	IPA	772	812	812	400000	3	1	V	V	V	V	
16	10016	ADHARULHAQ	IPA	735	802	791	400000	3	1	V	V	V	V	
17	10017	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
18	10018	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
19	10019	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
20	10020	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
21	10021	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
22	10022	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
23	10023	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
24	10024	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
25	10025	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
26	10026	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
27	10027	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
28	10028	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
29	10029	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	
30	10030	ADHARULHAQ	IPA	789	802	802	400000	3	1	V	V	V	V	

Gambar 3. Tampilan Kelompok Data

b. Tampilan Sign Up

Sign Up merupakan proses untuk melakukan pendaftaran yang ditujukan hanya untuk guru di SMAN 16 Bekasi, agar program aplikasi data mining ini tidak sembarang pengguna yang menggunakannya. Proses ini berisi data guru seperti id guru, nama, username, dan password yang digunakan untuk proses login yang dapat dilihat pada Gambar 2.

c. Tampilan Beranda

Setelah login, akan tampil halaman beranda yang merupakan halaman utama pada aplikasi data mining. Pada halaman utama terdapat limamenu, yaitu menu

kelompok, menu data, menu tentang, menu bantuan, dan menu keluar.

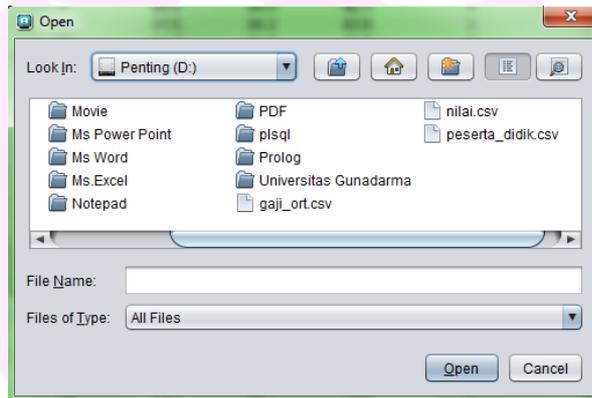
d. Tampilan Pengelompokan Data

Pada tampilan pengelompokan data adalah dimana proses pengelompokan data peserta didik dilakukan, data ditampilkan sesuai dengan kebutuhan guru di SMAN 16 Bekasi. Setelah data ditampilkan guru dapat melakukan proses pengelompokan data tersebut berdasarkan nilai akademik dan gaji orang tua untuk mendapatkan hasil yang akurat siapa saja peserta didik yang diprediksi mendapatkan program beasiswa untuk melanjutkan sekolah tingginya.

No	ID Peserta Didik	Nama	Jenjang	ID Nilai	Nilai UM	Nilai UN	Total Nilai
1	10002	KUT PRASETYO	PA	2	82,7	81,2	82,0
2	10003	AFI RAHMAN RAHMAN	PA	3	89,0	84,4	87,2
3	10004	PRAM MULYADIR	PA	4	81,0	80,2	80,6
4	10005	CIPTA FEBRIYATI ANINDY A	PA	5	81,0	81,7	81,4
5	10006	DIALIA KURNIAH	PA	5	79,0	80,2	80,4
6	10007	ULYAH ANAM KURNIA	PA	7	79,0	80,0	80,0
7	10008	DEVI ARIYANI	PA	8	78,0	80,7	80,1
8	10009	YANI ANISA KURNI	PA	8	80,0	84,0	82,0
9	10010	FAHRIYAH	PA	9	80,0	84,0	82,0
10	10011	FAZZAH	PA	11	79,0	81,0	80,0
11	10012	FAHRIYAH	PA	12	80,0	84,0	82,0
12	10013	REOVITA SIBEHING KUBIC	PA	13	83,0	84,0	84,0
13	10014	IGIYAH KURNI	PA	14	80,0	84,0	82,0
14	10015	IGIYAH PRABODOC	PA	15	77,0	81,0	81,0
15	10016	IGIYAH PRABODOC	PA	16	72,0	80,0	76,0
16	10017	IBRA AFFENDYALUR	PA	17	78,7	80,0	80,0
17	10018	ADARIFERUSALAH	PA	18	78,0	84,4	81,2
18	10019	BAGHYA A CHAN ANISA	PA	18	78,7	80,0	79,4
19	10020	REOVITA SIBEHING KUBIC	PA	20	80,0	84,4	82,2
20	10021	SELAMATI HUSNATI KURNIA S	PA	21	77,0	80,0	78,5
21	10022	YUSUF PURNAMA	PA	22	73,0	84,0	79,0
22	10023	YUSUF PURNAMA	PA	23	81,0	81,0	81,0
23	10024	POETRY RAHMAWATI LILAH	PA	24	73,1	84,1	78,6
24	10025	HELIMANORAH	PA	25	82,0	80,4	81,2
25	10026	SHAFITA RINI TURAH	PA	26	78,0	81,4	80,1

Gambar 4. Tampilan Halaman Data

Gambar 5. Tampilan Form Muat Data



Gambar 6. Tampilan File Direktori Komputer

e. Tampilan Data

Pada Halaman data, semua data peserta didik ditampilkan, dan guru dapat mengolah data peserta didik, seperti memuatdata dari file .csv ke MySQL yang digunakan untuk memasukan banyak data, sehingga lebih cepat dan mudah. Selain itu, dapat melakukan tambah data dan ubah data secara satu persatu, serta hapus data secara langsung semua data atau secara satu persatu (Gambar 4).

Pada Gambar 6, data yang dimasukkan dari file .csv sehingga dapat melakukan proses tambah data dengan data yang sudah ada di sekolah. Pada tampilan

form muatdata terdapat dua button, yaitu button memuat data ke MySQL dan button cari. Button cari membuka file direktori di komputer untuk mencari file .csv, sedangkan button memuat data ke MySQL untuk memproses file .csv yang dimasukkan ke database MySQL.

f. Tampilan Hapus Data

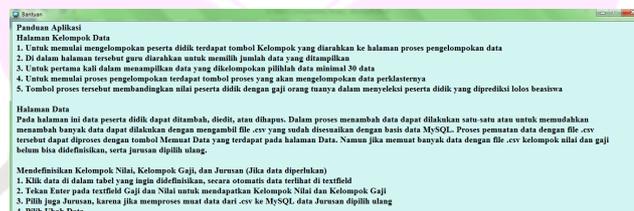
Tampilan ini menjelaskan mengenai proses penghapusan data peserta didik yang dapat dilakukan secara langsung semua data atau secara satu persatu. Pada tampilan hapusdata terdapat dua button dan satu textfield seperti pada Gambar 7.



Gambar 7. Tampilan Hapus Data



Gambar 8. Tampilan Tentang



Gambar 9. Tampilan Bantuan

g. Tampilan Tentang

Pada halaman tentang menjelaskan mengenai program aplikasi *data mining* yang digunakan untuk proses pengelompokan data peserta didik berdasarkan nilai dan gaji orang tua seperti pada Gambar 8.

fungsi pada *button* yang tersedia di dalam aplikasi.

h. Tampilan Bantuan

Tampilan bantuan ini menjelaskan mengenai prosedur dalam menjalankan aplikasi *data mining*, seperti fungsi-

SIMPULAN

Berdasarkan pada analisis dan pembahasan di bab sebelumnya, maka dapat diambil kesimpulan bahwa pembuatan aplikasi pengelompokan data peserta didik berdasarkan nilai dan gaji orang tua dengan menggunakan algoritma *K-Means clustering* telah berhasil dibuat. Dengan menggunakan algoritma ini membantu guru dalam memproses data

nilai akademik dan gaji orang tua peserta didik sehingga dalam menentukan calon penerima beasiswa dapat dengan mudah diprediksi, dan menghindari kesalahan dalam proses seleksinya. Bahasa yang digunakan dalam mengembangkan aplikasi ini adalah bahasa pemrograman Java dengan MySQL sebagai basis datanya.

Aplikasi ini telah diuji coba dengan menggunakan metode *black box testing*. Uji coba yang dilakukan yaitu pada fungsi halaman *login*, *sign up*, halaman beranda, halaman kelompok data peserta didik, dan halaman data peserta didik. Hasil dari pengujian pada setiap halaman ini, semua halaman berhasil dijalankan dengan baik.

Berdasarkan hasil kuisioner yang dibagikan kepada 10 responden secara acak bahwa dengan adanya aplikasi *data mining* dalam memprediksi beasiswa ini membantu guru dalam melakukan seleksi data peserta didik dalam memprediksi calon penerima beasiswa, hal ini ditunjukkan dengan hasil dari *User Acceptance Test* yang menghasilkan sikap guru sangat positif terhadap aplikasi ini.

DAFTAR PUSTAKA

- [1] A.S Rosa dan Salahuddin M. 2011. "*Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)*". Bandung : Modula.
- [2] Bertens, K. 2002. "*Etika*". Jakarta : Gramedia Pustaka.
- [3] Djaali. 2008. "*Skala likert*". Jakarta: PustakaUtama.
- [4] Dunham, M.H. 2003. "*Data mining introductory and advanced topics*". Upper Saddle River, NJ: Pearson Education, Inc.
- [5] Efraim, Turban. Rainer, Kelly R dan Potter, Richard. 2005. "*Introduction to Information Technology*". 3rd Edition. USA : John Willey & Sons, Inc.
- [6] Hariyadi, Teguh. Suprayogi. 2012. "*Penerapan Algoritma K-Means Untuk Pengelompokan Data Nilai Siswa*". Semarang : Universitas Dian Nuswantoro.
- [7] Hariyanto, Bambang. 2011. "*Esensi-esensi Bahasa Pemrograman JAVA*". Bandung : Informatika.
- [8] HM, Jogiyanto. 2003. "*Analisis dan Desain Sistem Informasi*". Yogyakarta : ANDI
- [9] Iswari, Ni Made Satvika. 2015. "*Review Perangkat Lunak StarUML Berdasarkan Faktor Kualitas McCall*". Universitas Multimedia Nusantara.
Volume: 7, No. 1, http://library.umn.ac.id/jurnal/index.php/paper/show_paper/224, April 2016.
- [10] Maimon, O. dan Last, M. 2000. "*Knowledge Discovery and Data Mining, The Info-Fuzzy Network (IFN) Methodology*". Dordrecht: Kluwer Academic.
- [11] Mulyana, Rohmat. 2004. "*Mengartikulasikan Pendidikan Nilai*". Bandung : Alfabeta CV.
- [12] Murniasih, Erny. 2009. "*Buku Pintar Beasiswa*", Jakarta : Gagasan Media
- [13] Nety, Cofriyanti Ervi, dan Angraini Renni. 2003. "*Perancangan Aplikasi Rental Mobil Pada CV Karya Bersama Palembang*". STMIK-MDP.
- [14] Ong, Johan Oscar, "*Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University*", Jurnal Ilmiah Teknik Industri, vol. 12, no. 1, <http://journals.ums.ac.id/index.php/jiti/article/download/651/392>, Juni 2013.
- [15] Pratama, I Putu Agus Eka. 2014. "*Sistem Informasi dan Implementasinya*". Bandung : Informatika.
- [16] Raharjo, Budi. 2011. "*Belajar Otodidak Membuat Database*

- Menggunakan MySQL*". Bandung : Informatika.
- [17] Rizky. 2011. "*Konsep Dasar Rekayasa Perangkat Lunak*". Jakarta : PT Prestasi Pustakarya.
- [18] Sulianta, Feri. 2008. "*Seri Referensi Praktis Manajemen IT*". Jakarta : PT Elex Media Komputindo.
- [19] Syukriyawati, Gusnia, et al. 2012. "*Penerapan metode k-mean clustering untuk pengelompokan mahasiswa universitas brawijaya tahun 2012*". Malang : Universitas Brawijaya.

