

PENERAPAN ALGORITMA K-MEANS PADA SISWA BARU SEKOLAH MENENGAH KEJURUAN UNTUK CLUSTERING JURUSAN

Fauziah Nur¹, Prof. M. Zarlis², Dr. Benny Benyamin Nasution³
Program Studi Magister Teknik Informatika, Universitas Sumatera Utara
Jl. Dr. T. Mansyur No. 9 Kampus USU Medan – 20155
nur4ziah@gmail.com¹, m.zarlis@yahoo.com², bnasution@yahoo.com³

Abstrak—Data mining merupakan teknik pengolahan data dalam jumlah besar untuk pengelompokan. Teknik ini digunakan dalam proses Knowledge Discovery in Database (KDD). Teknik tersebut mempunyai beberapa metode dalam pengelompokannya Naïve-Bayes dan Nearest Neighbour, pohon keputusan (KD-Tree), ID3, K-Means, text mining dan dbscan. Dalam hal ini penulis mengelompokan data siswa baru sekolah menengah kejuruan tahun ajaran 2014/2015. Pengelompokan tersebut berdasarkan kriteria – kriteria data siswa. Pada penelitian ini, penulis menerapkan algoritma K-Means Clustering untuk pengelompokan data siswa baru sekolah menengah kejuruan. Dalam hal ini, pada umumnya untuk memasuki jurusan hanya disesuaikan dengan nilai siswa saja namun dalam penelitian ini pengelompokan disesuaikan kriteria – kriteria siswa seperti penghasilan orang tua, tanggungan anak orang tua dan nilai tes siswa. Penulis menggunakan beberapa kriteria tersebut agar pengelompokan yang dihasilkan menjadi lebih optimal. Tujuan dari pengelompokan ini adalah terbentuknya kelompok jurusan pada siswa yang menggunakan algoritma K-Means clustering. Hasil dari pengelompokan tersebut diperoleh tiga kelompok yaitu kelompok tidak lulus, kelompok rekayasa perangkat lunak dan kelompok teknik komputer jaringan. Terdapat pusat cluster dengan Cluster-1=1.4;2.2;2.2, Cluster-2=2.28;1.64;4 dan Cluster-3=5;2;6. Pusat cluster tersebut didapat dari beberapa iterasi sehingga menghasilkan pusat cluster yang optimal.

Keywords— Kriteria, K-Means, Data Mining, Jurusan

I. PENDAHULUAN

Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*) dan suatu metode untuk mencari dan mengelompokan data yang memiliki kemiripan karakteristik antara satu data dengan data lain[1]. Menurut kategori kekompakan, pengelompokan terbagi menjadi dua, yaitu komplet dan parsial. Jika semua data dapat bergabung menjadi satu, dapat dikatakan semua data kompak menjadi satu kelompok[2]. Pada clustering ini terdapat beberapa algoritma pengelompokan untuk mengelompokan data secara mudah. Salah satunya adalah algoritma K-Means yang merupakan metode analisis kelompok yang mengarah pada partisian N objek pengamatan ke dalam K kelompok, di mana setiap objek pengamatan sebuah kelompok data dengan mean (rata-rata) terdekat[2].

Pengelompokan pada umumnya diterapkan untuk mengelompokan dokumen atau benda yang tidak tersusun dengan rapi dan tidak sesuai susunannya pada tempatnya. Namun fungsi dari pengelompokan tidak hanya sekedar mengelompokan dokumen atau benda. Pengelompokan (*clustering*) dapat diterapkan dalam hal penentuan jurusan sekolah agar jurusan tersebut sesuai dengan kemampuan siswa.

Banyak lembaga pendidikan seperti sekolah menengah kejuruan yang hanya menggunakan nilai

sebagai tolak ukur untuk memilih jurusan, sehingga pada akhirnya siswa menjadi salah mengambil jurusan dan tidak dapat menyesuaikan kemampuan yang siswa miliki dengan jurusan yang telah diambil. Dalam hal ini, untuk pengelompokan jurusan yang sesuai dengan kemampuan siswa adalah dapat ditentukan dengan data – data setiap siswa. Data yang dapat menjadi tolak ukur untuk pengelompokan jurusan yaitu nilai tes masuk sekolah menengah kejuruan, penghasilan orang tua, dan tanggungan anak orang tua. Dengan demikian pengelompokan jurusan yang diambil siswa akan lebih efektif bagi siswa itu sendiri.

Algoritma K-Means dalam hal ini akan mengelompokan data tersebut sesuai kriteria – kriteria yang dipilih dari data siswa.

II. PENDAHULUAN

A. Data Mining

Istilah data mining mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis hingga medis[2]. Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Dalam berbagai literatur, teori-teori pada data mining sudah ada sejak lama seperti antara lain Naïve-Bayes dan Nearest Neighbour, Pohon Keputusan, aturan asosiasi, K-Means Clustering dan text mining[3].

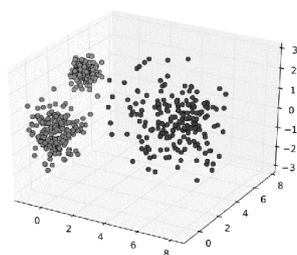
Data mining disebut juga dengan *knowledge-discovery in database (KDD)* ataupun *pattern recognition*. Istilah KDD atau disebut penemuan pengetahuan data karena tujuan utama data mining adalah untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna. Sedangkan istilah *pattern recognition* atau disebut pengenalan pola mempunyai tujuan pengetahuan yang akan digali dari dalam bongkahan data yang sedang dihadapi.

B. Dasar Cluster

Analisis kelompok (*cluster analysis*) adalah mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya[1]. Analisis *Cluster* sebagai metodologi untuk klasifikasi data secara otomatis menjadi beberapa kelompok dengan menggunakan ukuran asosiasi, sehingga data yang sama berada dalam satu kelompok yang sama dan data yang berbeda berada dalam kelompok data yang tidak sama.

Masukan (*input*) untuk sistem analisis *cluster* adalah seperangkat data dan kesamaan ukuran (atau perbedaan) antara dua data. Sedangkan keluaran (*output*) dari analisis *cluster* adalah sejumlah kelompok yang membentuk sebuah partisi atau struktur partisi dari kumpulan data. Salah satu hasil tambahan dari analisis *cluster* adalah deskripsi umum dari setiap *cluster* dan hal itu sangat penting untuk analisis lebih dalam dari karakteristik data set tersebut.

Ada saatnya di mana set data yang akan diproses dalam data mining belum diketahui label kelasnya. Pengelompokan data dilakukan dengan menggunakan algoritma yang sudah ditentukan dan selanjutnya data akan diproses oleh algoritma untuk dikelompokkan menurut karakteristik alaminya.



Gbr. 1 Pengelompokan dengan *clustering*

C. Algoritma K-Means

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *K cluster* yang sudah ditetapkan di awal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek[2]. *K-Means* dapat diterapkan pada data yang direpresentasikan dalam *r*-dimensi ruang tempat. *K-means* mengelompokkan set data *r*-dimensi, $X =$

$\{x_i | i=1, \dots, N\}$. Algoritma *K-Means* mengelompokkan semua titik data dalam *X* sehingga setiap titik x_i hanya jatuh dalam satu *K* partisi[4].

Tujuan pengelompokan ini adalah untuk meminimalkan fungsi objek yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antarkelompok. Parameter yang harus dimasukkan ketika menggunakan algoritma *K-Means* adalah nilai *K*. Nilai *K* yang digunakan pada umumnya didasarkan pada informasi yang diketahui sebelumnya mengenai sebenarnya berapa banyak *cluster* yang muncul dalam *X*, berapa banyak yang digunakan untuk penerapannya, atau jenis *cluster* dicari dengan melakukan percobaan dengan beberapa nilai *K*. Set representatif *cluster* dinyatakan $C = \{c_j | j=1, \dots, K\}$. sejumlah *K* representatif *cluster* tersebut sebagai *cluster centroid* (titik pusat *cluster*). Untuk set data dalam *X* dikelompokkan berdasarkan konsep kedekatan atau kemiripan, namun kuantitas yang digunakan untuk mengukurnya adalah ketidakmiripan. Metrik yang umum digunakan untuk ketidakmiripan tersebut adalah *Euclidean*.

Secara umum algoritma *K-Means* memiliki langkah-langkah dalam pengelompokan, diantaranya:

- 1) Inisialisasi: menentukan nilai *K centroid* yang diinginkan dan metrik ketidakmiripan (jarak) yang diinginkan.
- 2) Memilih *K* data dari set *X* sebagai *centroid*. Untuk menentukan *centroid* dapat menggunakan persamaan (1).

$$\frac{\text{Jumlah data}}{\text{Jumlah class} + 1} \quad (1)$$

- 3) Mengalokasikan semua data ke *centroid* terdekat dengan metrik jarak yang telah ditetapkan.
- 4) Menghitung kembali *centroid C* berdasarkan data yang mengikuti *cluster* masing – masing.
- 5) Mengulangi langkah 3 dan 4 hingga kondisi konvergen tercapai.

Berikut ini adalah rumus untuk menentukan jumlah *cluster*:

$$K = \sqrt{\frac{N}{2}} \quad (2)$$

Keterangan:

- K = klaster
- N = jumlah data

Menghitung jarak pada ruang jarak *Euclidean* menggunakan formula:

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \quad (3)$$

Keterangan:

- D* = euclidean distance
- x* = banyaknya objek
- Σ^p = jumlah data record

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengelompokkan data siswa baru sekolah menengah kejuruan dengan menggunakan algoritma *K-Means*. Agar tercapai tujuan tersebut, penulis akan melakukan pengujian dengan menggunakan data pada salah satu sekolah menengah kejuruan di medan yaitu SMK swasta medan area-1 tahun ajaran 2014/2015.

Adapun pengelompokan tersebut menggunakan langkah – langkah sebagai berikut:

A. Sumber Data

Dalam penelitian ini, sumber data diambil dari smk swasta medan area-1 dimana data tersebut merupakan data siswa baru tahun ajaran 2014/2015.

TABEL I
DATA SISWA BARU

No.	No. Registrasi	Nama	Tanggal Lahir	Jenis Kelamin	Alamat	Pekerjaan Orang Tua	Penghasilan orang tua	Tanggung jawab	Nilai tes
1	D320	Aman Gunadi	23-08-1998	L	Jln. PinangBari Gg. Bersama	Wiraswasta	1750000	2	80
2	D321	Desi Mulandani	04-12-1999	P	Jln. Binjai km 11	Wiraswasta	3000000	3	75
3	D322	Dita Risti Paralani	25-12-1999	P	Jln. Masjid Gg. Pibadi	Sales	2000000	1	80
4	D323	Fahrizal	23-09-1999	L	Jln. Serola Gg. Bersama	Wiraswasta	2500000	1	85
5	D324	Haisando	24-04-2000	L	Jln. Masjid km 2	Guru	2500000	2	70
6	D325	Jenika Ramadani	02-01-1999	P	Jln. Pungguk No. 30	Wiraswasta	2800000	2	75
7	D326	M. Abdu Mubarak	09-12-1999	L	Jln. Saetan Tanjung Gusta No. 8B	Wiraswasta	3500000	2	70
8	D327	Muhammad Ghopar	03-04-1999	L	Jln. Setia Maimun Sunggal Kanan	Kuli Bangunan	1500000	3	65
9	D328	M. Rinaldy A.S	11-04-2000	L	Jln. Kompos Gg. Asam	Wiraswasta	1800000	2	85
10	D329	M. Rizky Khalid	25-06-1999	L	Jln. Gelatik No. 8 Medan Sunggal	Wiraswasta	1700000	2	80
11	D330	M. Rizqi Pratama	03-07-1999	L	Jln. Si Guntung Block 8 No. 24	PNS	5000000	2	95
12	D331	M. Sidi Kadali	04-06-1999	L	Jln. Ganda No. 66	Wiraswasta	870000	2	60
13	D332	Maria Sri Rahayu	02-09-1999	P	Jln. PinangBari Gg. Bersama	PNS	4500000	1	70
14	D333	Nolizal	25-10-1998	L	Jln. Kamp. Lalong Gg. Amanah Medan Kiri	Supir	3000000	1	85
15	D334	Nugie Ramadhan	10-12-1999	L	Jln. Kertar Gg. Pinang Ayahanda	Wiraswasta	2000000	1	75
16	D335	Nurawandi	10-12-1999	L	Jln. Ganda No. 78 C	Wiraswasta	1600000	2	70
17	D336	Puji Sulaputri	06-07-1999	P	Jln. Beirigh	Wiraswasta	1500000	2	80
18	D337	Randika Frijanto	13-08-1999	L	Jln. Gaperia LK III No. 215 AA	Wiraswasta	300000	2	85
19	D338	Ricudidin	13-08-1999	L	Komplek PLH Paya Sari	Wiraswasta	350000	1	85
20	D339	Rinaldi Pane	27-11-1999	L	Jln. Balai Desa Gg. Wakaf	Buruh	2500000	2	85

B. Pengolahan Data

Sebelum data siswa baru tersebut dikelompokkan ke dalam jurusan yang sesuai kriteria. Data mentah tersebut akan ditransformasi dengan cara menginisialisasi data ke dalam bentuk angka yang dapat diolah dalam pengelompokan.

TABEL II
RANGE PENGELOMPOKAN JURUSAN

Cluster	Σ min	Σ max	cluster
1	0	7	Tidak Lulus
2	8	11	Rekayasa Perangkat Lunak (RPL)
3	12	15	Teknik Komputer Jaringan (TKJ)
4	16	18	Administrasi Perkantoran (AP)
5	19	20	Akuntansi (AK)
6	21	22	Multimedia (MM)

Berdasarkan *range* yang telah ditentukan pada pengelompokan jurusan, maka dapat disimpulkan bahwa pengelompokan data siswa baru ke dalam jurusan tanpa menggunakan metode.

Pada data tersebut, tidak semua kriteria yang dapat menjadi patokan. Dalam hal ini, kriteria yang diinisialisasikan adalah penghasilan orang tua menjadi

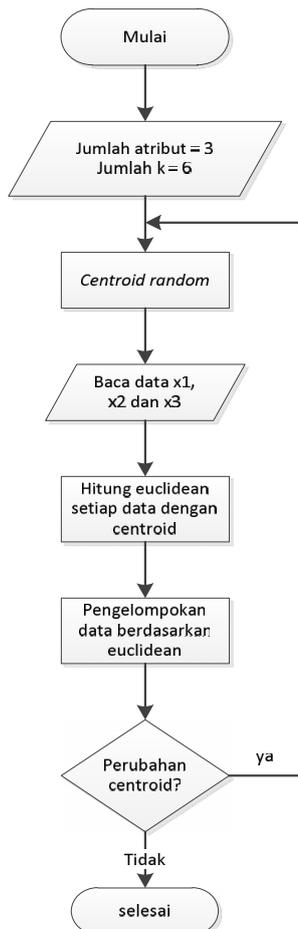
X1, Tanggungan anak diubah menjadi X2, dan hasil nilai tes diubah menjadi X3.

TABEL III
HASIL INISIALISASI DATA

No. pendaftar	X1	X2	X3	Jumlah	Cluster
D320	1	2	4	7	1
D321	3	3	3	9	2
D322	2	1	6	9	2
D323	2	1	5	8	2
D324	2	2	2	6	1
D325	3	2	3	8	2
D326	4	2	2	8	2
D327	1	3	1	5	1
D328	2	2	5	9	2
D329	2	2	4	8	2
D330	5	2	6	13	3
D331	1	2	1	4	1
D332	5	1	2	8	2
D333	3	1	5	9	2
D334	2	1	3	6	1
D335	1	2	2	5	1
D336	1	2	4	7	1
D337	1	2	5	8	2
D338	1	1	5	7	1
D339	2	2	1	5	1

C. Proses Data

Setelah data diolah, maka langkah selanjutnya adalah data diproses untuk membentuk pengelompokan data ke dalam jurusan sesuai kriteria yang telah ditentukan. Data yang telah ditransformasi tersebut akan diproses dengan menggunakan sebuah algoritma pengelompokan (*cluster*) yaitu algoritma *K-means*. Berikut ini adalah sebuah *flowchart* dalam pengelompokan jurusan pada siswa baru sekolah menengah kejuruan.



Gbr. 2 Flowchart Pengclusteran K-Means

Pada tahapan ini yaitu menentukan *centroid K-means* dari data siswa baru sekolah menengah kejuruan tahun ajaran 2014/2015. Dalam menentukan target *K-means* tersebut bertujuan untuk mendapatkan target data atau jarak kelompok yaitu titik pusat (*centroid*) kelompok awal untuk menghitung algoritma *K-means*.

Pada tabel 4 merupakan hasil perhitungan *K-means* untuk data sampel siswa baru tahun ajaran 2014/2015 pada sekolah menengah kejuruan medan area-1 yang berdasarkan kriteria – kriteria yang telah ditentukan untuk setiap kelompok (*cluster*).

Dalam hal ini, pengelompokan yang terbentuk terjadi sebanyak 3 kelompok (*cluster*). Untuk perhitungan secara lengkap terdapat pada tabel – tabel berikut:

TABEL IV
TARGET *K-MEANS* DATA SAMPEL SISWA BARU SMK
UNTUK C1

No. Pendaftaran	X1	X2	X3
D320	1	2	4
D324	2	2	2
D327	1	3	1
D331	1	2	1
D334	2	1	3
D335	1	2	2
D336	1	2	4
D338	1	1	5
D339	2	2	1
CLUSTER 1	1.33333	1.88889	2.55556

Setelah pengelompokan untuk kelompok pertama (*cluster-1*) telah terkelompok. Maka dapat dilakukan untuk pengelompokan data kelompok kedua (*cluster-2*).

TABEL V
TARGET *K-MEANS* DATA SAMPEL SISWA BARU SMK
UNTUK C2

No. Pendaftaran	X1	X2	X3
D321	3	3	3
D322	2	1	6
D323	2	1	5
D325	3	2	3
D326	4	2	2
D328	2	2	5
D329	2	2	4
D332	5	1	2
D333	3	1	5
D337	1	2	5
CLUSTER 2	2.7	1.7	4

Data yang dikelompokkan lainnya akan masuk pada kelompok 3 (*cluster 3*) yaitu hanya terdapat satu record data pada kelompok 3.

TABEL VI
TARGET *K-MEANS* DATA SAMPEL SISWA BARU SMK
UNTUK C3

No. Pendaftaran	X1	X2	X3
D330	5	2	6
CLUSTER 3	5	2	6

setelah semua data telah terkelompok, maka untuk menentukan titik pusat pada *cluster* dapat diambil nilai rata – rata dari setiap *cluster*.

Adapun tahapan proses algoritma *K-means* adalah sebagai berikut:

1. Menentukan nilai *k* dari jumlah *cluster* yang akan dibentuk. Pada tahapan ini *cluster* yang terbentuk sebanyak 3 *cluster*.
2. Menentukan titik pusat (*centroid*) awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan dengan menggunakan range dari jurusan.

TABEL VII
CENTROID AWAL CLUSTER

CLUSTER	X1	X2	X3
C1	1.33333	1.88889	2.55556
C2	2.7	1.7	4
C3	5	2	6

3. Setelah titik pusat *cluster* ditentukan, maka tahap selanjutnya adalah menghitung jarak terdekat atau *euclidean* dengan menggunakan persamaan 3, yaitu menghitung jarak dari data siswa baru pertama ke titik pusat *cluster*.

$$D(1,1)=\sqrt{(1-1.33)^2+(2-1.88)^2+(4-2.55)^2}=1.48657$$

Dari hasil perhitungan di atas didapatkan bahwa jarak terdekat bernilai 1.48657. Selanjutnya perhitungan jarak data pertama untuk *cluster* kedua.

$$D(1,2)=\sqrt{(1-2.7)^2+(2-1.7)^2+(4-4)^2}=1.72627$$

Dari hasil perhitungan tersebut didapatkan jarak terdekat bernilai 1.72627. Kemudian perhitungan jarak data pertama untuk *cluster* ketiga.

$$D(1,3)=\sqrt{(1-5)^2+(2-2)^2+(4-6)^2}=4.47214$$

Dari hasil perhitungan tersebut didapatkan jarak terdekat bernilai 4.47214. Berdasarkan perhitungan *euclidean* untuk data pertama diperoleh jarak terdekat *cluster* adalah 1.48657 yang terkelompok pada *cluster* kesatu (C1) dan untuk

TABEL VIII
PERHITUNGAN *EUCLIDEAN* ITERASI-1

No Pendaftaran	Kriteria			Cluster			Jarak Terdekat Ke Cluster
	X1	X2	X3	C1	C2	C3	
D320	1	2	4	1.48657	1.72627	4.47214	1
D321	3	3	3	2.0518	1.66733	3.74166	2
D322	2	1	6	3.61922	2.23159	3.16228	2
D323	2	1	5	2.68512	1.40712	3.31662	2
D324	2	2	2	0.87489	2.14009	5	1
D325	3	2	3	1.72848	1.08628	3.60555	2
D326	4	2	2	2.72619	2.40416	4.12311	2
D327	1	3	1	1.94047	3.68511	6.48074	1
D328	2	2	5	2.53616	1.25698	3.16228	2
D329	2	2	4	1.59474	0.76158	3.60555	2
D330	5	2	6	5.032	3.06268	0	3
D331	1	2	1	1.59474	3.46121	6.40312	1
D332	5	1	2	3.81356	3.1273	4.12311	2
D333	3	1	5	3.08921	1.25698	2.44949	2
D334	2	1	3	1.1967	1.40712	4.3589	1
D335	1	2	2	0.65734	2.64197	5.65685	1
D336	1	2	4	1.48657	1.72627	4.47214	1
D337	1	2	5	2.46957	1.99499	4.12311	2
D338	1	1	5	2.62232	2.09284	4.24264	2
D339	2	2	1	1.69604	3.09516	5.83095	1

Setelah data dikelompokkan pada iterasi pertama, maka langkah selanjutnya adalah membentuk titik pusat baru dengan menentukan nilai rata – rata dari setiap data yang sudah membentuk *cluster* untuk melanjutkan perhitungan jarak terdekat iterasi kedua dan akan menunjukkan *cluster* yang terbentuk selanjutnya dapat membentuk *cluster* yang konvergen.

TABEL IX
CENTROID BARU PERTAMA

CLUSTER	X1	X2	X3
C1	1.375	2	2.25
C2	2.54545	1.63636	4.09091
C3	5	2	6

Jika *cluster* belum konvergen, maka *centroid* akan dibangkitkan kembali dan menghitung kembali *euclidean* dari setiap data siswa. Langkah perhitungan jarak terdekat seperti langkah ketiga sebelumnya. Berikut perhitungan jarak terdekat untuk data kedua.

$$D(2,1)=\sqrt{(3-1.33)^2+(3-1.88)^2+(3-2.55)^2}=2.0518$$

Dari hasil perhitungan tersebut didapatkan jarak terdekat bernilai 2.0518.

Untuk perhitungan *euclidean* dapat dilihat secara lengkap pada tabel di bawah ini.

TABEL X
PERHITUNGAN *EUCLIDEAN* ITERASI-2

No Pendaftaran	Kriteria			Cluster			Jarak Terdekat Ke Cluster
	X1	X2	X3	C1	C2	C3	
D320	1	2	4	1.78973	1.59026	4.47214	2
D321	3	3	3	2.05015	1.80449	3.74166	2
D322	2	1	6	3.93105	2.08497	3.16228	2
D323	2	1	5	2.99218	1.2365	3.31662	2
D324	2	2	2	0.67315	2.19127	5	1
D325	3	2	3	1.78973	1.2365	3.60555	2
D326	4	2	2	2.63688	2.5729	4.12311	2
D327	1	3	1	1.64412	3.71506	6.48074	1
D328	2	2	5	2.82013	1.1208	3.16228	2
D329	2	2	4	1.85826	0.66183	3.60555	2
D330	5	2	6	5.21566	3.13076	0	3
D331	1	2	1	1.30504	3.47482	6.40312	1
D332	5	1	2	3.7687	3.28659	4.12311	2
D333	3	1	5	3.34711	1.19917	2.44949	2
D334	2	1	3	1.39754	1.3757	4.3589	2
D335	1	2	2	0.45069	2.62537	5.65685	1
D336	1	2	4	1.78973	1.59026	4.47214	2
D337	1	2	5	2.77545	1.82951	4.12311	2
D338	1	1	5	2.95011	1.90259	4.24264	2
D339	2	2	1	1.39754	3.15966	5.83095	1

Pada perhitungan jarak terdekat di atas *centroid* baru yang dibangkitkan ternyata belum konvergen, sehingga iterasi harus dilanjutkan. Dalam penelitian ini, iterasi *clustering* data siswa terjadi sebanyak 3 kali iterasi. Pada pengelompokan data di atas didapat titik *cluster* baru dengan nilai sebagai berikut:

TABEL XI
CENTROID BARU KEDUA

CLUSTER	X1	X2	X3
C1	1.4	2.2	1.4
C2	2.285714	1.6428571	4
C3	5	2	6

Centroid baru kedua yang dibangkitkan ternyata sudah konvergen, sehingga perhitungan jarak terdekat *clustering* berhenti pada iterasi-3. Adapun perhitungan secara lengkapnya sebagai berikut:

TABEL XII
PERHITUNGAN *EUCLIDEAN* ITERASI-3

No Pendaftaran	Kriteria			Cluster			Jarak Terdekat Ke Cluster
	X1	X2	X3	C1	C2	C3	
D320	1	2	4	2.638181	1.334396	4.47	2
D321	3	3	3	2.4	1.830858	3.74	2
D322	2	1	6	4.791659	2.120117	3.16	2
D323	2	1	5	3.841875	1.22266	3.32	2
D324	2	2	2	0.87178	2.05163	5	1
D325	3	2	3	2.271563	1.279748	3.61	2
D326	4	2	2	2.675818	2.658256	4.12	2
D327	1	3	1	0.979796	3.534812	6.48	1
D328	2	2	5	3.655133	1.099629	3.16	2
D329	2	2	4	2.675818	0.457366	3.61	2
D330	5	2	6	5.844656	3.390413	0	3
D331	1	2	1	0.6	3.283384	6.4	1
D332	5	1	2	3.841875	3.43229	4.12	2
D333	3	1	5	4.118252	1.386892	2.45	2
D334	2	1	3	2.088061	1.22266	4.36	2
D335	1	2	2	0.748331	2.40429	5.66	1
D336	1	2	4	2.638181	1.334396	4.47	2
D337	1	2	5	3.627671	1.667517	4.12	2
D338	1	1	5	3.815757	1.751093	4.24	2
D339	2	2	1	0.748331	3.034664	5.83	1

Pada iterasi-3 tersebut, titik pusat dari setiap *cluster* sudah tidak berubah dan tidak ada lagi terdapat data yang berpindah dari satu *cluster* ke *cluster* yang lain. Maka dari itu, data siswa baru SMK telah terkelompok sesuai kriteria yang ditentukan untuk memasuki

jurusan pada sekolah menengah kejuruan menggunakan algoritma *K-Means*.

IV. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, maka penulis dapat menarik beberapa kesimpulan, diantaranya sebagai berikut:

1. Pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* pada data siswa baru SMK terjadi sebanyak 3 kali iterasi.
2. Berdasarkan hasil *cluster* dengan menerapkan beberapa kriteria dari calon siswa menggunakan *K-Means* dapat diambil pengelompokan dengan rata – rata jurusan yang diambil adalah rekayasa perangkat lunak dan sedikit jumlah siswa yang tidak lulus. Bahkan ada beberapa jurusan yang tidak dibuka dikarenakan kriteria – kriteria siswa tidak dapat lulus dalam jurusan tersebut.

REFERENCE

- [1] Y. Agusta, "K-Means Penerapan, Permasalahan dan Metode Terkait," Jurnal Sistem dan Informatika, vol. 3, pp. 47-60, Pebruari 2007.
- [2] M. Kantardzic, J. Wiley and Sons, Data Mining: Concepts, Models, methods, and Algorithms, 2003
- [3] N. P. E. Merliana, Ernawati and A. J. Santoso, "Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering," *SENDI_U*.
- [4] Mardiani, "Perbandingan Algoritma K-Means dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah," *Citec Journal*, vol. 1, no. 4, pp. 316-325, Agustus-Oktober 2014.