

Rancang Bangun Aplikasi K-Means untuk Klasifikasi Kelulusan Siswa Sekolah Kepolisian Negara Daerah Kalimantan Selatan

Dwi Kartini¹, Muliadi²

^{1,2} Prodi Ilmu Komputer, Fakultas MIPA Universitas Lambung Mangkurat
Jl. A Yani km 36 Banjarbaru 70714

¹dwikartini@unlam.ac.id

²muliadiaziz@yahoo.com

Abstract -Student data and recapitulation data passing grade school students Regional State Police South Kalimantan each year increases, resulting in data that is very abundant in the form of the profile data of graduates and academic value data. The rapid growth of data accumulation has created conditions data-rich but minimal information. Data mining is a mining or the discovery of new information by looking for certain rules of a number of large amounts of data are expected to tackle the condition. By utilizing the students master data and data recapitulation of the value of the State Police School students Regional South Kalimantan in 2014, is expected to yield information on the classification of students passing through data mining techniques. The algorithm used within the classification graduation is K-Mean algorithm. The clustering of the results of the analysis carried out in the form of graduate classification based on the group. Group 1 with a recapitulation of the value of graduates being, group 2 with a recapitulation of the value of high graduates and group 3 with a recapitulation of the value of graduates is low. Group 1 with a value of clusters in the data center intellect 72.93 value, the value of personality 71.52, and 62.63 value physical health as much as 45%. Group 2 with a value of clusters in the data center intellect 73.33 value, the value of personality 73.79, and 70.80 value physical health as much as 35%. Group 3 with the center of the cluster in the data value 72.92 intellect, personality value of 69.95 and 53.64 value physical health as much as 20%.

Keywords: *clustering, K-Means algorithm, classification, graduation.*

I. Pendahuluan

Dalam dunia pendidikan, data yang berlimpah dan berkesinambungan mengenai siswa didik dan lulusan setiap tahunnya selalu bertambah. Salah satu lembaga pendidikan yang cukup besar di Indonesia adalah Sekolah Kepolisian Negara Daerah Kalimantan Selatan yang menghasilkan informasi yang berlimpah berupa jumlah kelulusan DIKTUK BRIGADIR Polri setiap tahunnya.

Data rekapitulasi nilai siswa hanya akan menjadi sekumpulan data yang tidak berguna jika tidak dilakukan penggalian data terhadapnya. Banyak informasi terpendam yang dapat diambil dari sekumpulan data tersebut sehingga dapat memberikan suatu pengetahuan untuk penentuan kebijakan. Penggalian data dapat dilakukan dengan cara pengelompokan data nilai siswa menjadi beberapa kelompok untuk dilakukan klasifikasi terhadap hasil pengelompokan tersebut.

Data mining sendiri merupakan sebuah proses ekstraksi informasi untuk menemukan pola (pattern recognition) yang penting dalam tumpukan data sehingga menjadi pengetahuan (knowledge discovery). Fungsi-fungsinya data mining antara lain

fungsi deskripsi, estimasi, predikis, clustering klasifikasi dan asosiasi [5]

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 (non hirarki) 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 [4]. Adapun 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 [3].

Penelitian ini akan melakukan pengelompokan menggunakan data rekapitulasi nilai siswa lulusan DIKTUK BRIGADIR Polri TA 2014. Atribut yang akan digunakan ialah nilai akhir intelek, kepribadian

dan kesehatan jasmani dengan jumlah kluster tiga. Berdasarkan latar belakang masalah diatas maka rumusan masalah pada penelitian ini adalah bagaimana menerapkan data mining menggunakan metode *K-Means* untuk menghasilkan informasi pengelompokan nilai lulusan siswa berdasarkan data rekapitulasi nilai akhir siswa DIKTUK BRIGADIR Polri TA 2014. Hasil klustering kelulusan siswa dapat dimanfaatkan oleh sekolah sebagai strategi dalam meningkatkan kualitas lulusan di masa akan datang.

II. Metodologi Penelitian

1.1 Kajian Pustaka

Beberapa Penelitian terdahulu mengenai datamining menggunakan k means telah dilakukan oleh Dian dengan judul Pengembangan Metode Klasifikasi berdasarkan k-means dan lvq [3]. Pada tahun 2016 dengan judul Klasifikasi Data Mahasiswa Menggunakan metode k-means untuk menunjang pemilihan strategi pemasaran [8] dan penelitian pada tahun 2014 oleh muzakir dengan judul analisa Pemanfaatan algoritma K-Means Clustering pada data Nilai Siswa Sebagai Penentu Penerima Beasiswa[1].

2.2 Pengertian Data Mining

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [8]. *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [6]. *Data mining*, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [7].

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* [4]. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Karakteristik *data mining* sebagai berikut :

1. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi [2].

2.3 Clustering

Clustering mengacu pada pengelompokan *record*, observasi, atau kasus ke dalam kelas dari objek serupa. Sebuah cluster adalah kumpulan *record* yang mirip satu sama lain, dan berbeda dengan *record* dari kelompok lain. Clustering berbeda dengan klasifikasi, yaitu tidak ada variabel target untuk clustering. Tugas *clustering* bukan mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel sasaran. Sebaliknya, algoritma *clustering* berusaha untuk mencari segmen seluruh data set menjadi subkelompok atau kelompok yang relatif homogen, di mana kesamaan *record* di dalam *cluster* dimaksimalkan dan kesamaan *record* di luar *cluster* diminimalkan [5]

2.4 Algoritma K-means

K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengkluster data besar dan data outlier dengan sangat cepat. Kelemahan metode ini memungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain [5]. Proses klustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1,\dots,n$; $j=1,\dots,m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1,\dots,k$; $j=1,\dots,m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean [5] seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m c_{ij} - c_{kj}^2} \quad (1)$$

Suatu data akan menjadi anggota dari cluster ke- k apabila jarak data tersebut ke pusat cluster ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap cluster.

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m c_{ij} - c_{kj}^2} \quad (2)$$

Nilai pusat cluster yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada cluster tersebut, dengan menggunakan rumus pada persamaan 3:

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

Dimana : x_{ij} = anggota kluster ke k
 P= banyaknya anggota kluster ke k

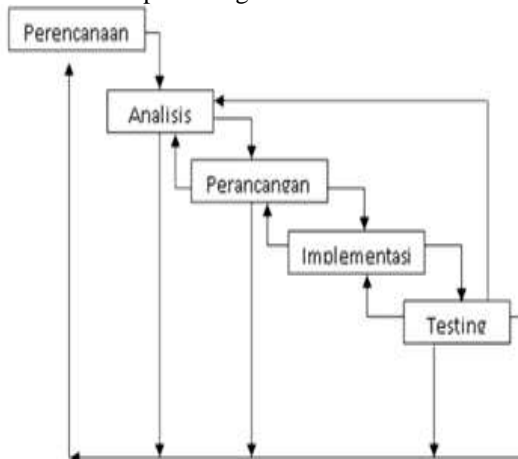
Algoritma dasar dalam k-means adalah

1. Tentukan jumlah kluster (k), tetapkan pusat kluster sembarang.
2. Hitung jarak setiap data ke pusat kluster menggunakan persamaan 1.
3. Kelompokkan data ke dalam kluster yang dengan jarak yang paling pendek menggunakan persamaan 2.
4. Hitung pusat kluster yang baru menggunakan persamaan 3.

Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

2.4 Metode Pengembangan Sistem

Perancangan sistem pada penelitian ini menggunakan metode *waterfall* (Sommerville) yang memiliki tahapan sebagai berikut :



Gambar 1. Metode Waterfall

III. Hasil dan Pembahasan

Penelitian ini menggunakan teknik kluster dengan K-Means untuk mengelompokkan data kategorikal hingga menghasilkan klaster. Selanjutnya dilakukan analisis terhadap hasil klasterisasi tersebut dalam bentuk pengelompokan lulusan untuk mendapatkan pengetahuan baru.

3.1 Pengumpulan Data

Data sampel awal yang digunakan dalam penelitian ini ialah data rekapitulasi nilai siswa DIKTUK BRIGADIR TA 2014 sebanyak 277 siswa dengan menggunakan parameter nilai intelek, nilai

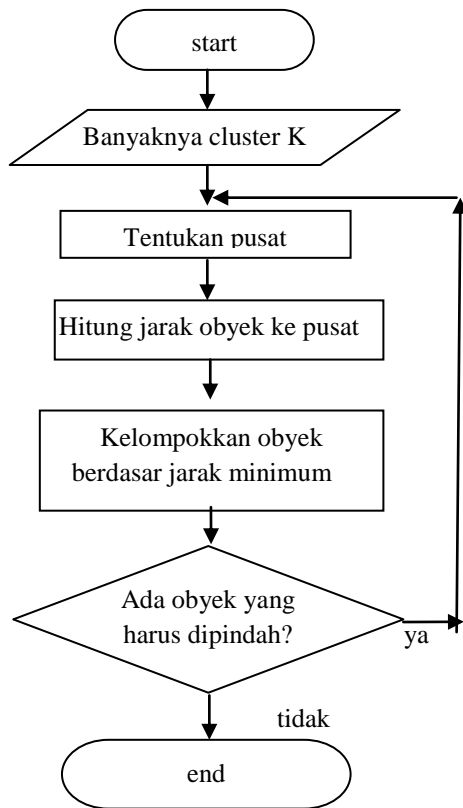
kepribadian, dan nilai kesehatan jasmani yang dapat dilihat pada Tabel .1

Tabel 1. Rekapitulasi nilai siswa DIKTUK BRIGADIR Polri

No siswa	Kepribadian	Intelek	Kes Jas
001	73,94	78,85	55,35
002	75,91	75,39	77,37
003	71,31	72,50	68,12
004	71,98	72,79	65,15
005	73,06	74,45	55,50
006	71,04	70,77	60,42
007	70,87	69,90	60,25
008	70,40	72,50	60,36
009	72,89	71,92	55,50
010	70,67	77,69	68,00
011	72,11	70,19	75,00
012	73,67	70,78	69,37
013	72,60	66,45	66,25
014	70,97	73,08	59,00
015	74,04	69,33	62,37
016	73,37	72,50	61,00
017	73,97	75,23	77,12
018	71,07	73,71	56,00
019	72,31	79,42	61,25
020	73,26	73,08	72,25
021	72,52	71,35	65,12
022	70,98	68,75	64,12
023	69,70	68,46	54,87
024	70,94	71,35	65,87
025	71,21	70,19	65,50
026	71,24	67,60	65,00
027	72,96	73,70	64,62
028	72,63	67,40	67,00
029	71,14	68,30	62,62
030	71,99	71,60	61,37
035	70,56	76,50	66,12
036	73,46	78,80	72,87
.....
.....
.....
274	71,86	71,30	63,75
275	73,59	77,90	60,00

276	74,48	72,30	63,50
277	73,91	65,90	58,00

3.2 Proses Klustering



Gambar 2. Flowchart algoritma K-Means

Selanjutnya diproses menggunakan algoritma K-Means sesuai dengan flowchart di atas.

Flowcart algoritma *K-means*:

- a. Penetapan jumlah *cluster* (K) pada penelitian ini yaitu berjumlah 3 *cluster*.

Tabel 2. Titik Pusat Awal Tiap Cluster

Titik Pusat Awal	nilai inteleg	nilai kepribadian	nilai kesehatan jasmani
cluster 1	73,94	78,85	55,35
cluster 2	75,12	66,20	61,37
cluster 3	73,91	65,90	58,00

- b. menghitung jarak setiap data ke pusat *cluster* antara objek ke *centroid* dengan perhitungan jarak *Euclidean* menggunakan persamaan 1.

$$d_{ik} = \sqrt{\sum_{j=1}^m c_{ij} - c_{kj}^2}$$

$$d_{11} = \sqrt{73,94 - 73,94 + 78,85 - 78,85 + 55,35 - 55,35}^2$$

$$d_{11} = 0$$

$$d_{12} = \sqrt{73,94 - 75,12 + 78,85 - 66,20 + 55,35 - 61,37}^2$$

$$d_{12} = 14,06$$

$$d_{13} = \sqrt{73,94 - 73,91 + 78,85 - 65,90 + 55,35 - 58,00}^2$$

$$d_{13} = 13,22$$

- c. Kelompokkan data ke dalam *cluster* dengan jarak minimal menggunakan persamaan 2.

Tabel 3. hasil perhitungan setiap data ke setiap cluster

NO Siswa	C1	C2	C3	JARAK TERDEKAT
1	0	14,06	13,22	C1
2	22,38	18,47	21,66	C2
3	14,5	9,99	12,36	C2
4	11,69	8,22	10,12	C2
5	4,49	10,33	8,95	C1
6	9,97	6,2	6,15	C3
7	10,66	5,75	5,51	C3
8	8,83	7,93	7,84	C3
9	7,01	8,49	6,6	C3
10	13,12	13,99	15,8	C1
11	21,55	14,52	17,62	C2
12	16,18	9,33	12,38	C2
1	0	14,06	13,22	C1
.....
.....
.....
272	15,16	8,86	11,81	C2
273	15,75	2,29	5,43	C2
274	11,48	6,5	8,15	C2
275	4,76	11,88	12,17	C1
276	10,47	6,49	8,46	C2
277	13,22	3,59	0	C3

- d. 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 menggunakan persamaan 3. Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga

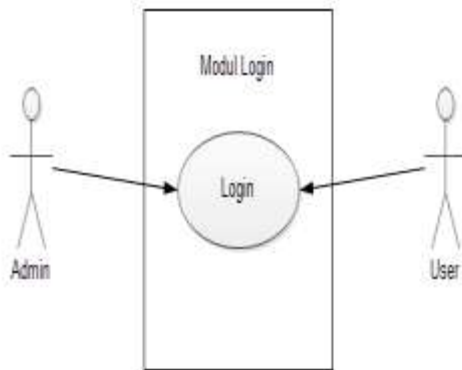
hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

a. Perancangan Sistem

1. Use Case Diagram

a. Login

Use Case Diagram ini menggambarkan tentang hal yang dapat dilakukan oleh user/admin pada proses *Login* seperti gambar 5.3.



Gambar 3 Use Case Diagram Login

Berikut *Use Case Description* yang menjelaskan penggambaran proses *Login*:

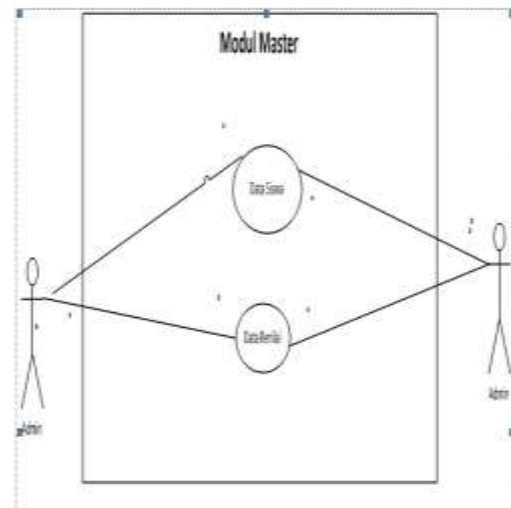
1. Login

Tabel 4 Use Case Description Login

Atribut	Keterangan
Name	Login
Short Description	Use case ini untuk proses login pengguna sistem
Actor	Admin, Pengguna
Pre-Condition	-
Basic Flow	Actor melalui proses login agar bisa masuk dan menggunakan sistem
Post-Condition	Actor berhasil masuk ke dalam sistem
Alternate Flow	-
Exception Flow	Actor gagal masuk ke dalam sistem dan kembali ke halaman login

b. Modul Master

Use Case Diagram ini menggambarkan tentang Modul Master seperti pada Gambar 5.4.



Gambar 4 Use Case Diagram Modul Master

Berikut *Use Case Description* yang menjelaskan penggambaran proses Modul Master:

1. Data siswa

Tabel 5. Use Case Description Siswa

Atribut	Keterangan
Name	Data siswa
Short Description	Use case ini untuk menyimpan, mengedit dan menghapus data siswa ke dalam sistem
Actor	Admin ataupun user
Pre-Condition	Halaman user ditampilkan.
Post-Condition	Actor berhasil menyimpan, mengedit dan menghapus data siswa

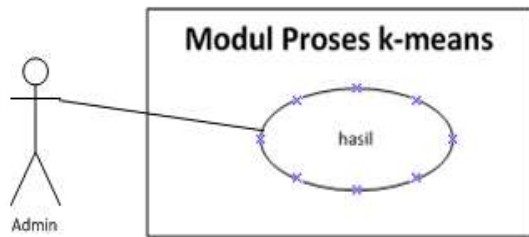
2. Data Renilai

Tabel 6. Use Case Description Data renilai

Atribut	Keterangan
Name	Data renilai
Short Description	Use case ini untuk menyimpan, mengedit dan menghapus data renilai ke dalam sistem
Actor	Admin
Pre-Condition	Halaman data pengguna ditampilkan.
Post-Condition	Actor berhasil menyimpan, mengedit dan menghapus data renilai

c. Modul proses k-means

Use case diagram ini menggambarkan tentang modul proses *k-means* seperti pada gambar 5.5.



Gambar 5 Use Case Diagram Modul proses k-means

Berikut use case description yang menjelaskan penggambaran proses modul proses k-means:

1. Data proses k-means

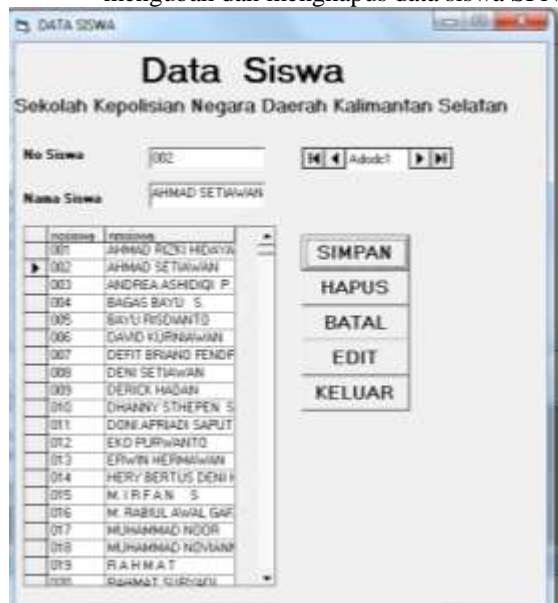
Tabel 7. Use Case Description Data proses k-means

Atribut	Keterangan
Name	Proses kluster
Short Description	Use case ini untuk menyimpan, mengedit dan menghapus data hasil ke dalam sistem
Actor	Admin, user
Pre-Condition	Halaman data hasil ditampilkan.
Post-Condition	Actor berhasil menyimpan, mengedit dan menghapus data hasil.

3.2 Implementasi Rancangan Aplikasi

Berikut akan ditampilkan implementasi dari perancangan sistem yang telah dibuat sebelumnya menggunakan use case diagram. Form Siswa

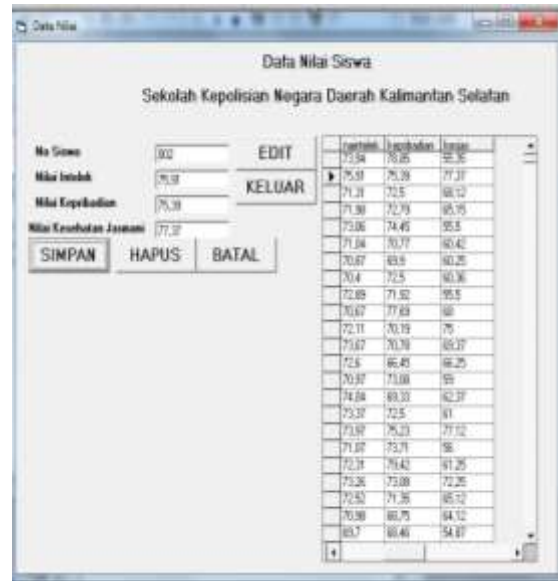
1. Form siswa merupakan halaman yang digunakan untuk melihat, menambah, mengubah dan menghapus data siswa SPN.



Gambar 6. Form Data siswa

2. Form nilai

Data nilai merupakan halaman yang digunakan untuk melihat, menambah, mengubah dan menghapus data nilai siswa SPN.



Gambar 7. form Data Nilai Siswa

3. Data Proses K-means

Tampilan proses K-Mean Clustering merupakan tampilan yang berguna untuk melakukan proses pengelompokan data rekapitulasi nilai akhir siswa SPN berdasarkan nilai kepribadian, intelek, dan kesehatan jasmani. Tampilan awal K-Mean Clustering dapat dilihat pada Gambar 8. Jumlah kalsifikasi dengan menginputkan jumlah pusat kluster awal sebanyak tiga dengan nilai yang tampilan berdasarkan random, kemudian user dapat menekan tombol proses yang menampilkan hasil proses k-means dan menyajikan informasi banyak anggota dari masing-masing klasifikasi dari 277 rekapitulasi nilai siswa yang ada.



Gambar 8. Form proses k-means

4. Form Pengujian K-Means Clustering Pada Iterasi Terakhir

Proses perhitungan *K-Means Clustering* berakhir ketika posisi *cluster* setiap siswa tidak mengalami perubahan posisi dengan posisi *cluster* pada proses iterasi sebelumnya, proses perhitungan telah selesai dilakukan pada iterasi ke 9.

Gambar 9. Tampilan Hasil Pengujian *K-Means Clustering* Iterasi Terakhir

5. Tampilan Hasil Laporan Akhir Pengujian Pada bagian ini akan menampilkan laporan akhir pengujian data yang sudah mengalami proses pengelompokan dengan metode *K-Means Clustering*. Tampilan hasil laporan akhir pengujian dapat dilihat pada gambar 10.

Hasil Klasi/Kan Kelulusan Siswa

No Siswa	Klasi 1	Klasi 2	Klasi 3	Kel/Hasil
001	10,301782010	10,20082202004	8,1177620100001	C1
002	10,500888004	7,2004200000000	24,53200004400	C2
003	5,805700000	3,5024200011400	14,702010014000	C2
004	2,970070100	5,0012704011000	11,00010000040	C1
005	7,712014000	10,5710000000000	4,0004010000001	C3
006	3,004010000	11,5400000000000	7,0000000000000	C1
007	3,540000000	11,50000014000	6,0040001000001	C1
008	3,007700100	10,0101000000000	7,0000000000000	C1
009	7,140000000	10,41000000000	3,7000010000000	C3
010	3,407100000	3,4000000000000	10,4000000000000	C2
011	12,400000000	5,0000174010001	21,3000000000000	C2
012	6,017040000	3,5471000000000	10,77000000000	C2
013	6,234000000	6,0040000000000	11,0000000000000	C1
014	4,410100000	12,0000000000000	6,0000000000001	C1
015	2,400000000	5,0000000000000	6,0000000000000	C1
016	1,000000000	6,0000000000000	7,0000000000000	C1
017	14,000000000	6,5170000000000	24,0000000000000	C2
018	7,200000000	14,00000000000	4,0000000000000	C3
019	6,047000000	11,1000000000000	12,0000000000000	C1
020	6,700000000	1,0000000000000	10,0000000000000	C2
021	2,520000000	6,2000000000000	11,5700000000000	C1

Gambar 10. Tampilan Laporan Akhir

5.4 Pembahasan

Pada proses k-means yang telah dilakukan, proses iterasi 19 dihentikan ketika hasilnya menunjukkan adanya setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain pada iterasi ke-18 dan ke-19. Dari hasil perhitungan didapatkan 3 kelompok *cluster* dan pengelompokan data berdasarkan kemiripan antar data dalam suatu kelompok dan meminimalkan kemiripan data pada kelompok lain. Tabel dibawah ini merupakan pusat kluster akhir yang didapat dari iterasi 19. Tabel dibawah ini menunjukkan banyaknya jumlah anggota pada setiap kluster pada iterasi 19.

Tabel 8. Pusat *cluster* akhir pada iterasi ke-19

Titik Pusat Awal	nilai intelek	nilai kepribadian	nilai kesehatan jasmani
cluster 1	72,93	71,52	62,63
cluster 2	73,33	73,79	70,80
cluster 3	72,92	69,95	53,64

Tabel 9. Pengelompokan Rekapitulasi Nilai

Cluster	Jumlah Anggota (Siswa)
1	124
2	98
2	55
	277

Kelompok 1 dengan nilai pusat kluster pada data nilai intelek 72,93, nilai kepribadian 71,52, dan nilai kesehatan jasmani 62,63 sebanyak 45%. Kelompok 2 dengan nilai pusat kluster pada data nilai intelek 73,33, nilai kepribadian 73,79, dan nilai kesehatan jasmani 70,80 sebanyak 35%. Kelompok 3 dengan nilai pusat kluster pada data nilai intelek 72,92, nilai kepribadian 69,95, dan nilai kesehatan jasmani 53,64 sebanyak 20%.

Dari hasil klasterisasi tersebut dilakukan analisis dalam bentuk klasifikasi lulusan berdasarkan kelompoknya. Kelompok 1 dengan rekapitulasi nilai lulusan sedang, Kelompok 2 dengan rekapitulasi nilai lulusan tinggi dan kelompok 3 dengan rekapitulasi nilai lulusan rendah.

IV. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan algoritma K-Means dapat dilakukan analisis dalam menentukan klasifikasi rekapitulasi nilai kelulusan siswa DIKTUK BRIGADIR TA 2014 dengan menggunakan parameter nilai intelek, nilai kepribadian, dan nilai kesehatan jasmani yang didapatkan 3 kelompok yaitu: kelompok rekapitulasi nilai lulusan sedang sebanyak 45%, kelompok rekapitulasi nilai lulusan tinggi sebanyak 35% dan

kelompok dengan rekapitulasi nilai lulusan rendah sebanyak 20%.

Daftar Pustaka:

- [1] Ari Muzakir, *Analisa dan Pemanfaatan Algoritma K-Means Clustering pada Data Nilai Siswa Sebagai Penentu Penerimaan Beasiswa*, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 ISSN: 1979-911X Yogyakarta, 15 November 2014.
- [2] Davies and Paul Beynon, 2004, *Database System Third Edition*. Palgrave Macmillan : New York
- [3] Dian Eka Ratnawati, dkk, Pengembangan Metode Klasifikasi berdasarkan K-Means dan LVQ, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* Vol. 1, No. 1, 1-4, 2014.
- [4] Han, Jiawei; & Kamber, Micheline. 2001. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kauffman.
- [5] Larose, D. T . 2005. *Discovering Knowledge in Data : An Introduction to Data mining*. Wiley-Interscience A John Wiley & Sons, Inc Publication.
- [6] Pramudiono, I, 2003, Pengantar *Data Mining*. Diakses 10 Mei 2016, dari <http://ikc.depsos.go.id/umum/ikodatamining.php>.
- [7] Santosa, B, 2007, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu : Yogyakarta.
- [8] Sommerville, I. 2001. *Software Engineering*. Addison Wesley.
- [9] Totok Suprawoto, 2016, *Klasifikasi Data Mahasiswa Menggunakan Metode K-Means untuk Menunjang Pemilihan Strategi Pemasaran*. *Jurnal Informatika dan Komputer (JIKO)*, Vol. 1, No. 1, 12-18, 2016