

PROYEK TUGAS AKHIR
IMPLEMENTASI METODE *K-MEANS* PADA PENERIMAAN SISWA
BARU

(Studi Kasus : SMK Pembangunan Nasional Purwodadi)



Disusun oleh:
Novian Hari Pratama
5130411335

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN ELEKTRO
UNIVERSITAS TEKNOLOGI YOGYAKARTA
YOGYAKARTA
2018

Naskah Publikasi

**IMPLEMENTASI METODE *K-MEANS* PADA PENERIMAAN SISWA
BARU**
(Studi Kasus : SMK Pembangunan Nasional Purwodadi)

Disusun oleh:
Novian Hari Pratama
5130411335

Telah disetujui oleh pembimbing

Pembimbing

Suhirman,Ph.D.

tanggal:.....

IMPLEMENTASI METODE K-MEANS PADA PENERIMAAN SISWA BARU (Studi Kasus : SMK Pembangunan Nasional Purwodadi)

Novian Hari Pratama

*Program Studi Informatika, Fakultas Teknologi Informasi dan Elektro
Universitas Teknologi Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
Email: novianpratama26@gmail.com*

ABSTRAK

Saat ini, belum banyak sekolah di Indonesia menyelenggarakan Penerimaan Siswa Baru berbasis komputer. Penerimaan siswa juga masih belum menggunakan perhitungan khusus agar calon siswa tidak salah masuk jurusan. Kmeans merupakan salah satu model matematis yang digunakan untuk mengelompokkan data yang awalnya acak sehingga data menjadi terstruktur. K-Means dalam pengelompokan data sangat efisien dan tidak membutuhkan waktu yang lama dan akurasi kebenaran dari metode K-Means dapat dikatakan di atas standar. Maka dari itu, peneliti membuat aplikasi yang dapat mengklasifikasikan penerimaan siswa berdasarkan nilai (Grade) yang telah ditentukan. Adapun hasil dari penelitian yang dibuat ini berupa saran, jika siswa memilih jurusan 'A' sedangkan nilainya hanya termasuk di jurusan 'B', maka sistem akan memberikan siswa saran untuk masuk di Jurusan 'B'. Namun jika sebaliknya, siswa dapat memilih salah satu dari jurusan dipilih dan jurusan yang diterima.

Kata kunci: *K-Means*, Penerimaan siswa baru, Sistem

1. PENDAHULUAN

Saat ini belum banyak sekolah di Indonesia menyelenggarakan Penerimaan Siswa Baru berbasis komputer. Dengan manfaat dan kemudahan yang ada, sudah seharusnya sistem ini dikembangkan oleh tiap sekolah. Hal ini sejalan dengan kemajuan teknologi informasi dan komunikasi seperti teknologi yang mampu mendukung proses input dan output data secara cepat dan akurat, khususnya dalam pelaksanaan ujian penerimaan siswa baru. Nugroho (2011) membahas tentang Sistem Penerimaan Siswa Baru Berbasis Web dengan PHP dan SQL bahwa sistem mampu mengelola pelaksanaan Penerimaan Siswa Baru yaitu pendaftaran, seleksi, penjurjanaan, pengumuman, dan pendaftaran ulang.

Berdasarkan dari pemaparan di atas, karena sistem penerimaan siswa di sekolah belum mengimplementasikan metode untuk kelulusan dan tidak lulusnya siswa pada saat ujian, maka peneliti akan membuat IMPLEMENTASI METODE K-MEANS PADA SISTEM PENERIMAAN SISWA PADA SMK PEMBANGUNAN NASIONAL PURWODADI.

2. LANDASAN TEORI

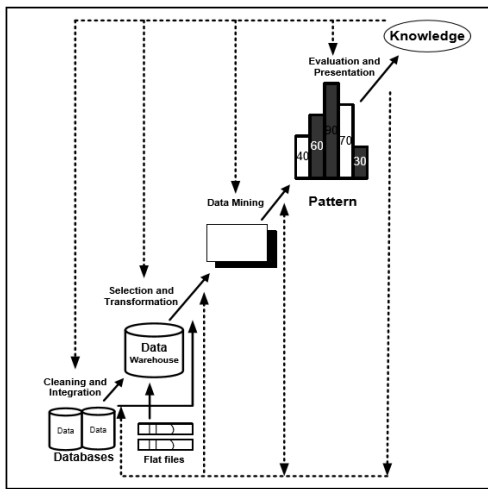
2.1. Data Mining

Data Mining merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, association rule, dan visualisasi (Han dan Kamber, 2006).

Secara garis besar, data mining terdiri dari dua kategori utama, yaitu:

- a. Deskripsi Mining, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk deskripsi mining adalah clustering, association, dan sequential mining.
- b. Prediksi Mining, yaitu proses untuk menemukan pola dari data dengan menggunakan variabel lain dimasa depan. Salah satu teknik yang terdapat dalam prediksi mining adalah klasifikasi.

Secara sederhana data mining biasa dikatakan sebagai proses menyaring atau menambang pengetahuan dari sejumlah data yang sangat besar. Istilah lain untuk data mining adalah Knowledge Discovery in Database atau KDD. Walaupun sebenarnya data mining sendiri adalah bagian dari tahapan proses dalam KDD seperti yang terlihat pada Gambar 1 (Han, 2006).

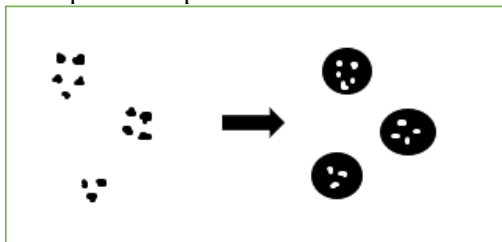


Gambar 1 Proses data mining (Han, 2006)

2.2. Definisi Clustering

Clustering merupakan salah satu bagian dari teknik data mining yaitu sekumpulan objek yang mempunyai “kesamaan” di antara anggotanya dan memiliki “ketidaksamaan” dengan objek lain pada cluster lainnya, dengan kata lain sebuah cluster adalah sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya.

Clustering adalah proses membuat pengelompokan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matriks tertentu. Selain dengan menggunakan similaritas (kesamaan) berdasarkan bentuk dan warna, clustering juga bisa dilakukan dengan menggunakan similaritas berdasarkan jarak, artinya data yang memiliki jarak berdekatan akan membentuk satu cluster, contohnya seperti dapat dilihat pada Gambar 2.



Gambar 2 Clustering berdasarkan jarak

2.3. Algoritma K-Means

Algoritma k-means adalah metode clustering non-hierarchical berbasis jarak yang membagi data ke dalam cluster dan algoritma ini bekerja pada atribut numerik. Algoritma k-means termasuk dalam partitioning clustering yang memisahkan data ke daerah bagian yang terpisah. Algoritma k-means sangat terkenal karena kemudahannya dan

kemampuannya untuk meng-cluster data besar dan outlier dengan sangat cepat.

K-Means merupakan metode clustering yang sangat terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk meng-cluster data yang besar, mampu menangani data outlier dan kompleksitas waktunya linear $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah cluster, dan T adalah jumlah iterasi. Dalam Algoritma k-means, setiap data harus termasuk dapat berpindah ke cluster yang lain. Pada dasarnya penggunaan Algoritma k-means dalam melakukan proses clustering tergantung dari data yang ada dan konklusi yang ingin dicapai. Untuk itu digunakan K-Means Algorithm yang di dalamnya memuat aturan sebagai berikut:

- a. Jumlah input dari cluster
- b. Memiliki atribut yang bersifat numerik

Algoritma k-means pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat cluster awal. Pada step ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya k-means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat cluster. Posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap cluster dan terakhir akan terbentuk posisi cluster baru.

2.4. Proses K-Means

Algoritma k-means pada dasarnya melakukan 3 proses yaitu, proses pendeteksian lokasi pusat cluster, proses pencarian anggota dari tiap-tiap cluster dan proses pencarian anggota dari tiap-tiap cluster. Proses Algoritma k-means sebagai berikut:

a. Penentuan pusat cluster awal

Dalam menentukan n buah pusat cluster awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data nilai max dan nilai min bukan dengan menentukan titik baru secara dengan random pusat awal dari data.

b. Perhitungan jarak dengan pusat cluster

Untuk mengukur jarak antar data dengan pusat cluster digunakan algoritma euclidian distance, algoritma perhitungan jarak data dengan pusat cluster:

1. Ambil nilai data dan nilai pusat cluster.

- 2 Hitung *city block distance* dengan data tiap pusat cluster menggunakan persamaan (0.1).

$$d_{ij} = \sqrt{\sum_{k=1}^n |x_{ik} - x_{jk}|} \quad (0.1)$$

Di mana x_{ik} adalah pusat cluster baru, x_{jk} pada kasus ini adalah nilai dari siswa d_{ij} merupakan jarak, i merupakan pusat cluster di mana $i = 1, 2, \dots, n$, j merupakan data nilai dari setiap siswa, di mana $j = 1, 2, \dots, n$. n merupakan jumlah sampel.

c. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat.

Algoritma pengelompokan data:

- 1 Ambil nilai jarak tiap pusat cluster dengan data.
- 2 Cari nilai jarak terkecil.
- 3 Kelompokkan data dengan pusat cluster yang memiliki jarak yang terkecil.

d. Penentuan pusat cluster baru

Untuk mendapatkan pusat cluster baru bisa dihitung dari rata-rata nilai anggota cluster dan pusat cluster. Pusat cluster yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah konvergen (pusat cluster baru sama dengan pusat cluster lama).

Algoritma penentuan pusat cluster:

- 1 Cari jumlah anggota tiap cluster.
- 2 Hitung pusat cluster baru menggunakan persamaan (0.2) :

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (0.2)$$

Di mana $\frac{1}{M}$ merupakan jumlah data dari setiap cluster, C_i merupakan fitur ke- i dalam sebuah cluster, dan X merupakan data dari setiap cluster.

e. Ulangi langkah ke tiga, jika cluster masih berubah

Karakteristik algoritma k-means berdasarkan cara kerjanya sebagai berikut:

- a. K-Means sangat cepat dalam proses clustering.
- b. K-Means sangat sensitif pada proses pembangkitan centroid awal secara random.

- c. Memungkinkan suatu cluster tidak mempunyai anggota.
- d. Hasil clustering dengan k-means bersifat tidak unik (selalu berubah), terkadang baik, dan terkadang buruk.

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. Ada dua cara pengalokasian data kembali ke dalam masing-masing cluster pada saat proses iterasi clustering. K-Means dalam pengalokasian data ke dalam masing-masing cluster dapat dilakukan dengan dua cara yaitu hard k-means dan fuzzy k-means. Perbedaan dari kedua metode tersebut terletak pada asumsi yang dipakai sebagai dasar dari pengalokasian data. Hard, dalam artian suatu data secara tegas atau pasti dinyatakan sebagai anggota satu cluster tertentu dan tidak menjadi anggota cluster yang lain. Sedangkan fuzzy diartikan masing-masing data mempunyai nilai kemungkinan untuk dapat bergabung ke setiap cluster yang ada.

3. METODOLOGI PENELITIAN

3.1. Obyek Penelitian

Obyek dari penelitian ini adalah penerimaan siswa baru. Data lama penerimaan siswa baru di jadikan sumber untuk mengelempokan siswa yang lulus dan tidak lulus menggunakan metode *K-Means* karena saat ini, belum banyak sekolah yang menggunakan metode untuk penerimaan siswa baru dan prinsip kerja metode ini adalah mengambil nilai ujian masuk yang paling tinggi jika calon siswa tidak ada yang memenuhi target yang di tentukan dari sekolah untuk standar nilai kelulusan.

3.2. Tahapan Penelitian

a. Pengumpulan Data

Pertama, peneliti melakukan pengumpulan data. Adapun pengumpulan data yang penulis gunakan untuk menyelesaikan penelitian ini adalah:

1. Wawancara
Pengumpulan data dengan menggunakan metode wawancara langsung dengan petugas Tata Usaha yaitu bapak Muh Dul Wahid
2. Observasi

Pada Observasi ini, penulis mengumpulkan data yang berhubungan dengan sistem yang akan dibuat mulai dari pengecekan nilai siswa dan standar kelulusan.

3. Literatur

Penulis melakukan literatur penelitian dengan membaca dan mempelajari buku-buku, skripsi serta artikel yang mendukung dengan topik yang dibahas dalam penyusunan penelitian kerja praktik ini. Penulis juga mengumpulkan data-data dari situs internet yang berhubungan dengan penelitian kerja praktik penulis.

3.3. Analisis Sistem

Sistem yang akan dibangun pada penelitian ini adalah sistem penerimaan siswa dengan menggunakan metode *K-Means* sebagai acuan untuk penerimaan siswa dari SMK Pembangunan Nasional Purwodadi. Adapun tahapan dari sistem ini adalah antara lain:

- Menginputkan nilai siswa sebanyak-banyaknya untuk digunakan sebagai proses pelatihan metode *K-Means* terlebih dahulu dengan menggunakan $k = 7$, yaitu lulus dan tidak lulus.
- Ujian penerimaan siswa dilakukan dengan cara berbasis komputer.
- Hasil dari ujian penerimaan siswa akan diuji dengan menggunakan bobot yang telah didapatkan pada saat proses pelatihan.
- Hasil dari proses pengujian berupa diterima di jurusan tertentu, misalnya pada jurusan TKR, TSM, TITL, TKJ, TPMP, TPBO, dan TAV.

Untuk input nilai siswa untuk dijadikan data pelatihan, menggunakan nilai ujian tahun kemarin atau satu tahun sebelumnya. Data ini nantinya akan dijadikan acuan untuk proses penerimaan siswa pada SMK Pembangunan Nasional Purwodadi. Jumlah *cluster* dari metode *K-Means* adalah 2 *cluster* yaitu lulus dan tidak lulus. Proses pelatihan digunakan untuk mencari bobot yang tepat untuk dijadikan acuan pada proses pengujian nantinya.

Kedua, yaitu ujian berbasis komputer. Di mana sistem yang akan dibuat telah menyediakan ujian untuk pengambilan nilai dari setiap siswa. Hasil dari ujian penerimaan ini nantinya akan dilakukan proses pengujian untuk menentukan siswa yang berhak lulus dan siswa yang tidak lulus. Adapun algoritma dari metode *K-Means* adalah sebagai berikut:

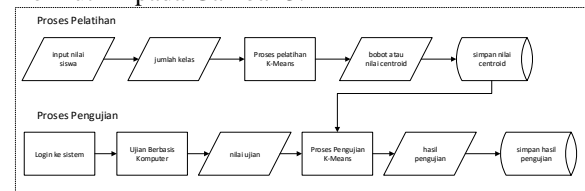
- Penentuan pusat cluster awal, cluster awal merupakan nilai centroid yang didapatkan secara acak dengan syarat jumlah data sebagai jumlah cluster dan jumlah kelas sebagai jumlah cluster.
- Menghitung jarak dengan pusat cluster menggunakan persamaan (0.1).

- Pengelompokan data, lalu mengambil nilai jarak yang paling terkecil.
- Penentuan pusat cluster baru dengan menggunakan persamaan (0.2).
- Ulangi langkah 2 sampai pusat *cluster* tidak berpindah.

3.4. Perancangan Sistem

Perancangan alur proses sistem dimulai dari tahap awal penginputan nilai siswa untuk dijadikan data pelatihan kemudian menentukan jumlah kelas. Jumlah kelas pada kasus ini ada dua yaitu, lulus dan tidak lulus. Setelah itu dilakukan proses pelatihan untuk mendapatkan bobot dari metode *k-means*.

Untuk proses pengujian sistem siswa terlebih dahulu melaksanakan ujian berbasis komputer. Sebelum ujian siswa terlebih dahulu login ke sistem dengan username menggunakan nis dan password menggunakan tanggal lahir. Setelah siswa melakukan proses ujian, pengumuman akan otomatis keluar dari sistem. Berikut ini pada Gambar 3.



Gambar 3 Alur perancangan sistem

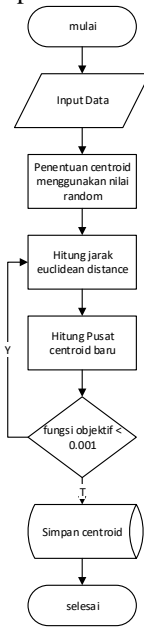
3.5. Perancangan metode *K-Means* untuk proses pelatihan sistem

Algoritma dari metode *k-means* terdiri dari beberapa proses. Proses tersebut adalah sebagai berikut:

- Penentuan jumlah kelompok atau jumlah kelas.
- Mengalokasikan data kelompok secara acak atau menentukan pusat *cluster* awal menggunakan nilai random.
- Menghitung jarak menggunakan persamaan (0.1).
- Hitung pusat cluster baru dengan menggunakan persamaan (0.2).
- Alokasikan kembali data ke centroid terdekat.
- Ulangi langkah C,
 - Apabila masih ada data yang berpindah cluster.
 - Apabila perubahan nilai centroid ada yang di atas nilai threshold yang ditentukan.
 - Apabila perubahan nilai pada fungsi obyektif yang digunakan masih di atas nilai threshold yang ditentukan.

Jika nilai centroid akhir telah ditemukan, nilai centroid tersebut akan disimpan ke dalam database dan akan digunakan sebagai proses pengujian sistem. Adapun

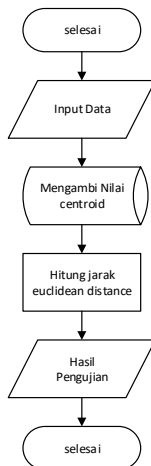
alur proses dari metode *k-means* untuk proses pelatihan dapat dilihat pada Gambar 4



Gambar 4 Alur proses peatihan *k-means*

3.6. Perancangan metode *K-Means* untuk proses pengujian

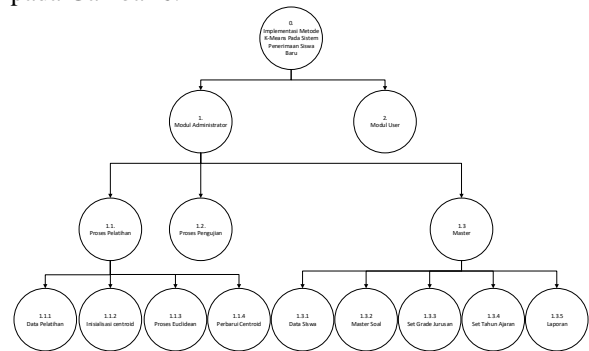
Nilai Centroid untuk data pengujian metode *k-means* dengan menggunakan nilai centroid yang telah disimpan ke dalam database pada sub-sub bab 4.2.1 di atas. Alur dari proses pengujian metode *k-means* tidak jauh beda dengan proses pelatihnnya, bedanya hanya pada di nilai centroid. Pada proses ini dilai centroid diambil langsung dari database dengan menggunakan nilai centroid yang telah ditemukan pada proses pelatihan sebelumnya. Alur dari proses perancangan metode *k-means* untuk proses pengujian dapat dilihat pada Gambar 5.



Gambar 5 Alur proses pengujian *k-means*

3.7. Diagram Jenjang

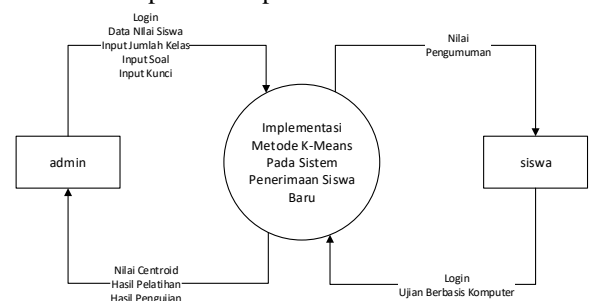
Perancangan diagram berjenjang untuk sistem yang akan dibuat terdiri dari beberapa proses yaitu dimulai dari level 0 yaitu proses pada diagram konteks. Level 1 menunjukkan proses DFD pada level 1 yaitu modul administrator, dan modul user. Modul administratif kemudian dipecah menjadi beberapa proses, yaitu proses pelatihan, proses pengujian dan data master. Proses pelatihan menghasilkan nilai centroid, sedangkan pada proses pengujian sistem menggunakan nilai centroid yang telah ditemukan pada saat proses pelatihan. Kemudian data master yang berfungsi untuk menginputkan data siswa, data jurusan, dan data soal. Adapun perancangan diagram berjenjang dapat dilihat pada Gambar 6.



Gambar 6 Diagram jenjang

3.8. Diagram Konteks

Diagram konteks atau biasa juga disebut diagram alir data level 0 merupakan gambaran umum dari sistem yang akan dibuat. Diagram konteks pada kasus ini terdiri dari dua *external entity* yaitu, admin dan user. Admin bertugas untuk melakukan proses pelatihan ataupun proses pengujian pada sistem, sedangkan user yang dimaksud pada sistem ini adalah siswa. Untuk masuk ke sistem, siswa terlebih dahulu melakukan proses login, kemudian mengerjakan soal ujian yang telah disediakan. Gambaran diagram konteks pada sistem ini dapat dilihat pada Gambar 7.

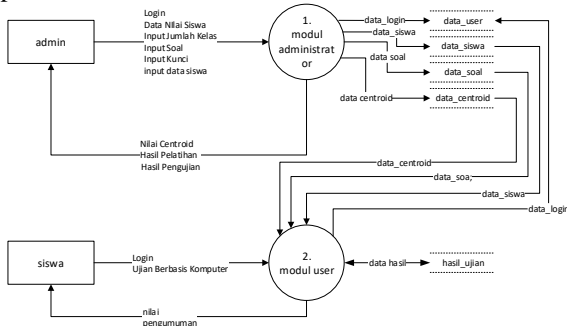


Gambar 7 Diagram konteks

3.9. Diagram alir data level 1

Diagram alir data (DAD) level 1 pada sistem ini terdapat dua proses yaitu proses masuk ke sistem sebagai administrator atau user. Adapun arus data yang mengalir ke sistem pada admin yaitu login, input nilai siswa untuk proses pelatihan input jumlah kelas, input soal dan input kunci jawaban dari soal, sedangkan arus data yang mengalir dari sistem ke admin yaitu nilai centroid, hasil pelatihan dan hasil pengujian dari sistem. Nilai centroid ditemukan ketika proses pelatihan telah dilakukan.

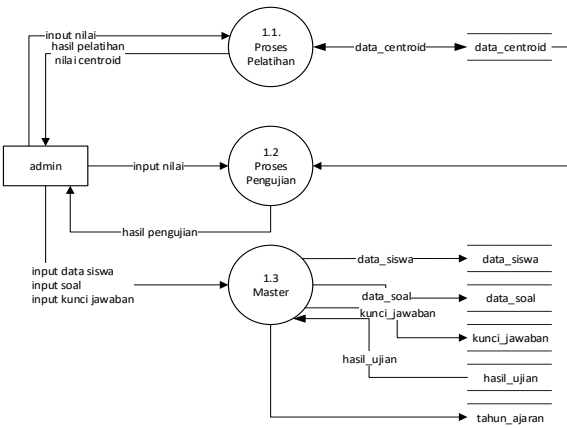
Arus data dari siswa ke modul user yaitu, login dan ujian berbasis komputer, sedangkan arus data dari modul user ke siswa yaitu berupa output nilai dan hasil pengumuman. Gambaran DAD level 2 dapat dilihat pada Gambar 8.



Gambar 8 DAD level 1

3.10. Diagram alir data level 2 proses 1

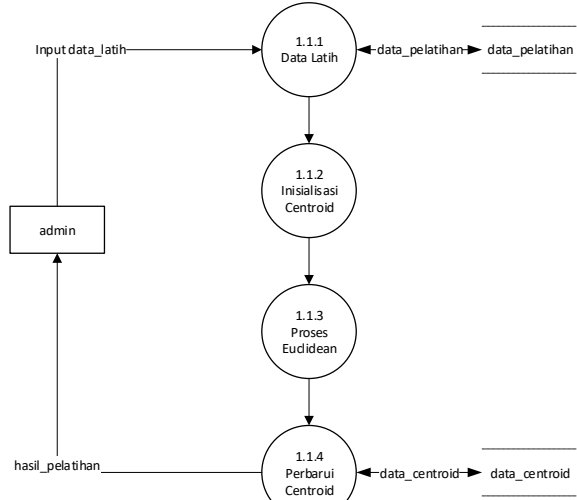
Diagram alir data level 2 proses 1 merupakan pecahan dari DAD level 1 pada proses 1. Diagram alir data level 2 proses 1 terdapat tiga proses yaitu, proses pelatihan, proses pengujian dan proses penginputan data master. Perancangan DAD level 2 proses 1 dapat dilihat pada Gambar 9.



Gambar 9 DAD level 2 proses 1

3.11. Diagram alir data level 3 proses 1.1

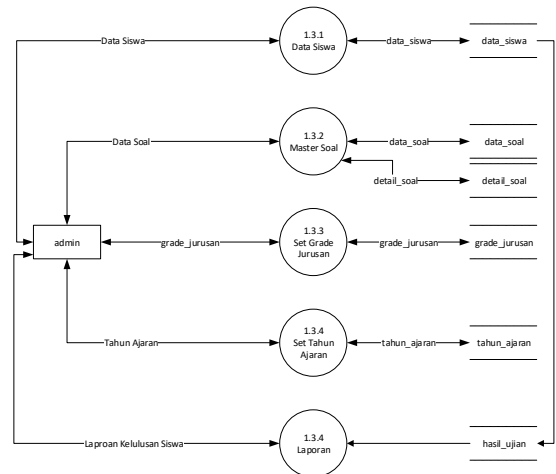
DAD level 3 proses 1.1 merupakan dahapan dari proses pelatihan dengan menggunakan metode *k-means*. Pada Gambar 10 terdapat empat proses yakni, data latih, inialisasi nilai centroid, proses perhitungan euclidean dengan menggunakan persamaan (0.1), dan proses memperbarui nilai centroid dengan menggunakan persamaan (0.2).



Gambar 10 DAD level 3 proses 1.1

3.12. Diagram alir data proses 3 level 1.3

DAD level 3 proses 1.3 merupakan proses-proses yang berhubungan dengan penginputan data-data penting seperti data siswa dll. Pada Gambar 11 menunjukkan bahwa terdapat lima proses yaitu, proses memasukkan data siswa, data untuk soal ujian, *setting* grade jurusan, *setting* tahun ajaran, dan print out hasil seleksi siswa baik yang lulus maupun yang tidak lulus.



Gambar 11 DAD level 3 proses 1.3

4. HASIL DAN PEMBAHASAN

Proses implementasi dari perancangan aplikasi yang dilakukan pada bab sebelumnya akan dijelaskan pada bab ini. Implementasi bertujuan untuk menerjemahkan keperluan perangkat lunak ke dalam bentuk sebenarnya yang dimengerti oleh komputer atau dengan kata lain tahap implementasi ini merupakan tahapan lanjutan dari tahap perancangan yang sudah dilakukan. Dalam tahap implementasi ini akan dijelaskan mengenai perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan dalam membangun sistem ini, *file-file* yang digunakan dalam membangun sistem, tampilan halaman *form* beserta potongan-potongan *script* program.

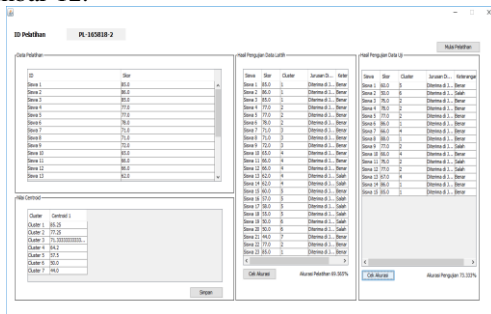
Sistem penerimaan yang akan di buat menggunakan *Algoritma K-Means* dengan menggunakan *7 cluster* sebagai target. Target adalah macam – macam *output* yang akan di hasilkan oleh komputer berupa saran dan informasi ketidaklulusan.

4.1. Implementasi halaman proses pelatihan data simulasi

Nilai yang telah diinputkan sebagai data pelatihan akan di gunakan untuk proses pelatihan *K-Means*. Pada proses pelatihan inilah akan di dapatkan nilai centroid yang akan digunakan sebagai acuan pada saat pengujian siswa. Proses data pelatihan ada 3 proses yaitu:

Proses simulasi data

Proses simulasi data merupakan proses pelatihan untuk data simulasi atau untuk mencari centroid dengan menggunakan data acak yang diambil dari rekap nilai siswa. Proses pelatihan ini digunakan untuk membuktikan proses apakah proses perhitungan manual dengan yang ada diprogram sudah sesuai. Implementasi halaman simulasi bisa dilihat pada Gambar 12.



Gambar 12 Implementasi halaman proses pelatihan data simulasi

Iterasi I

Data	X	C1	C2	C3	C4	C5	C6	C7	Hasil cluster
1	85	5	10	15	20	27	33	45	C1
2	86	6	11	16	21	28	34	46	C1
3	85	5	10	15	20	27	33	45	C1
4	77	3	2	7	12	19	25	37	C2
5	77	3	2	7	12	19	25	37	C2
6	78	2	3	8	13	20	26	38	C1
7	71	9	4	1	6	13	19	31	C3
8	71	9	4	1	6	13	19	31	C3
9	72	8	3	2	7	14	20	32	C3
10	65	15	10	5	0	7	13	25	C4
11	66	14	9	4	1	8	14	26	C4
12	66	14	9	4	1	8	14	26	C4
13	62	18	13	8	3	4	10	22	C4
14	62	18	13	8	3	4	10	22	C4
15	60	20	15	10	5	2	8	20	C5
16	57	23	18	13	8	1	5	17	C5
17	58	22	17	12	7	0	6	18	C5
18	55	25	20	15	10	3	3	15	C5
19	50	30	25	20	15	8	2	10	C6
20	50	30	25	20	15	8	2	10	C6
21	44	36	31	26	21	14	8	4	C7

μ1	80
μ2	75
μ3	70
μ4	65
μ5	58
μ6	52
μ7	40

Cluster 1 = {85,86,85,78}	83,5
Cluster 2 = {77,77,}	77
Cluster 3 = {71,71,72}	71,33 333
Cluster 4 = {65,66,66,62,62}	64,2
Cluster 5 = {60,57,58,55}	57,5
Cluster 6 = {50,50}	50
Cluster 7 = {44}	44

Iterasi II

data	X	C1	C2	C3	C4	C5	C6	C7	Hasil cluster
1	85	1,5	8	13,667	20,8	27,5	35	41	C1
2	86	2,5	9	14,667	21,8	28,5	36	42	C1
3	85	1,5	8	13,667	20,8	27,5	35	41	C1
4	77	6,5	0	5,667	12,8	19,5	27	33	C2
5	77	6,5	0	5,667	12,8	19,5	27	33	C2
6	78	5,5	1	6,667	13,8	20,5	28	34	C2
7	71	12,5	6	0,333	6,8	13,5	21	27	C3
8	71	12,5	6	0,333	6,8	13,5	21	27	C3
9	72	11,5	5	0,667	7,8	14,5	22	28	C3
10	65	18,5	12	6,333	0,8	7,5	15	21	C4
11	66	17,5	11	5,333	1,8	8,5	16	22	C4
12	66	17,5	11	5,333	1,8	8,5	16	22	C4
13	62	21,5	15	9,333	2,2	4,5	12	18	C4
14	62	21,5	15	9,333	2,2	4,5	12	18	C4
15	60	23,5	17	11,333	4,2	2,5	10	16	C5
16	57	26,5	20	14,333	7,2	0,5	7	13	C5
17	58	25,5	19	13,333	6,2	0,5	8	14	C5
18	55	28,5	22	16,333	9,2	2,5	5	11	C5
19	50	33,5	27	21,333	14,2	7,5	0	6	C6
20	50	33,5	27	21,333	14,2	7,5	0	6	C6
21	44	39,5	33	27,333	20,2	13,5	6	0	C7

μ1	83,5
μ2	77
μ3	71,333
μ4	64,2
μ5	57,5
μ6	50
μ7	44

Cluster 1 = {85,86,85}	85,33333
Cluster 2 = {77,77,78}	77,33333
Cluster 3 = {71,71,72}	71,33333
Cluster 4 = {65,66,66,62,62}	64,2
Cluster 5 = {85,86,85}	57,5
Cluster 6 = {85,86,85}	50
Cluster 7 = {85,86,85}	44

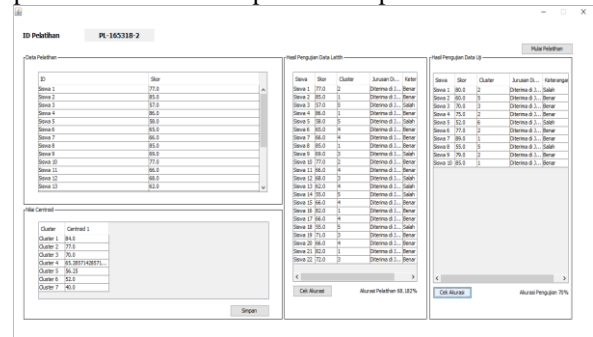
Iterasi III

data	X	C1	C2	C3	C4	C5	C6	C7	Hasil cluster
1	85	0.333	7.667	13.667	20.8	27.5	35	41	C1
2	86	0.667	8.667	14.667	21.8	28.5	36	42	C1
3	85	0.333	7.667	13.667	20.8	27.5	35	41	C1
4	77	8.333	0.333	5.667	12.8	19.5	27	33	C2
5	77	8.333	0.333	5.667	12.8	19.5	27	33	C2
6	78	7.333	0.667	6.667	13.8	20.5	28	34	C2
7	71	14.333	6.333	0.333	6.8	13.5	21	27	C3
8	71	14.333	6.333	0.333	6.8	13.5	21	27	C3
9	72	13.333	5.333	0.677	7.8	14.5	22	28	C3
10	65	20.333	12.333	6.333	0.8	7.5	15	21	C4
11	66	19.333	11.333	5.333	1.8	8.5	16	22	C4
12	66	19.333	11.333	5.333	1.8	8.5	16	22	C4
13	62	23.333	15.333	9.333	2.2	4.5	12	18	C4
14	62	23.333	15.333	9.333	2.2	4.5	12	18	C4
15	60	25.333	17.333	11.333	4.2	2.5	10	16	C5
16	57	28.333	20.333	14.333	7.2	0.5	7	13	C5
17	58	27.333	19.333	13.333	6.2	0.5	8	14	C5
18	55	30.333	22.333	16.333	9.2	2.5	5	11	C5
19	50	35.333	27.333	21.333	14.2	7.5	0	6	C6
20	50	35.333	27.333	21.333	14.2	7.5	0	6	C6
21	44	41.333	33.333	27.333	20.2	13.5	6	0	C7

$$= \frac{15}{21} \times 100\% = 71.429\%$$

4.2. Implementasi Data Pelatihan sebanyak 22 data

Data latih kecil merupakan Proses pelatihan data merupakan proses pelatihan untuk data kecil atau untuk mencari centroid dengan menggunakan data yang berurutan sebanyak 22 data dan untuk data uji sebanyak 10 data yang diambil dari rekam nilai siswa dari tahun sebelumnya. Adapun implementasi proses pelatihan data kecil dapat dilihat pada Gambar 13



Gambar 13 Implementasi data latih kecil

Pada Gambar 13 terdapat 22 data latih dengan 15 data yang benar dan sisanya data tidak sesuai dengan target. Jadi, untuk menghitung akurasi dari proses data latih kecil sebagai berikut:

$$akurasi = \frac{15}{22} \times 100\% = 68.182\%$$

Adapun data benar dari proses pelatihan data kecil yaitu, data ke 1, 2, 4, 6, 7, 8, 10, 11, 15, 16, 17, 19, 20, 21, 22.

Untuk akurasi data uji pada proses pelatihan data kecil menggunakan 10 pengujian dengan 7 jumlah data benar dari 10 data uji. Jadi, untuk menghitung akurasi dari proses data uji sebagai berikut

$$akurasi = \frac{7}{10} \times 100\% = 70\%$$

Adapun data benar dari proses pelatihan data kecil yaitu, data ke 2,3,4,6,7,9,10

4.3. Implementasi Data Pelatihan sebanyak 34 data

Data latih kecil merupakan Proses pelatihan data merupakan proses pelatihan untuk data besar atau untuk mencari centroid dengan menggunakan data yang berurutan sebanyak 34 data dan untuk data uji

μ1	83.33333
μ2	77.33333
μ3	71.33333
μ4	64.2
μ5	57.5
μ6	50
μ7	44

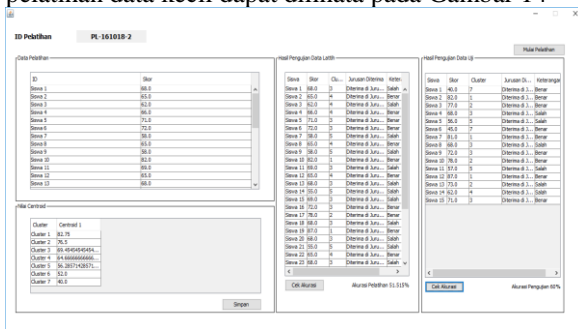
Cluster 1 = {85,86,85}	85,33333
Cluster 2 = {77,77,78}	77,33333
Cluster 3 = {71,71,72}	71,33333
Cluster 4 = {65,66,66,62,62}	64,2
Cluster 5 = {85,86,85}	57,5
Cluster 6 = {85,86,85}	50
Cluster 7 = {85,86,85}	44

Hasil kesimpulan

Data	X	Hasil seharusnya	Hasil Cluster	Keterangan
1	85	C1	C1	Benar
2	86	C1	C1	Benar
3	85	C1	C1	Benar
4	77	C2	C2	Benar
5	77	C2	C2	Benar
6	78	C2	C2	Benar
7	71	C3	C3	Benar
8	71	C3	C3	Benar
9	72	C3	C3	Benar
10	65	C4	C4	Benar
11	66	C4	C4	Benar
12	66	C4	C4	Benar
13	62	C5	C4	Salah
14	62	C5	C4	Salah
15	60	C5	C5	Benar
16	57	C6	C5	Salah
17	58	C6	C5	Salah
18	55	C6	C5	Salah
19	50	C7	C6	Salah
20	50	C7	C6	Salah
21	44	C7	C7	Benar

$$akurasi = \frac{\text{jumlah data benar}}{\text{jumlah data}} \times 100\%$$

sebanyak 15 data yang diambil dari rekap nilai siswa dari tahun sebelumnya. Adapun implementasi proses pelatihan data kecil dapat dilihat pada Gambar 14



Gambar 14 Implementasi data besar

Pada Gambar 14 terdapat 34 data latih dengan 17 data yang benar dan sisanya data tidak sesuai dengan target. Jadi, untuk menghitung akurasi dari proses data latih kecil sebagai berikut

$$\begin{aligned} \text{akurasi} &= \frac{17}{34} \times 100\% \\ &= 51.515\% \end{aligned}$$

Adapun data benar dari proses pelatihan data kecil yaitu, data ke 2, 4, 5, 6, 8, 10, 12, 16, 17, 19, 15, 28, 29, 30, 31, 32, 34.

Untuk akurasi data uji pada proses pelatihan data besar menggunakan 15 pengujian dengan 9 jumlah data benar dari 15 data uji. Jadi, untuk menghitung akurasi dari proses data uji sebagai berikut:

$$\begin{aligned} \text{akurasi} &= \frac{9}{15} \times 100\% \\ &= 60\% \end{aligned}$$

Adapun data benar dari proses pelatihan data kecil yaitu, data ke 1, 2, 3, 6, 7, 9, 10, 12, 15.

5. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian penerimaan siswa dengan menggunakan metode *k-means* diperoleh kesimpulan sebagai berikut:

- Praposes dari sitem yaitu dengan cara mengambil bobot nilai dari masing-masing soal ujian, kemudian bobot nilai tersebut dilatih dengan menggunakan metode *k-means*. Proses pelatihan *k-means* digunakan untuk mencari nilai centroid dari data yang awalnya tidak beraturan menjadi data yang beraturan. Nilai centroid yang didapatkan pada saat proses pelatihan, digunakan untuk menghitung bobot jawaban yang benar pada saat ujian dilaksanakan. Pada penelitian ini, jumlah kelas yang digunakan sebanyak delapan

yaitu, 'Diterima pada kelompok 1, kelompok 2, kelompok 3 sampai dengan kelompok 7.'

- Adapun kesimpulan yang dikeluarkan oleh sistem dengan cara membandingkan jurusan yang dipilih oleh calon siswa dan jurusan yang diterima. Untuk membandingkan jurusan yang dipilih oleh calon siswa dengan jurusan yang diterima digunakan *grade* (peringkat) yang telah ditentukan oleh pihak sekolah.
- Pada proses pelatihan dengan menggunakan 22 data latih menghasilkan akurasi sebesar 68.182% , sedangkan pada proses pengujian dengan menggunakan 10 data uji menghasilkan akurasi sebesar 70%

Pada proses pelatihan data besar dengan menggunakan 34 data latih menghasilkan akurasi sebesar 51.515% sedangkan pada proses pengujian dengan menggunakan data uji sebanyak 15 menghasilkan presentase sebesar 60%.

5.2. Saran

Berdasarkan hasil penelitian dan kesimpulan, maka dapat diberikan saran untuk kepentingan peneliti selanjutnya adalah sebagai berikut:

- Pada saat proses pelatihan, terkadang system mengluarkan nilai centroid yang kurang tepat karena nilai centroid dibangkitkan secara random (acak). Maka dari itu, peneliti selanjutnya dapat menggunakan beberapa metode klasifikasi seperti Naïve Bayes, SVM (Support Vector Machine) dan Jaringan Saraf Tiruan sehingga dapat membandingkan hasil akurasi dari system yang telah dibuat.
- Peneliti selanjutnya juga diharapkan menambahkan parameter jurusan agar dapat diterapkan di sekolah-sekolah maupun kampus, sehingga siswa atau calon mahasiswa tidak salah dalam menentukan jurusan.

DAFTAR PUSTAKA

- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- Ong, J. O. (2013). IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY. *Jurnal Ilmiah Teknik Industri*, 12(1), 10 – 20.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*, 2nd ed. Kanada: Morgan Kaufmann.
- Sudrajat, C. (2011). Pengembangan Sistem Informasi Pendaftaran Siswa Baru Secara Online Berbasis Web.

