



**PENERAPAN ALGORITMA K-MEANS *CLUSTERING*
UNTUK MENENTUKAN PENERIMAAN PEGAWAI
PT. KERETA API INDONESIA DIVRE 1 SUMATERA UTARA**

Disusun Dan Diajukan Untuk Memenuhi Persyaratan Ujian Akhir
Memperoleh Gelar Sarjana Komputer Pada Fakultas Sains Dan Teknologi
Universitas Pembangunan Panca Budi
Medan

SKRIPSI

OLEH

**NAMA : Firmansyah
N.P.M : 1314370081
PROGRAMSTUDI : SISTEM KOMPUTER**

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN PANCA BUDI
MEDAN
2019**

LEMBAR PENGESAHAN

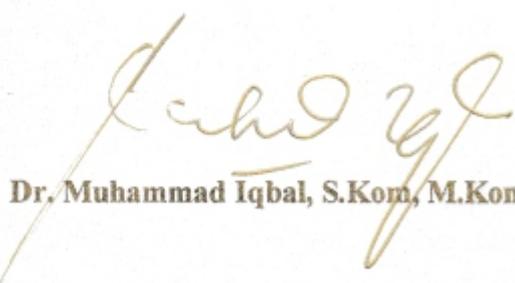
**PENERAPAN ALGORITMA K-MEANS *CLUSTERING*
UNTUK MENENTUKAN PENERIMAAN PEGAWAI
PT. KERETA API INDONESIA DIVRE 1 SUMATERA UTARA**

DISUSUN OLEH :

**Nama : Firmansyah
Npm : 13143
Program Studi : Sistem Komputer**

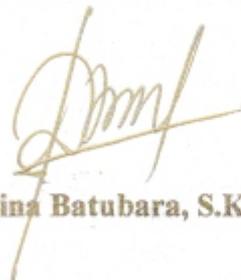
**Skripsi telah disetujui oleh Dosen Pembimbing Skripsi
Pada Tanggal : 27 Agustus 2019**

Dosen Pembimbing I



Dr. Muhammad Iqbal, S.Kom, M.Kom

Dosen Pembimbing II



Supina Batubara, S.Kom, M.Kom

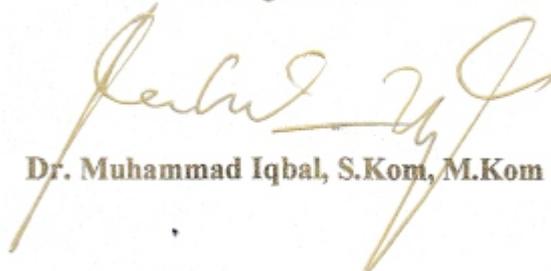
Mengetahui,

Dekan Fakultas Sains & Teknologi



Sri Shanti Indira, S.T, M.Sc

Ketua Program Studi



Dr. Muhammad Iqbal, S.Kom, M.Kom

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

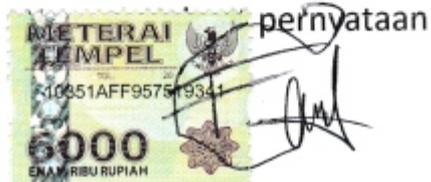
Nama : FIRMANSYAH
NPM : 1314370081
Prodi : SAINS DAN TEKNOLOGI
Konsentrasi : SISTEM KOMPUTER
Judul Skripsi : PENERAPAN ALGORITMA K-MEANS CLUESTERING
UNTUK MENENTUKAN PENERIMAAN PEGAWAI
PT.KERETA API INDONESIA DIVRE 1 SUMATERA UTARA

Dengan ini menyatakan bahwa :

1. Tugas Akhir/Skripsi saya bukan hasil Plagiat
2. Sayat tidak akan menuntut perbaikan nilai indeks Prestasi Kumulatif (IPK) setelah ujian Sidang Meja Hijau
3. Skripsi saya dapat dipublikasikan oleh pihak lembaga, dan saya tidak akan menuntut akibat publikasi tersebut

Demikian pernyataan ini saya perbuat dengan sebenar-benarnya, terima kasih

Medan, 31 Agustus 2019



FIRMANSYAH
NPM:1314370081

PERNYATAAN ORISINALITAS

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam skripsi ini dan di sebutkan dalam daftar pustaka.

Medan, 31 Agustus 2019
Yang membuat pernyataan



FIRMANSYAH
NPM:1314370081

UNIVERSITAS PEMBANGUNAN PANCA BUDI FAKULTAS ILMU KOMPUTER

Jl. Jend. Gatot Subroto Km 4,5 Medan Fax. 061-8458077 PO.BOX : 1099 MEDAN

PROGRAM STUDI SISTEM KOMPUTER
PROGRAM STUDI TEKNIK KOMPUTER

(TERAKREDITASI)
(TERAKREDITASI)

PERMOHONAN MENGAJUKAN JUDUL SKRIPSI

Peranda tangan di bawah ini :

Nama : FIRMANSYAH
 Tanggal Lahir : BINJAI / 24 Januari 1996
 NIM / Mahasiswa : 1314370081
 Jurusan : Sistem Komputer
 Bidang yang telah dicapai : Keamanan Jaringan Komputer

yang mengajukan judul skripsi sesuai dengan bidang ilmu, dengan judul:

Judul Skripsi	Persetujuan
ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI PT. KERETA API INDONESIA SUMATERA UTARA	<input type="checkbox"/>
ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN BEASISWA BAGI SISWA KURANG MAMPU PADA MEDAN	<input type="checkbox"/>
ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN PENERIMAAN PEGAWAI PT. KERETA API SUMATERA UTARA	<input checked="" type="checkbox"/> <i>ff</i>

Disetujui oleh Kepala Program Studi diberikan tanda

Rektor I,
(In. Bhakti Alamsyah, M.T., Ph.D.)

Medan, 11 Agustus 2017
 Pemohon,

 (FIRMANSYAH)

Komor :
 Tanggal : 02-10-2017
 Disahkan oleh :

 (In. Bhakti Alamsyah, M.T., Ph.D.)

Tanggal : 15 Sept 2017
 Disetujui oleh :
 Dosen Pembimbing I :

 (M. Irfan)

Tanggal :
 Disetujui oleh:
 Ka. Prodi Sistem Komputer

 (Eko Hariyanto, S.Kom., M.Kom)

Tanggal :
 Disetujui oleh:
 Dosen Pembimbing II:

 (Kama Saputra S.)

Dokumen: FM-LPPM-08-01	Revisi: 02	Tgl. Eff: 20 Des 2015
------------------------	------------	-----------------------

No. 1870 / Perp / Op / 2019

Dinyatakan tidak ada sangkut paut dengan UPT. Perpustakaan

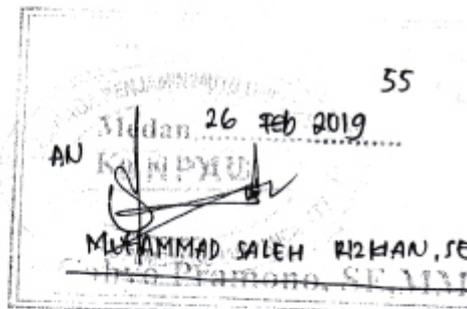
Medan, 26 FEB 2019



Medan, 26 Februari
Kepada Yth : Ba
Fakultas SAINS &
UNPAB Medan
Di -
Tempat

Hal : Permohonan Meja Hijau

Dengan hormat, saya yang bertanda tangan di bawah ini.
Nama : FIRMANSYAH
Tempat/Tgl. Lahir : BINJAI / 24 Januari 1996
Nama Orang Tua : BARUSTAM
N. P. M : 1314370081
Fakultas : SAINS & TEKNOLOGI
Program Studi : Sistem Komputer
No. HP : 082166084285
Alamat : Jln.A.R.Hakim Gg.raden



Datang bernohon kepada Bapak/Ibu untuk dapat diterima mengikuti Ujian Meja Hijau dengan judul PENERAPAN ALGO CLUSTERING UNTUK MENENTUKAN PENERIMAAN PEGAWAI PT. KERETA API INDONESIA DIVRE I SUMATERA UTARA, Se menyatakan :

1. Melampirkan KKM yang telah disahkan oleh Ka. Prodi dan Dekan
2. Tidak akan menuntut ujian perbaikan nilai mata kuliah untuk perbaikan indek prestasi (IP), dan mohon diterbit lulus ujian meja hijau.
3. Telah tercap keterangan bebas pustaka
4. Terlampir surat keterangan bebas laboratorium
5. Terlampir pas photo untuk ijazah ukuran 4x6 = 5 lembar dan 3x4 = 5 lembar Hitam Putih
6. Terlampir foto copy STTB SLTA dilegalisir 1 (satu) lembar dan bagi mahasiswa yang lanjutan D3 ke S1 lampirkan sebanyak 1 lembar.
7. Terlampir pelunasan kwintasi pembayaran uang kuliah berjalan dan wisuda sebanyak 1 lembar
8. Skripsi sudah dijilid lux 2 exemplar (1 untuk perpustakaan, 1 untuk mahasiswa) dan jilid kertas jeruk 5 exampl. dan warna penjilidan diserahkan berdasarkan ketentuan fakultas yang bertaku) dan lembar persetujuan sudah pembimbing, prodi dan dekan
9. Soft Copy Skripsi disimpan di CD sebanyak 2 disc (Sesuai dengan Judul Skripsinya)
10. Terlampir surat keterangan BKKOL (pada saat pengambilan ijazah)
11. Setelah menyelesaikan persyaratan point-point diatas berkas di masukan kedalam MAP
12. Bersedia melunaskan biaya-biaya uang dibebankan untuk memproses pelaksanaan ujian dimaksud, dengan peri

1. [102] Ujian Meja Hijau	: Rp.	250,000
2. [170] Administrasi Wisuda	: Rp.	1,500,000
3. [202] Bebas Pustaka	: Rp.	100,000
4. [221] Bebas LAB	: Rp.	5,000

Total Biaya : Rp. 1,605,000

Uk.-T.50%

1.855.000
2.500.000 + 26/02-19
Rp. 4.355.000

Ukuran Toga :

M

Diketahui/Diseetujui oleh :
Sri Shidiq, S.T., M.Sc.
Dekan Fakultas SAINS & TEKNOLOGI

Telah di terima
berkas persyaratan
dapat di proses
Medan, 26 FEB 2019
a.w. Ka. BPAA
TEGUH WAHYONO, SE., NIM.

Hormat saya
FIRMANSYAH
1314370081

Catatan :

- 1. Surat permohonan ini sah dan bertaku bila ;
 - a. Telah dicap Bukti Pelunasan dari UPT Perpustakaan UNPAB Medan.

Plagiarism Detector v. 1092 - Originality Report:

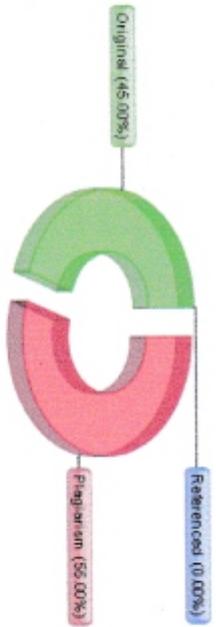
Analyzed document: 26-12-18 8:04:02 AM

"FIRMANSYAH_1314370081_SYSTEM KOMPUTER.docx"

Licensed to: Universitas Pembangunan Panca Budi_License2



Relation chart:



Distribution graph:





UNIVERSITAS PEMBANGUNAN PANCA BUDI
FAKULTAS ILMU KOMPUTER
 Jl. Jend. Gatot Subrot Km. 4,5 Telp (061)-
 Medan - Indonesia

FM-BPAA-2012-038

Universitas : Universitas Pembangunan Panca Budi
 Fakultas : Ilmu Komputer
 Dosen Pembimbing I : Muhammad Iqbal, S.Kom, M.Kom
 Dosen Pembimbing II : Kana Safutra, S.S.Pd, M.Kom
 Nama Mahasiswa : Fitman Syah
 Jurusan/Program Studi : Keamanan Jaringan Komputer
 Nomor Pokok Mahasiswa : 21A370081
 Bidang Pendidikan : Strata Satu (S-1) / Diploma Tiga (D-III)
 Judul Tugas Akhir /Skripsi : Penerapan algoritma k-means clustering untuk
 menentukan penentuan Pesawat PI-kebeta api
 Indonesia, DivRE 1 Sumatera Utara

Tanggal	Pembahasan Materi	Paraf	Keterangan
11/17	~ + 2 referensi (jurnal) dalam latar belakang	✓	
11/18	~ perbaikan rumusan masalah	✓	
11/18	~ Aca BAB I	✓	
11/18	~ Aca BAB II	✓	
11/18	~ penjelasan mengenai pembagian metode k-means	✓	
11/18	~ Aca BAB III	✓	
11/18	~ + penjelasan pada setiap subbab	✓	
11/18	~ Aca BAB IV	✓	
11/18	~ Aca BAB V dan seminar	✓	
11/19	~ Aca Friday	✓	
11/19	~ Aca Jumat	✓	

Medan,
 Diketahui/Disetujui
 oleh:
 Dekan





UNIVERSITAS PEMBANGUNAN PANCA BUDI
FAKULTAS ILMU KOMPUTER
 Jl. Jend. Gatot Subrot Km. 4,5 Telp (061)-
 Medan - Indonesia

FM-BPAA-2012-038

Universitas : Universitas Pembangunan Panca Budi
 Fakultas : Ilmu Komputer
 Dosen Pembimbing I : Muhammad Lahal, S.Kom., M.Kom
 Dosen Pembimbing II : Kana Salotra S.S.P., M.Kom
 Nama Mahasiswa : Firmansyah
 Jurusan/Program Studi : Keamanan Jaringan Komputer
 Nomor Pokok Mahasiswa : 1314370001
 Bidang Pendidikan : Strata Satu (S-1) / Diploma Tiga (D-III) *)
 Judul Tugas Akhir /Skripsi : Penerapan Algoritma K-means Clustering untuk
 menentukan Penetiran Pegawai PT Kereta Api
 Indonesia Divisi I Sumatera Utara

Tanggal	Pembahasan Materi	Paraf	Keterangan
2014	Penyusunan Bab I - K-means Clustering	[Signature]	
2014	Arahan Bab II	[Signature]	
2014	Penyusunan Bab II	[Signature]	
2014	Arahan Bab II, Penyusunan Bab III	[Signature]	
2014	Arahan Bab III	[Signature]	
2014	Arahan Bab IV	[Signature]	
2014	Arahan Bab V	[Signature]	
2014	Arahan Bab VI	[Signature]	
2014	Arahan Bab VII	[Signature]	
2014	Arahan Bab VIII	[Signature]	

Medan,
 Diketahui/Disetujui
 oleh :
 Dekan





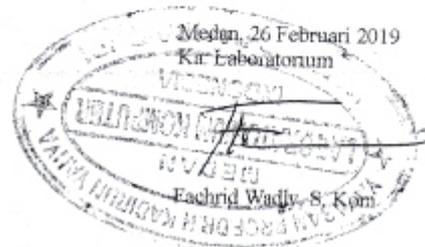
YAYASAN PROF. DR. H. KADIRUN YAHYA
UNIVERSITAS PEMBANGUNAN PANCA BUDI
LABORATORIUM KOMPUTER
Jl. Jend. Gatot Subroto Km 4,5 Sei Sikambing Telp. 061-8455571
Medan - 20122

KARTU BEBAS PRAKTIKUM

Yang bertanda tangan dibawah ini Ka. Laboratorium Komputer dengan ini menerangkan bahwa :

Nama : FIRMANSYAH
N.P.M : 1314370081
Tingkat/Semester : Akhir
Fakultas : SAINS & TEKNOLOGI
Jurusan/Prodi : Sistem Komputer

Benar dan telah menyelesaikan urusan administrasi di Laboratorium Komputer Universitas Pembangunan Panca Budi Medan.



ABSTRAK

Penumpukan data calon pegawai secara menerus akan memperlambat proses penerimaan pegawai baru pada PT. Kereta Api Indonesia (Persero). Salah satu metode yang dapat digunakan untuk dapat menyelesaikan permasalahan tersebut adalah metode K-Means *Clustering*. K-Means *Clustering* merupakan salah satu metode data *clustering* non hirarki yang membagi data yang ada kedalam bentuk satu atau lebih *cluster* / kelompok. Dengan menggunakan metode K-Means *clustering* dimana di- *cluster* berdasarkan nilai kriteria kelulusan yang telah ditentukan, yaitu dengan penentuan nilai awal titik pusat (*centroid*) dimana nilai *centroid* awal mempengaruhi nilai *centroid* berikutnya dan penentuan nilai *cluster* berikutnya, apabila *cluster* sebelumnya memiliki pola yang sama dengan pola *cluster* selanjutnya maka perhitungan dihentikan. Sehingga diperoleh hasil *cluster* akhir, tetapi sebaliknya jika berbeda dilakukan perhitungan kembali sampai batasi terasi yang telah ditentukan. Dari hasil uji coba yang telah dilakukan diperoleh clustering data dengan lima kriteria dan dua kategori.

Kata Kunci: Clustering, K-Means

DAFTAR ISI

LEMBAR JUDUL	
LEMBAR PENGESAHAN	
ABSTRAK	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL.....	vii
DAFTAR LAMPIRAN	viii
BAB I : PENDAHULUAN	
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	5
1.7 Sistematika Penulisan	6
BAB II : LANDASAN TEORI	
1 Data Mining	8
1.1 Pengelompokan Data Mining	8
1.2 Proses Data Mining	11
1.3 Proses Pencarian Pola dalam Data Mining	13
1.4 Transformasi Data	13
2. Metode Clustering	18
3. K-Means	21
3.1 Metode K-Means	25
3.2 Algoritma K-Means	26
BAB III : ANALISA MASALAH DAN RANCANGAN PROGRAM	
1. Data yang Digunakan	35
2. Metode Penelitian	35
2.a. Studi Pustaka	36
2.b. Wawancara	36
2.c. Analisa Data	36
3. Perancangan Sistem.....	37
3.a Algoritma K-Means Clustering	37
3.b. Desain Input dan Output.....	37
4. Analisa Data Mining.....	39
4.a Nilai Centroid Awal.....	40
4.b Perhitungan Jarak	41
BAB IV : IMPLEMENTASI DAN HASIL ANALISA UJI COBA	
1. Implementasi	47
1.a Spesifikasi Sistem.....	47

1.b.	Aplikasi Metode K-Means.....	48
2.	Hasil Uji Coba	48
2.a	Implementasi Algoritma K-Means Clustering	48
2.a.1	Nilai Centroid Awal.....	48
2.a.2.	Perhitungan Jarak	50
2.b.	Implementasi Algoritma K-Means Clustering dengan Software .	57
2.b.1.	Form Login.....	57
2.b.2.	Form Data Calon Pegawai.....	58
2.b.3.	Form Hasil Clustering	58
 BAB V : PENUTUP		
5.1	Kesimpulan.....	60
5.2	Saran	60
 DAFTAR PUSTAKA		61

DAFTAR GAMBAR

Gambar 1.	Sistematika Penulisan	5
Gambar 2.	Algoritma K-Means Secara Umum	29
Gambar 3.	Metodologi Penelitian	35
Gambar 4.	Form Login	38
Gambar 5.	Form Menu	39
Gambar 6.	Form Login	57
Gambar 7.	Form Data Calon Pegawai	58
Gambar 8.	Hasil Akhir dari Tes Calon Pegawai	59

DAFTAR TABEL

Tabel 1.	Sampel Data	30
Tabel 2.	Data Centroid Awal.....	30
Tabel 3.	<i>Cluster</i> Iterasi 1	31
Tabel 4.	Data Centroid Iterasi 1	32
Tabel 5.	<i>Cluster</i> Iterasi 2	32
Tabel 6.	Atribut Kategori Kelulusan Calon Pegawai	39
Tabel 7.	Data Kriteria Kelulusan Calon Pegawai.....	40
Tabel 8.	Data Centroid Awal.....	41
Tabel 9.	Cluster Iterasi 1	43
Tabel 10.	Data Centroid Baru Iterasi 1.....	44
Tabel 11.	<i>Cluster</i> Iterasi 2	44
Tabel 12.	Hasil <i>Clustering</i> Data dari Iterasi 1 dan Iterasi 2.....	46
Tabel 13.	Data Kriteria Kelulusan Calon Pegawai.....	49
Tabel 14.	Data Centroid Awal.....	50
Tabel 15.	Cluster Iterasi 1	52
Tabel 15.	Cluster Iterasi 1 (Lanjutan)	53
Tabel 16.	Data Centroid Baru Iterasi 1.....	54
Tabel 17.	<i>Cluster</i> Iterasi 2	54
Tabel 17.	<i>Cluster</i> Iterasi 2 (Lanjutan)	55
Tabel 18.	Hasil <i>Clustering</i> Data dari Iterasi 1 dan Iterasi 2	55
Tabel 18.	Hasil <i>Clustering</i> Data dari Iterasi 1 dan Iterasi 2 (Lanjutan)	56
Tabel 19.	Hasil Akhir Perekrutan Pegawai	56

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Lembar Pengesahan Tugas Akhir.....	L-1
Lampiran 2 Coding.....	L-2
Lampiran 3 Biografi Penulis	L-3
Lampiran 4 Berita Acara Bimbingan Skripsi	L-4

BAB I

PENDAHULUAN

1. Latar Belakang

Kemajuan teknologi informasi sudah semakin berkembang pesat di segala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, politik, dan pendidikan serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pekerjaan juga dapat menghasilkan data yang berlimpah mengenai proses penerimaan pegawai yang dihasilkan.

PT Kereta Api Indonesia (Persero) atau PT KAI merupakan salah satu Badan Usaha Milik Negara Indonesia yang menyediakan produk jasa angkutan dengan menggunakan kereta api. Adapun layanan PT KAI adalah angkutan penumpang dan barang. Historis data PT. Kereta Api Indonesia (Persero) dapat diperoleh berdasarkan data akan bertambah secara terus menerus, misalnya data penerimaan pegawai. Proses penerimaan pegawai baru pada PT. Kereta Api Indonesia (Persero) sangat tinggi menghasilkan data yang berlimpah berupa profil dari calon pegawai baru tersebut. Tahap selanjutnya calon pegawai akan melakukan proses seleksi kesehatan awal, sehingga dapat diketahui data setiap calon pegawai yang akan lanjut ke tahap kesehatan awal. Hal ini akan terjadi secara berulang pada PT. Kereta Api Indonesia. Penumpukan data calon pegawai secara menerus akan memperlambat proses penerimaan pegawai baru pada PT. Kereta Api Indonesia (Persero).

Proses seleksi penerimaan pegawai baru secara manual yaitu dengan menginputkan satu per satu hasil seleksi ke dalam file *spreadsheet* kemudian melakukan *sorting* data calon pegawai baru seringkali menimbulkan beberapa permasalahan, antara lain membutuhkan waktu yang lama dan ketelitian yang tinggi. Selain itu, transparansi serta ketidakjelasan metodologi yang digunakan dalam proses komputasi penerimaan calon pegawai baru juga menjadi salah satu permasalahan, sehingga dibutuhkan suatu sistem yang dapat membantu dalam proses pengambilan keputusan siapa saja calon pegawai baru yang lulus pada setiap tahap, berdasarkan kriteria-kriteria yang telah ditentukan secara cepat dan tepat sasaran.

Salah satu metode yang dapat digunakan untuk dapat menyelesaikan permasalahan tersebut adalah metode K-Means *Clustering*. K-Means *Clustering* merupakan salah satu metode data *clustering* non hirarki yang membagi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini membagi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama (Agusta,2007).

Pada penelitian sebelumnya, Asroni (2015) melakukan pembahasan mengenai penerapan metode K-Means untuk *clustering* mahasiswa berdasarkan nilai akademik dengan *weka interface* studi kasus pada Jurusan Teknik Informatika UMM Magelang. Dari penelitian tersebut didapatkan bahwa algoritma K-Means dapat digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan beberapa atribut mata kuliah.

Penelitian sebelumnya, Wicaksono (2016) melakukan 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). Dari penelitian tersebut didapatkan bahwa algoritma K-Means dapat digunakan untuk mengelompokkan penerima beasiswa berdasarkan nilai akademik di sekolah dan gaji orang tua siswa.

Pada penerimaan calon pegawai baru yang dilakukan oleh pihak PT. Kereta Api Indonesia, dinyatakan lulus dan berhak menjadi pegawai baru harus memenuhi kriteria sebagai berikut seleksi administrasi, kesehatan awal, psikologi, wawancara, dan kesehatan akhir. Berdasarkan latar belakang yang telah penulis uraikan di atas, maka penulis tertarik untuk memilih judul ***“Penerapan Algoritma K-Means Clustering Untuk Menentukan Penerimaan Pegawai PT. Kereta Api Indonesia Sumatera Utara”***.

2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas dapat penulis simpulkan bahwa yang menjadi pokok permasalahan dalam pembahasan ini adalah sebagai berikut:

- a. Bagaimana menerapkan metode algoritma *K-Means Clustering* dalam menentukan penerimaan pegawai pada PT. Kereta Api Indonesia.
- b. Bagaimana merancang sebuah perangkat lunak yang dapat digunakan untuk menentukan penerimaan pegawai dengan menggunakan algoritma *K-Means Clustering*.

3. Batasan Masalah

Berdasarkan perumusan masalah diatas maka penulis melakukan pembatasan masalah yang akan dibahas sebagai berikut:

- a. Penerimaan pegawai dihitung berdasarkan hasil administrasi, kesehatan awal, psikologi, wawancara, dan kesehatan akhir.
- b. Metode pengelompokan data yang digunakan adalah *K-Means Clustering*.
- c. Program yang dibahas menggunakan pemrograman Visual C# 2010.

4. Tujuan Penelitian

Adapun tujuan dari penelitian ini dengan menggunakan algoritma *K-Means Clustering* adalah sebagai berikut:

- a. Untuk menerapkan cara kerja algoritma *K-Means Clustering* dalam melakukan pengelompokan terhadap data berdasarkan input yang dimasukkan.
- b. Untuk menganalisis dan merancang sistem penentuan pegawai baru dengan menggunakan *K-Means Clustering*.

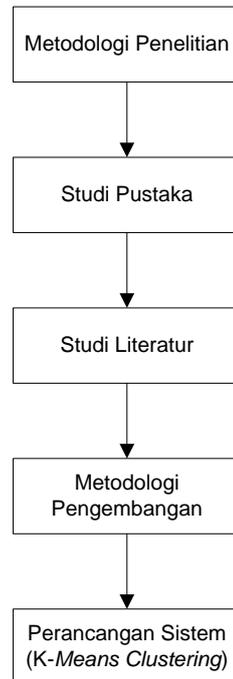
5. Manfaat Penelitian

Adapun Manfaat dari penelitian ini dengan menggunakan algoritma *K-Means Clustering* adalah sebagai berikut:

- a. Perangkat lunak dapat digunakan oleh PT. Kereta Api Indonesia untuk menentukan pegawai baru yang akan bekerja.
- b. Menciptakan proses penentuan calon pegawai pada PT. Kereta Api Indonesia menjadi terkomputerisasi.

6. Metodologi Penelitian

Dalam metodologi penelitian ini peneliti menggunakan beberapa metode dalam pengumpulan data untuk melengkapi hasil penelitian ini. Adapun metode tersebut sebagai berikut:



Gambar 1. Sistematika Penulisan

a. Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Studi Pustaka yaitu pengumpulan data yang diperoleh dari sumber tertulis berupa buku-buku, artikel ilmiah, dan penelitian-penelitian yang berkaitan dengan judul penelitian.

2. Studi literature yaitu pengumpulan data yang diperoleh dari literature, jurnal, paper, dan bacaan-bacaan dari berbagai sumber yang berkaitan dengan judul penelitian.

b. Metode pengembangan dan perancangan sistem pada kasus ini menggunakan algoritma K-Means *Clustering* yang merupakan salah satu contoh metode yang mempartisi data ke dalam *cluster* atau 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..

7. Sistematika Penulisan

Secara garis besar sistematika penulisan tugas akhir ini terdiri dari lima bab yaitu sebagai berikut:

BAB I PENDAHULUAN

Pada bab pendahuluan ini, akan menguraikan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab landasan teori ini, akan memaparkan teori-teori yang digunakan sebagai panduan dalam menyelesaikan

skripsi sesuai dengan judul yang diteliti dan dapat diperoleh dari berbagai sumber.

BAB III ANALISA MASALAH DAN RANCANGAN PROGRAM

Pada bab analisa masalah dan rancangan program ini, menjelaskan tentang gambaran pembahasan permasalahan yang terjadi serta perancangan sistem yang ingin diselesaikan.

BAB IV IMPLEMENTASI DAN ANALISA HASIL

ASIL UJI COBA PROGRAM

Pada bab implementasi dan analisa hasil uji coba program ini, membahas tentang hasil implementasi yang dibuat serta melakukan analisa terhadap hasil tersebut.

BAB V PENUTUP

Pada bab penutup ini, berisi tentang kesimpulan dan saran yang diperoleh dari hasil penelitian untuk pengembangan serta perbaikan yang di perlukan dari hasil penelitian ini.

BAB II

LANDASAN TEORI

1. *Data Mining*

Data mining merupakan suatu metode untuk meolah data yang mana menentukan pola tersembunyi dari data tersebut. Pengolahan data yang dihasilkan dari suatu metode data mining ini dapat digunakan untuk mengambil beberapa keputusan. Pengolahan data yang membutuhkan skala besar, sehingga data mining mempunyai peranan penting dalam berbagai bidang seperti industri, keuangan, cuaca, ilmu dan teknologi (Ong, 2013). Data mining dapat diterapkan pada database dan menyimpan berbagai informasi, tetapi jenis pola yang akan ditemukan dapat ditentukan oleh berbagai fungsi data mining seperti deskripsi class/konsep, asosiasi, analisa korelasi, klasifikasi, prediksi, analisa cluster dan lain-lain (Dash et al., 2010).

Ada berbagai penerapan algoritma yang digunakan dalam data mining untuk dapat menentukan pola dari data yang ingin dianalisa, yang mana data yang diproses secara otomatis akan menentukan pola yang sederhana terhadap data yang besar dengan menggunakan beberapa hasil yang telah ditetapkan oleh pengguna (Ndehedehe et al., 2013).

1) Pengelompokan Data mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005) :

a.Deskripsi (Description)

Terkadang penelitian analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

b.Estimasi (Estimation)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

c.Prediksi (Prediction)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa datang. Contoh prediksi dalam bisnis

dan penelitian adalah prediksi harga beras dalam tiga bulan yang akan datang, prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan untuk prediksi.

d. Klasifikasi (Classification)

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan, memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk, dan mendiagnosis penyakit seorang pasien untuk mendapatkan kategori penyakit apa.

e. Pengklusteran (Clustering)

Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam cluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam Pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma Pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data

menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal. Contoh Pengklusteran dalam bisnis dan penelitian adalah melakukan Pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar, dan untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam keadaan baik atau mencurigakan.

f. Asosiasi (Assosiation)

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan, dan meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons positif terhadap penawaran upgrade layanan yang diberikan.

2. Proses Data Mining

Fase-fase dimulai dari data mentah dan berakhir dengan pengetahuan atau informasi yang telah diolah, yang didapatkan sebagai hasil dari tahapan-tahapan berikut:

1. Data Cleansing, juga dikenal sebagai data cleansing, ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih relevan dapat digunakan untuk diproses ulang untuk penggalian pengetahuan(discovery knowledge)
2. Data Integration, pada tahap ini terjadi integrasi data,dimana sumber-sumber data yang berulang(multiple data), file-file yang berulang(multiple file), dapat dikombinasikan dan digabungkan kedalam suatu sumber.
3. Data Selection, pada langkah ini, data yang relevan terhadap analisis dapat dipilih dan diterima dari koleksi data yang ada.
4. Data Transformation, juga dikenal sebagai data consolidation. Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan kedalam bentuk-bentuk yang cocok untuk prosedur penggalian (meaning procedure) dengan cara melakukan normalisasi dan agregasi data.
5. Data Mining, tahap ini adalah tahap yang paling penting, dengan menggunakan teknik-teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna.
6. Pattern Evaluation, pada tahap ini, pola-pola menarik dengan jelas mempresentasikan pengetahuan telah diidentifikasi berdasarkan measure yang telah diberikan.
7. Knowledge Representation, ini merupakan tahap terakhir dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada

user. Tahap penting ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.

3. Proses Pencarian Pola dalam Data Mining

Proses pencarian pola atau bisa dikatakan sebagai proses penambangan data penting. Seperti menambang pada umumnya yang memerlukan pencarian untuk mendapatkan sesuatu yang penting. Nah berikut proses pencarian pola dalam menemukan data penting.

1. Pembersihan data

Proses pencarian pola yang pertama adalah proses pembersihan data. Pembersihan data berupa penghapusan data pengganggu atau data yang tidak penting serta mengisi data yang hilang.

2. Integrasi data

Setelah pembersihan data, proses selanjutnya yaitu integrasi data. Integrasi data merupakan proses penggabungan beberapa sumber data yang ada.

3. Pemilihan data

Pencarian pola selanjutnya dalam adalah pemilihan data. Data-data yang relevan nantinya dipilih dan dikumpulkan.

4. Transformasi data

Setelah itu proses selanjutnya adalah transformasi data. Jadi dari banyaknya data, nantinya akan diproses dan ditransformasi ke dalam format tertentu, format yang akan digunakan dalam penggalian data. 5. Penggalian data
Dalam proses satu ini, data akan diolah menggunakan metode yang cerdas dan canggih sehingga akan menghasilkan ekstraksi pola tertentu.

6. Evaluasi pola

Proses selanjutnya yaitu evaluasi pola. Dari pola-pola yang ditemukan, nantinya akan dikenali pola-pola yang menarik. Pola-pola menarik tersebut lah yang akan diambil.

7. Penyajian pola

Setelah ditemukan pola yang menarik, pola tersebut kemudian akan divisualisasikan ke pengguna.

5. Kelebihan Data Mining

Data mining merupakan sebuah proses interatif dan interaktif untuk mendapatkan sebuah pola baru yang menarik. Pola tersebut tentunya akan sangat bermanfaat. Model yang dihasilkan dari proses data mining biasanya sudah sempurna sehingga dapat digeneralisasi untuk kepentingan di masa depan. Karena prosesnya yang cukup panjang dan rumit, maka dari proses awal biasanya akan menghasilkan sesuatu yang baru, yang tidak diketahui sebelumnya. Sesuatu yang baru ini akan menambah pengetahuan para pengguna ataupun peneliti dan tentunya akan sangat bermanfaat karena dapat digunakan untuk melakukan tindakan tertentu. Penggalan data juga sering dikatakan sebagai proses interaktif dan interatif. Proses interaktif maksudnya yaitu proses yang masih memerlukan interaksi manusia agar bisa terlaksana. Sedangkan proses interatif, maksudnya adalah proses yang tidak hanya dilakukan sekali, perlu proses yang berulang-ulang untuk mendapatkan data penting yang dimaksud. Kelebihannya membuat analisa suatu data besar menjadi semakin mudah. Pencarian pola baru atau trend

baru bisa dilakukan dengan mudah sehingga bisa membantu mengambil keputusan di masa yang akan datang atau bisa memprediksi data tertentu sehingga bisa menganalisis apa yang harus dilakukan.

Suatu proses pengelompokan data serupa ke dalam beberapa kelompok yang berbeda atau membentuk partisi dari suatu data set ke dalam bentuk data sub set, sehingga data yang telah dibentuk dalam sub set tersebut memiliki arti yang bermanfaat yang mana disebut sebagai *clustering*. Suatu *cluster* terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan suatu benda yang terdapat pada *cluster* lainnya.

Pada proses ini dilakukan pengelompokan objek ke dalam bentuk subset yang memiliki arti dalam konsep masalah tertentu itulah yang disebut dengan analisis *clustering*. *Clustering* tidak sama dengan klasifikasi yang mana *clustering* tidak berdasarkan pada kelas yang sudah ada. Metode yang mempelajari tentang *unsupervised* karena tidak adanya informasi yang disajikan dalam bentuk pernyataan benar untuk objek apa pun, hal tersebut disebut dengan *clustering*. Sehingga dapat ditemukan hubungan dari sebelumnya yang tidak diketahui di dalam suatu *data set* yang kompleks.

Analisis *clustering* merupakan teknik analisis yang multivariasi dimana dilakukan untuk mencari dan mengorganisir suatu informasi tentang variabel, sehingga secara relatif dapat ditentukan kelompoknya ke dalam bentuk kelompok yang homogen atau ter-*cluster*. *Cluster* yang dibentuk merupakan metode kedekatan yang secara internal harus berupa homogen yang mana anggota sama

dengan anggota yang lain dan secara eksternal tidak sejenis yang mana anggota tidak sama dengan anggota yang lain.

Syarat *Clustering*

Menurut Han dan Kamber, 2012, syarat sekaligus tantangan yang harus dipenuhi oleh suatu algoritma *clustering* adalah :

1. Skalabilitas

Suatu metode *clustering* harus mampu menangani data dalam jumlah yang besar. Saat ini data dalam jumlah besar sudah sangat umum digunakan dalam berbagai bidang misalnya saja suatu database. Tidak hanya berisi ratusan objek, suatu database dengan ukuran besar bahkan berisi lebih dari jutaan objek.

2. Kemampuan analisa beragam bentuk data

Algoritma klasterisasi harus mampu diimplementasikan pada berbagai macam bentuk data seperti data nominal, ordinal maupun gabungannya.

3. Menemukan *cluster* dengan bentuk yang tidak terduga

Banyak algoritma *clustering* yang menggunakan metode *Euclidean* atau *Manhattan* yang hasilnya berbentuk bulat. Padahal hasil *clustering* dapat berbentuk aneh dan tidak sama antara satu dengan yang lain. Karenanya dibutuhkan kemampuan untuk menganalisa *cluster* dengan bentuk apapun pada suatu algoritma *clustering*.

3. Kemampuan untuk dapat menangani noise

Data tidak selalu dalam keadaan baik. Ada kalanya terdapat data yang rusak, tidak dimengerti atau hilang. Karena system inilah, suatu algoritma *clustering* dituntut untuk mampu menangani data yang rusak.

4. Sensitifitas terhadap perubahan input

Perubahan atau penambahan data pada input dapat menyebabkan terjadi perubahan pada *cluster* yang telah ada bahkan bisa menyebabkan perubahan yang mencolok apabila menggunakan algoritma *clustering* yang memiliki tingkat sensitifitas rendah.

5. Mampu melakukan *clustering* untuk data dimensi tinggi

Suatu kelompok data dapat berisi banyak dimensi ataupun atribut. Untuk itu diperlukan algoritma *clustering* yang mampu menangani data dengan dimensi yang jumlahnya tidak sedikit.

6. Interpretasi dan kegunaan

Hasil dari *clustering* harus dapat diinterpretasikan dan berguna.

2. Metode Clustering

Metode Clustering merupakan salah satu metode analisis data utama untuk membantu mengidentifikasi pengelompokan objek data dari dataset. Clustering merupakan klasifikasi tanpa pengawasan dan merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma, yaitu dengan Euclidean Distance (Venkateswarlu & Raju, 2013: 9). Mempartisi dataset menjadi beberapa subset atau kelompok sedemikian serupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set properti yang dibagikan bersama, dengan tingkat similaritas yang tinggi dalam satu kelompok dan tingkat similaritas antar kelompok yang rendah. Disebut juga dengan unsupervised learning. Jika diberikan sejumlah titik data yang masing-masing mempunyai sejumlah atribut, dan dengan menggunakan satu ukuran similaritas, dapat ditemukan klaster-klaster sedemikian sehingga titik-titik data dalam satu klaster mempunyai similaritas yang lebih besar. Titik-titik data dalam klaster yang berbeda mempunyai similaritas yang kecil. Ukuran similaritas yang digunakan adalah Euclidean Distance jika atributnya continue (Hermawati, 2013: 16). Analisis cluster merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis cluster mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama. Kelompok-kelompok yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Fokus dari analisis cluster adalah

membandingkan objek berdasarkan set variabel, hal inilah yang menyebabkan para ahli mendefinisikan set variabel sebagai tahap kritis dalam analisis cluster. Set variabel cluster adalah suatu set variabel yang merepresentasikan karakteristik yang dipakai objek-objek. Menurut Han dkk. (2011: 444), analisis cluster adalah proses partisi sekumpulan objek data ke subset. Masing-masing bagian adalah cluster, sehingga objek dalam sebuah cluster mirip satu sama lain. Namun berbeda dengan objek dalam cluster lainnya. Set cluster yang dihasilkan dari analisis cluster dapat disebut sebagai pengelompokan. Dalam konteks ini, metode pengelompokan yang berbeda dapat menghasilkan Clustering yang berbeda pada set data yang sama. Partisi tidak dilakukan oleh manusia, tetapi dengan algoritma Clustering, maka, Clustering berguna dalam mengarahkan pada penemuan kelompok yang sebelumnya tidak diketahui dalam data. Metode Clustering merupakan proses untuk menemukan kelompok dalam data. Tujuannya bukan untuk memprediksi variabel kelas target, namun untuk sekedar memperoleh pengelompokan pada data. Misalnya, pelanggan dari perusahaan dapat dikelompokkan berdasarkan perilaku konsumen. Proses membagi data menjadi kelompok-kelompok yang bermakna disebut Clustering. Dalam banyak kasus tidak dapat diketahui kelompok apa yang harus dicari dan dengan demikian kelompok tersebut sulit untuk diidentifikasi. Kelompok kelompok yang diidentifikasi disebut sebagai klaster. Tugas data mining Clustering dapat digunakan dalam dua kelas yang berbeda untuk menggambarkan satu set data yang diberikan dan sebagai langkah preprocessing pada algoritma prediksi lainnya (Kotu & Deshpande, 2015: 217). Terdapat perbedaan antara metode Clustering

dan algoritma Clustering. Sebuah metode Clustering merupakan strategi umum yang diterapkan untuk memecahkan masalah Clustering. Sedangkan algoritma Clustering hanya sebuah contoh dari metode. Semua algoritma Clustering pada dasarnya dapat dikategorikan ke dalam dua kategori utama. Yaitu partisi dan hirarki. Salah satu algoritma yang termasuk ke dalam partisi adalah K-Means (Soni & Ganatra, 2012: 64).

Kekurangan dan Kelebihan K-Means

Ada beberapa kelebihan pada algoritma k-means, yaitu:

- a. Mudah untuk diimplementasikan dan dijalankan.
- b. Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat.
- c. Mudah untuk diadaptasi.
- d. Umum digunakan.

Algoritma k-means memiliki beberapa kelebihan, namun ada kekurangannya juga.

Kekurangan dari algoritma tersebut yaitu :

- a.) Sebelum algoritma dijalankan, k buah titik diinisialisasi secara random sehingga pengelompokan data yang dihasilkan dapat berbeda-beda. Jika nilai random untuk inisialisasi kurang baik, maka pengelompokan yang dihasilkan pun menjadi kurang optimal.
- b.) Dapat terjebak dalam masalah yang disebut curse of dimensionality. Hal ini dapat terjadi jika data pelatihan memiliki dimensi yang sangat tinggi (Contoh jika

data pelatihan terdiri dari 2 atribut maka dimensinya adalah 2 dimensi. Namun jika ada 20 atribut, maka akan ada 20 dimensi). Salah satu cara kerja algoritma ini adalah mencari jarak terdekat antara k buah titik dengan titik lainnya. Jika mencari jarak antar titik pada 2 dimensi, masih mudah dilakukan. Namun bagaimana mencari jarak antar titik jika terdapat 20 dimensi. Hal ini akan menjadi sulit.

c.)Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan k titik yang diinisialisasi secara random. Namun jika terdapat banyak sekali titik data (misalnya satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti kD -Tree atau hashing.

3. K-Means

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan

variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Objective function yang berusaha diminimalkan oleh k-means adalah:

$$J(U, V) = \sum_{k=1}^N \sum_{i=1}^c (a_{ik} * (x_k, v_i)^2)$$

dimana:

U : Matriks keanggotaan data ke masing-masing cluster yang berisikan nilai 0 dan 1

V : Matriks centroid/rata-rata masing-masing cluster

N : Jumlah data

c : Jumlah cluster

a_{ik} : Keanggotaan data ke-k ke cluster ke-i

x_k : data ke-k

v_i : Nilai centroid cluster ke-i

Prosedur yang digunakan dalam melakukan optimasi menggunakan k-means adalah sebagai berikut:

Step 1. Tentukan jumlah cluster

Step 2. Alokasikan data ke dalam cluster secara random

Step 3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster.

Step 4. Alokasikan masing-masing data ke centroid/rata-rata terdekat

Step 5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan

atau apabila perubahan nilai pada objective function yang digunakan, di atas nilai threshold yang ditentukan

Centroid/rata-rata dari data yang ada di masing-masing cluster yang dihitung pada Step 3. didapatkan menggunakan rumus sebagai berikut:

$$v_{ij} = \text{SUM} (k=0 \text{ to } N_i) (x_{kj}) / N_i$$

dimana:

i, k : indeks dari cluster

j : indeks dari variabel

v_{ij} : centroid/rata-rata cluster ke- i untuk variabel ke- j

x_{kj} : nilai data ke- k yang ada di dalam cluster tersebut untuk variabel ke- j

N_i : Jumlah data yang menjadi anggota cluster ke- i

Sedangkan pengalokasian data ke masing-masing cluster yang dilakukan pada Step 4. dilakukan secara penuh, dimana nilai yang memungkinkan untuk a_{ik} adalah 0 atau 1. Nilai 1 untuk data yang dialokasikan ke cluster dan nilai 0 untuk data yang dialokasikan ke cluster yang lain. Dalam menentukan apakah suatu data teralokasikan ke suatu cluster atau tidak, dapat dilakukan dengan menghitung jarak data tersebut ke masing-masing centroid/rata-rata masing-masing cluster. Dalam hal ini, a_{ik} akan bernilai 1 untuk cluster yang centroidnya terdekat dengan data tersebut, dan bernilai 0 untuk yang lainnya.

Referensi:

J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press,

1:281-297

J. A. Hartigan (1975) "Clustering Algorithms". Wiley.

J. A. Hartigan and M. A. Wong (1979). "A K-Means Clustering Algorithm".
Applied Statistics 28 (1): 100–108.

Kelebihan k-means

1. Mudah dilakukan saat pengimplementasian dan di jalankan.
2. Waktu yang di butuhkan untuk melakukan pembelajaran relatif lebih cepat.
3. Sangat fleksibel, adaptasi yang mudah untuk di lakukan
4. Sangat umum penggunaannya.
5. Menggunakan prinsip yang sederhana dapat di jelaskan dalam non-statistik.

Kekurangan dari k-means:

1. Sebelum algoritma di jalankan, titik K diinisialisasikan secara random sehingga pengelompokan data yang di dapatkan bisa berbeda-beda. Namun apabila nilai yang diperoleh acak untuk penginisialisasi kurang baik maka pengelompokan yang didapatkn menjadi tidak optimal.
2. Apabila terjebak dalam kasus yang biasanya di sebut dengan curse of dimensionality. Hal ini pun akan terjadi apabila salah satu data untuk melakukan pelatihan mempunyai dimensi yang sangat banyak, sebagai contoh; jika ada data pelatihan yang terdiri dari 2 buah atribut saja maka dimensinya ada 2 dimensi pula, namun akan berbeda jika ada 20 atribut

maka akan ada 20 dimensi yang di miliki. Adapun salah satu dari cara kerja algoritma cluster ini ialah untuk mencari jarak terdekat dari antara k titik dengan titik lainnya. Apabila ingin mencari jarak untuk antar titik dari 2 dimensi hal itu masih mudah untuk di lakukan, namun bagaimana dengan 20 buah dimensi hal tersebut akan menjadi lebih sulit untuk di lakukan pencarian jarak.

3. Apabila hanya ada terdapat beberapa buah titik sampel data yang ada, maka hal yang mudah untuk melakukan penghitungan dan mencari jarak titik terdekat dengan k titik yang telah di lakukan inialisasi yang secara acak. Namun jika ada banyak titik data, misalkan satu juta data, maka perhitungan dan pencarian titik terdekat akan sangat membutuhkan waktu yang lama. Proses tersebut dapat dipercepat namun dibutuhkan sebuah struktur data yang lebih rumit seperti kD-tree atau hashing untuk melakukan proses tersebut.
4. Adanya penggunaan k buah random, tidak ada jaminan untuk menemukan kumpulan cluster yang optimal.

3.1 Metode *K-Means*

Dari beberapa teknik yang paling sederhana dan umum dikenal adalah clustering k-means. Dalam teknik ini, objek-objek dikelompokkan ke dalam K kelompok atau cluster. Untuk melakukan clustering ini, nilai k harus ditentukan terlebih dahulu. Biasanya user atau pemakai sudah mempunyai informasi awal tentang objek yang sedang dipelajari, termasuk berapa jumlah cluster yang paling tepat.

Secara detail dapat digunakan ukuran ketidakmiripan untuk pengelompokan objek. Ketidak miripan dapat diterjemahkan dalam konsep jarak. Jika dua objek atau data titik cukup dekat, maka dua objek itu mirip. Semakin dekat semakin tinggi kemiripannya. Hasil klastering dikatakan baik jika nilai interclass similarity (kesamaan antar kelas/ WVC (withing cluster variation) tinggi dan nilai interclass similarity (kesamaan antar kelas/ BVC (between cluster variation) rendah. Nilai WVC dapat dilihat dari nilai SSE-nya sedangkan BVC dihitung dengan rumus : $BVC = d(m_1, m_2, \dots, m_n)$ dimana d adalah jarak antara centroid .

3.2. Algoritma K-Means

Pertama kali algoritma K-Means diperkenalkan oleh Mac Queen JB pada tahun 1976. Algoritma ini merupakan salah satu algoritma yang *non hierarchi* yang mana umum digunakan. Algoritma ini juga merupakan bentuk teknik penyekatan atau partisi yang membagi dan memisahkan suatu objek ke suatu daerah tertentu ke dalam daerah yang terpisah. Di dalam algoritma K-Means, pada setiap objek yang telah dimasukkan ke dalam suatu kelompok tertentu yang selanjutnya ke satu tahapan objek berikutnya yang akan berpindah ke kelompok lain.

Algoritma K-Means merupakan salah satu algoritma data *clustering* yang tidak hirarki, untuk membentuk suatu partisi data ke dalam bentuk satu atau lebih *cluster*. Algoritma ini membagi objek ke dalam *cluster*, sehingga data yang mempunyai karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan juga mempunyai karakteristik tidak memiliki kesamaan atau berbeda,

maka dikelompokkan pada kelompok lain yang mempunyai kesaamaan dengan kelompok yang berbeda itu juga. (Giyanto dan Heribertus, 2008).

Adapun proses *clustering* dimulai dengan melakukan identifikasi pada data yang akan dianalisa secara *cluster*, X_{ij} ($I = 1, \dots, n$; $j = 1, \dots, m$), dimana n adalah jumlah data yang akan di-*cluster* dan m adalah jumlah variabel. Pada iterasi awal, *centroid* pada setiap *cluster* ditetapkan secara bebas (*random*), c_{kj} ($k = 1, \dots, k$; $j = 1, \dots, m$). Selanjutnya, menghitung jarak pada data ke - I (x_i), *centroid cluster* pada data ke - k (c_k), disebut dengan (d_{ik}), sehingga formula yang digunakan adalah formula Euclidean sebagai berikut:

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

(2.1)

Dimana : d_{ik} = *centroid cluster* pada data ke - k

c_{ij} = data yang akan di-*cluster*; $k = 1, \dots, k$; $j = 1, \dots, m$.

c_{kj} = *cluster* yang dilakukan secara *random*;

$k = 1, \dots, k$; $j = 1, \dots, m$.

Di dalam suatu data yang akan menjadi anggota dari suatu *cluster* ke - k yang jika data tersebut ke *cluster centroid* pada data ke - k yang memiliki nilai paling kecil apabila dibandingkan dnegan jarak ke *cluster centroid* lainnya. Sehingga, kasus tersebut dapat dihitung dengan menggunakan persamaan berikut:

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

(2.2)

Nilai pada *cluster centroid* yang dihasilkan dapat dihitung dengan cara menghitung nilai rata-rata dari data-data tersebut yang merupakan anggota pada *cluster itu*:

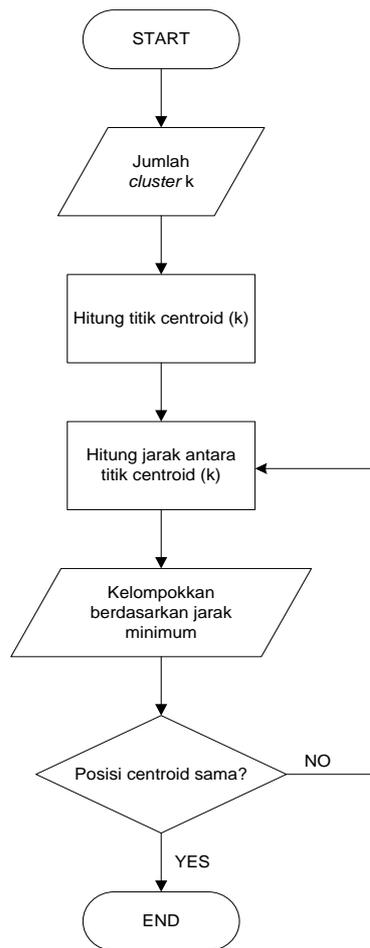
$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \dots\dots\dots$$

(2.3)

Dimana : x_{ij} = \in cluster data ke – k

p = jumlah anggota cluster data ke – k

Pada dasarnya, algoritma K-Means merupakan algoritma *clustering* yang secara umum dilakukan secara beberapa tahap adalah sebagai berikut:



Gambar 2. Algoritma K-Means Secara Umum

1. Menentukan berapa jumlah anggota *cluster* (k) yang akan dibentuk.
2. Menentukan nilai k yaitu *centroid* (titik pusat dari suatu *cluster*) awal secara *random*.
3. Menghitung jarak pada setiap data ke dalam masing-masing *centroid*.
4. Menentukan posisi *centroid* yang baru dengan menghitung nilai rata-rata dari suatu data pada *centroid* yang sama.
5. Apabila posisi *centroid* yang baru dan *centroid* sebelumnya tidak sama, maka lakukan kembali tahap 3.

Berikut adalah contoh kasus *cluster* dengan menggunakan algoritma K-Means:

Suatu kasus terdiri dari 10 data yang dikelompokkan ke dalam x dan y dengan jumlah *cluster* (K) adalah 3.

Tabel 1: Sampel Data

Data ke-i	Data Cluster	
	x	y
1	2	1
2	1	2
3	4	3
4	5	4
5	3	1
6	2	1
7	1	2
8	4	3
9	5	4
10	3	1

Langkah penyelesaian kasus dengan menggunakan algoritma K-Means adalah sebagai berikut:

1. Ada 2 cluster yang akan dibentuk.
2. Menentukan nilai k yaitu *centroid* (titik pusat dari suatu *cluster*) awal secara *random*.

Tabel 2: Data Centroid Awal

Data ke-i	Cluster	
	x	y
1	2	1
3	4	3

3. Menghitung jarak pada setiap data ke dalam masing-masing *centroid*.

a. *Cluster* pertama

$$d_{11} = \sqrt{(2 - 2)^2 + (1 - 1)^2} = 0$$

$$d_{12} = \sqrt{(1 - 4)^2 + (2 - 3)^2} = 2.83$$

Sehingga diperoleh iterasi pertama adalah sebagai berikut:

Tabel 3: Cluster Iterasi 1

Data ke-i	Data Cluster		Iterasi 1		Jarak Terpendek	Kelompok
	x	Y	C1	C2		
2	1	2	1.414214	3.162278	1.414214	C1
3	4	3	2.828427	0	0	C2
4	5	4	4.242641	1.414214	1.414214	C2
5	3	1	1	2.236068	1	C1
6	2	1	0	2.828427	0	C1
7	1	2	1.414214	3.162278	1.414214	C1
8	4	3	2.828427	0	0	C2
9	5	4	4.242641	1.414214	1.414214	C2
10	3	1	1	2.236068	1	C1

4. Menentukan posisi *centroid* yang baru dengan menghitung nilai rata-rata dari suatu data pada centroid yang sama.

$$C_{11} = \frac{(2 + 1 + 3 + 1 + 2 + 3)}{6} = 2$$

$$C_{12} = \frac{(1 + 2 + 1 + 1 + 2 + 1)}{6} = 1.33$$

$$C_{21} = \frac{(4 + 5 + 4 + 5)}{4} = 4.5$$

$$C_{22} = \frac{(3 + 4 + 3 + 4)}{6} = 3.5$$

Tabel 4: Data Centroid Iterasi 1

Data ke-i	Cluster	
	X	y
1	2	1.33
3	4.5	3.5

Untuk perhitungan jarak pada setiap masing-masing data centroid adalah sebagai berikut:

Tabel 5: Cluster Iterasi 2

Data ke-i	Data Cluster		Output		Jarak Terpendek	Kelompok
	X	y	C1	C2		
1	2	1	0.33	3.535534	0.33	C1
2	1	2	1.203703	3.807887	1.203702621	C1
3	4	3	2.605552	0.707107	0.707106781	C2
4	5	4	4.01608	0.707107	0.707106781	C2
5	3	1	1.053043	2.915476	1.053043209	C1
6	2	1	0.33	3.535534	0.33	C1
7	1	2	1.203703	3.807887	1.203702621	C1
8	4	3	2.605552	0.707107	0.707106781	C2
9	5	4	4.01608	0.707107	0.707106781	C2
10	3	1	3.162278	2.915476	2.915475947	C1

Dari penyelesaian kasus di atas, dapat dianalisa bahwa cluster dari iterasi 1 dan cluster 2 memiliki kesamaan kelompok cluster, sehingga untuk menghitung cluster selanjutnya dihentikan.

Ada beberapa kelebihan pada algoritma k-means, yaitu:

1. Mudah untuk diimplementasikan dan dijalankan.
2. Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat.
3. Mudah untuk diadaptasi.
4. Umum digunakan.

Algoritma k-means memiliki beberapa kelebihan, namun ada kekurangannya juga.

Kekurangan dari algoritma tersebut yaitu :

1. Sebelum algoritma dijalankan, k buah titik diinisialisasi secara *random* sehingga pengelompokan data yang dihasilkan dapat berbeda-beda. Jika nilai *random* untuk inisialisasi kurang baik, maka pengelompokan yang dihasilkan pun menjadi kurang optimal.
2. Dapat terjebak dalam masalah yang disebut *curse of dimensionality*. Hal ini dapat terjadi jika data pelatihan memiliki dimensi yang sangat tinggi (Contoh jika data pelatihan terdiri dari 2 atribut maka dimensinya adalah 2 dimensi. Namun jika ada 20 atribut, maka akan ada 20 dimensi). Salah satu cara kerja algoritma ini adalah mencari jarak terdekat antara k buah titik dengan titik lainnya. Jika mencari jarak antar titik pada 2 dimensi, masih mudah dilakukan. Namun bagaimana mencari jarak antar titik jika terdapat 20 dimensi. Hal ini akan menjadi sulit.
3. Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan k titik yang diinisialisasi secara *random*. Namun jika terdapat banyak sekali titik data (misalnya

satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti *kD-Tree* atau *hashing*.

BAB III

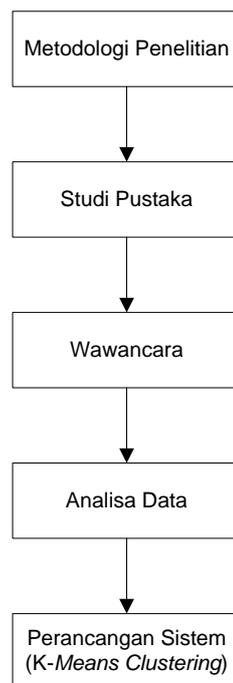
ANALISA DAN PERANCANGAN PROGRAM

1. Data yang Digunakan

Adapun data yang digunakan adalah data dari penerimaan calon pegawai dari PT Kereta Api Indonesia (Persero), dengan menggunakan sampel 30 calon pegawai, dan 5 kategori standar kelulusan.

2. Metode Penelitian

Pada penelitian ini, dalam penyelesaian masalah dilakukan beberapa metode penelitian adalah sebagai berikut:



Gambar 3. Metodologi Penelitian

2.a. Studi Pustaka

Kegiatan ini dilakukan mengacu pada buku-buku yang saling berkaitan dengan permasalahan atau kasus yang dibahas dalam tugas akhir ini dan yang mendukung beberapa kelengkapan informasi sesuai dengan yang dibutuhkan. Adapun langkah-langkah studi pustaka yang dilakukan adalah mencari gambaran atau rancangan mengenai aplikasi yang akan dibuat dan juga mendapatkan beberapa informasi, beberapa teori pendukung dan juga beberapa data yang diperlukan.

2.b. Wawancara

Selanjutnya setelah mendapatkan dan mempelajari beberapa bahan referensi yang dibutuhkan, maka proses pengumpulan data dilakukan dengan menggunakan wawancara untuk mengetahui kondisi nyata dari kasus dalam tugas akhir ini.

2.c. Analisa Data

Pada tahap ini dilakukan analisis data dari hasil wawancara yang telah dilakukan sebelumnya.. Adapun data-data yang digunakan untuk analisis dengan menggunakan metode k-means *clustering* adalah sebagai berikut:

1. Standar kelulusan calon pegawai
2. Kategori data dalam metode k-means *clustering*, yaitu LULUS dan TIDAK LULUS.

3. Perancangan Sistem

Adapun perancangan sistem yang digunakan dalam tulisan ini adalah sebagai berikut:

1. *Flowchart*
2. *Desain input dan output*

3.a. Algoritma K-Means Clustering

Adapun tahapan yang dilakukan dengan menggunakan algoritma K-Means adalah sebagai berikut:

1. Menentukan berapa jumlah anggota *cluster* (k) yang akan dibentuk.
2. Menentukan nilai k yaitu *centroid* (titik pusat dari suatu *cluster*) awal secara *random*.
3. Menghitung jarak pada setiap data ke dalam masing-masing *centroid*.
4. Menentukan posisi *centroid* yang baru dengan menghitung nilai rata-rata dari suatu data pada *centroid* yang sama.
5. Apabila posisi *centroid* yang baru dan *centroid* sebelumnya tidak sama, maka lakukan kembali tahap 3.

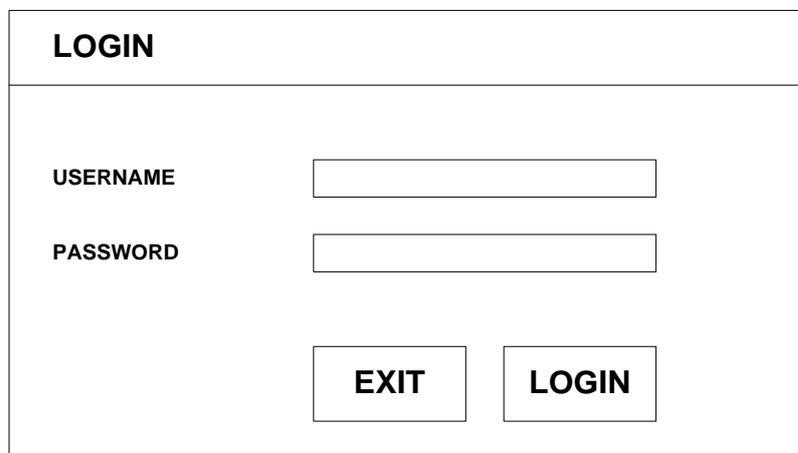
3.b. Desain Input dan Output

Adapun desain input yang dilakukan setelah mendapatkan informasi-informasi apa saja yang mengenai sistem. Untuk melakukan perancangan *interface* ini diinginkan dapat memenuhi aspek-aspek, seperti dapat

mempermudah dan sederhana, tidak harus melalui prosedur yang lama. Untuk tampilan awal interface ini, maka sebagai berikut adalah bentuk desain input dan output sistem.

1. Form Login

Tampilan form ini digunakan untuk login bagi admin dimana admin dapat mengakses form utama. Dengan melakukan proses login, maka admin dapat melakukan tugasnya masing-masing sesuai dengan hak akses yang telah diberikan.



The image shows a login form with the following elements:

- LOGIN** (Title)
- USERNAME** (Label) and an input field
- PASSWORD** (Label) and an input field
- EXIT** (Button)
- LOGIN** (Button)

Gambar 4. Form Login

2. Form Menu

Tampilan form ini digunakan untuk meng-input data tes dari calon pegawai yang telah dilakukan, seperti nomor ujian, nama, nilai tes awal yaitu administrasi berkas, kesehatan awal, psikotes, wawancara, dan kesehatan akhir, serta analisis perhitungan dengan menggunakan metode *k-means clustering*.

Gambar 5. Form Menu

4. Analisis Data Mining

Pada penelitian ini akan dicari nilai lulus dan tidak lulus dari hubungan kategori kelulusan dengan data calon pegawai. Data yang digunakan berdasarkan dari nilai tes yang dilakukan oleh PT Kereta Api Indonesia (Persero). Adapun atribut kategori kelulusan adalah sebagai berikut:

Tabel 6: Atribut Kategori Kelulusan Calon Pegawai

No	Atribut Kategori
1	Administrasi
2	Kesehatan Awal
3	Psikologi
4	Wawancara
5	Kesehatan Akhir

4.a. Nilai *Centroid* Awal

Nilai-nilai yang diperoleh dari beberapa tes yang dilakukan berdasarkan kategori yang telah ditentukan, maka diperoleh data sebagai berikut:

Tabel 7: Data Kriteria Kelulusan Calon Pegawai

Calon Pegawai	Kriteria Kelulusan				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
1	5	5	6	7	6
2	6	7	7	7	6
3	7	7	8	7	8
4	8	8	9	9	9
5	5	5	6	7	6
6	5	5	6	7	6
7	6	7	7	7	6
8	7	7	8	7	8
9	8	8	9	8	9
10	5	5	5	7	6
11	5	5	6	7	6
12	6	7	7	7	6
13	7	7	8	7	8
14	8	8	9	8	9
15	5	5	6	7	6
16	5	5	6	7	6
17	6	7	7	7	6
19	8	8	9	8	9
20	5	5	6	7	6
21	5	5	6	7	6
22	6	7	7	7	6
23	7	7	8	7	8
24	8	8	9	8	9
25	5	5	6	7	6
26	5	5	6	7	6
27	6	7	7	7	6
28	7	7	8	7	8
29	8	8	9	8	9
30	5	5	6	7	6

4.b. Perhitungan Jarak

Untuk menghitung jarak pada setiap centroid terlebih dahulu menentukan nilai k yaitu *centroid* (titik pusat dari suatu *cluster*) awal secara *random*, adalah sebagai berikut:

Tabel 8: Data Centroid Awal

Data ke-i	Cluster				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
4	8	8	9	9	9
10	5	5	5	7	6

Jumlah cluster adalah dengan dimana LULUS (C1) dan TIDAK LULUS (C2), jumlah data yang digunakan adalah 30 orang calon pegawai dan 5 jumlah atribut, yaitu administrasi, kesehatan awal, psikologi, wawancara, dan kesehatan akhir.

a. Cluster Iterasi 1

$$d_{11} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{12} = \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{13} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83 \quad d_{14} =$$

$$d_{15} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (9-9)^2 + (9-9)^2} = 0$$

$$d_{16} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{17} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{18} = \sqrt{(6-8)^2 + (5-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$\begin{aligned}
d_{18} &= \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83d_{19} = \\
&\sqrt{(8-8)^2 + (7-8)^2 + (9-9)^2 + (9-9)^2 + (9-9)^2} = 1 \quad d_{1-10} = \\
&\sqrt{(5-8)^2 + (5-8)^2 + (5-9)^2 + (7-9)^2 + (6-9)^2} = 6.86 \\
d_{1-11} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-12} &= \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69 \\
d_{1-13} &= \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83 \\
d_{1-14} &= \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1 \\
d_{1-15} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-16} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-17} &= \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69 \\
d_{1-18} &= \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.82 \\
d_{1-19} &= \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1d_{1-20} = \\
&\sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-21} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-22} &= \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69 \\
d_{1-23} &= \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.82 \\
d_{1-24} &= \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1 \\
d_{1-25} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-26} &= \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \\
d_{1-27} &= \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69
\end{aligned}$$

$$d_{1-28} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83$$

$$d_{1-29} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1$$

$$d_{1-30} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

Sehingga diperoleh iterasi pertama adalah sebagai berikut:

Tabel 9: Cluster Iterasi 1

Calon Pega-wai	Kriteria Kelulusan					C1	C2	Jrk Ter-pendek
	Adminis-trasi	Kesehatan Awal	Psi-kologi	Wawan-cara	Kesehatan Akhir			
1	5	5	6	7	6	6.323	1	1
2	6	7	7	7	6	4.690	3	3
3	7	7	8	7	8	2.828	4.583	2.828
4	8	8	9	9	9	0	6.856	0
5	5	5	6	7	6	6.325	1	1
6	5	5	6	7	6	6.325	1	1
7	6	7	7	7	6	4.690	3	3
8	7	7	8	7	8	2.828	4.583	2.828
9	8	8	9	8	9	1	6.633	1
10	5	5	5	7	6	6.856	0	0
11	5	5	6	7	6	6.325	1	1
12	6	7	7	7	6	4.690	3	3
13	7	7	8	7	8	2.828	4.583	2.828
14	8	8	9	8	9	1	6.633	1
15	5	5	6	7	6	6.325	1	1
16	5	5	6	7	6	6.325	1	1
17	6	7	7	7	6	4.690	3	3
18	7	7	8	7	8	2.828	4.583	2.828
19	8	8	9	8	9	1	6.633	1
20	5	5	6	7	6	6.325	1	1
21	5	5	6	7	6	6.325	1	1
22	6	7	7	7	6	4.690	3	3
23	7	7	8	7	8	2.828	4.583	2.828
24	8	8	9	8	9	1	6.633	1
25	5	5	6	7	6	6.325	1	1
26	5	5	6	7	6	6.325	1	1
27	6	7	7	7	6	4.690	3	3
28	7	7	8	7	8	2.828	4.583	2.828
29	8	8	9	8	9	1	6.633	1
30	5	5	6	7	6	6.325	1	1

Sehingga nilai posisi *centroid* yang baru dengan menghitung nilai rata-rata dari suatu data pada centroid yang sama, adalah sebagai berikut:

$$C_{11} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 7.273$$

$$C_{12} = \frac{(7 + 8 + 7 + 8 + 7 + 8 + 7 + 7 + 7 + 7 + 7 + 8 + 7 + 8)}{11} = 7.364$$

$$C_{13} = \frac{(7 + 8 + 7 + 8 + 7 + 7 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 8.273$$

$$C_{14} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 7.455$$

$$C_{15} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 8.182$$

Maka, diperoleh data centroid baru untuk iterasi 1 yang dilanjutkan ke pencarian data iterasi 2.

Tabel 10: Data Centroid Baru Iterasi 1

Data ke-i	Cluster				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
4	7.273	7.364	8.273	7.455	8
10	5.579	5.842	6.526	7.105	6.316

Untuk perhitungan jarak pada setiap masing-masing data centroid adalah sebagai berikut:

Tabel 11: Cluster Iterasi 2

Calon Pegawai	Kriteria Kelulusan					C1	C2	Jrk Terpendek
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir			
1	5	5	6	7	6	4.486	1.1967	1.197
2	6	7	7	7	6	2.753	1.361	1.361
3	7	7	8	7	8	1.220	2.895	1.230
4	8	8	9	9	9	2.203	5.238	2.203
5	5	5	6	7	6	4.486	1.197	1.197

6	5	5	6	7	6	4.486	1.197	1.197
7	6	7	7	7	6	2.753	1.361	1.361
8	7	7	8	7	8	0.698	2.895	0.698
9	8	8	9	8	9	1.661	4.964	1.661
10	5	5	5	7	6	5.066	1.867	1.867
11	5	5	6	7	6	4.486	1.197	1.197
12	6	7	7	7	6	2.753	1.361	1.361
13	7	7	8	7	8	0.698	2.895	0.698
14	8	8	9	8	9	1.661	4.964	1.661
15	5	5	6	7	6	4.486	1.197	1.197
16	5	5	6	7	6	4.486	1.197	1.197
17	6	7	7	7	6	2.753	1.361	1.361
18	7	7	8	7	8	0.698	2.895	0.698
19	8	8	9	8	9	1.661	4.964	1.661
20	5	5	6	7	6	4.486	1.197	1.197
21	5	5	6	7	6	4.486	1.197	1.197
22	6	7	7	7	6	2.753	1.361	1.361
23	7	7	8	7	8	0.698	2.895	0.698
24	8	8	9	8	9	1.661	4.964	1.661
25	5	5	6	7	6	4.486	1.197	1.197
26	5	5	6	7	6	4.486	1.197	1.197
27	6	7	7	7	6	2.753	1.361	1.361
28	7	7	8	7	8	0.698	2.895	0.698
29	8	8	9	8	9	1.661	4.964	1.661
30	5	5	6	7	6	4.486	1.197	1.197

Dari penyelesaian kasus di atas, dapat dianalisis bahwa *cluster* dari iterasi 1 dan cluster 2 memiliki kesamaan kelompok cluster, sehingga untuk menghitung *cluster* selanjutnya dihentikan. Dengan menggunakan algoritma K-Means untuk *clustering* penerimaan pegawai di PT Kereta Api Indonesia (Persero), adalah sebagai berikut:

Tabel 12: Hasil *Clustering* Data dari Iterasi 1 dan Iterasi 2

Calon Pegawai	Iterasi 1		Iterasi 2	
	C1	C2	C1	C2
1		1		1
2		1		1
3	1		1	
4	1		1	
5		1		1
6		1		1
7		1		1
8	1		1	
9	1			1
10		1		1
11		1		1
12		1		1
13	1		1	
14		1	1	
15		1		1
16		1		1
17	1			1
18	1		1	
19		1	1	
20		1		1
21		1		1
22		1		1
23	1		1	
24	1		1	
25		1		1
26		1		1
27		1		1
28	1		1	
29	1		1	
30		1		1

BAB IV

IMPLEMENTASI DAN ANALISA HASIL UJI COBA

Dalam bab ini dibahas mengenai penerapan atau implementasi sistem yang spesifikasi sistem dan cara menggunakan aplikasi clustering dengan menggunakan metode K-Means langkah demi langkah supaya memudahkan penggunaan aplikasi ini, serta dibahas juga mengenai hasil ujicoba yang dilakukan di dalam aplikasi ini.

1. Implementasi

1.a. Spesifikasi Sistem

Di dalam mengoperasikan aplikasi ini digunakan metode K-Means yang diperlukan di dalam sistem yang mendukungnya. Adapun sistem pendukung yang digunakan adalah perangkat keras (hardware) dan perangkat lunak (software) dimana dibahas sebagai berikut:

- a. Spesifikasi perangkat keras (hardware)
 1. PC dengan prosesor Intel core i5
 2. RAM 4GB
 3. Hard disk 500GB
 4. VGA card on board
 5. Monitor
 6. Keyboard dan mouse.

- b. Spesifikasi perangkat lunak (software)
 - 1. Sistem operasi Windows 7 Professional SP 2
 - 2. Microsoft Access 2007
 - 3. Microsoft Visual Studio C#

1.b Aplikasi Metode K-Means

Aplikasi dengan menggunakan metode K-Means akan dijelaskan tahap demi tahap adalah sebagai berikut:

- a. Form Login
- b. Form Menu

2. Hasil Uji Coba

Di dalam tahap ini dilakukan beberapa hasil uji coba terhadap sistem yang telah dirancang dan telah dibuat. Dilakukan dengan 30 calon pegawai dengan 5 kriteria penilaian dalam seleksi perekrutan calon pegawai. Adapun pengujian yang dilakukan adalah sebagai berikut:

2.a. Implementasi Algoritma *K-Means Clustering*

2.a.1. Nilai *Centroid* Awal

Nilai-nilai yang diperoleh dari beberapa tes yang dilakukan berdasarkan kategori yang telah ditentukan, maka diperoleh data sebagai berikut:

Tabel 13: Data Kriteria Kelulusan Calon Pegawai

Calon Pegawai	Kriteria Kelulusan				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
1	5	5	6	7	6
2	6	7	7	7	6
3	7	7	8	7	8
4	8	8	9	9	9
5	5	5	6	7	6
6	5	5	6	7	6
7	6	7	7	7	6
8	7	7	8	7	8
9	8	8	9	8	9
10	5	5	5	7	6
11	5	5	6	7	6
12	6	7	7	7	6
13	7	7	8	7	8
14	8	8	9	8	9
15	5	5	6	7	6
16	5	5	6	7	6
17	6	7	7	7	6
18	6	7	7	7	6
19	8	8	9	8	9
20	5	5	6	7	6
21	5	5	6	7	6
22	6	7	7	7	6
23	7	7	8	7	8
24	8	8	9	8	9
25	5	5	6	7	6
26	5	5	6	7	6
27	6	7	7	7	6
28	7	7	8	7	8
29	8	8	9	8	9
30	5	5	6	7	6

2.a.2. Perhitungan Jarak

Untuk menghitung jarak pada setiap centroid terlebih dahulu menentukan nilai k yaitu *centroid* (titik pusat dari suatu *cluster*) awal secara *random*, adalah sebagai berikut:

Tabel 14: Data Centroid Awal

Data ke-i	Cluster				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
4	8	8	9	9	9
10	5	5	5	7	6

Jumlah cluster adalah dengan dimana LULUS (C1) dan TIDAK LULUS (C2), jumlah data yang digunakan adalah 30 orang calon pegawai dan 5 jumlah atribut, yaitu administrasi, kesehatan awal, psikologi, wawancara, dan kesehatan akhir.

a. Cluster Iterasi 1

$$d_{11} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{12} = \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69 \quad d_{13} =$$

$$\sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83 \quad d_{14} =$$

$$\sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (9-9)^2 + (9-9)^2} = 0$$

$$d_{15} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \quad d_{16} =$$

$$\sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{17} = \sqrt{(6-8)^2 + (5-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{18} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83 \quad d_{19} =$$

$$\sqrt{(8-8)^2 + (7-8)^2 + (9-9)^2 + (9-9)^2 + (9-9)^2} = 1 \quad d_{1-10} =$$

$$\sqrt{(5-8)^2 + (5-8)^2 + (5-9)^2 + (7-9)^2 + (6-9)^2} = 6.86 \quad d_{1-11} =$$

$$\sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32 \quad d_{1-12} =$$

$$\sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{1-13} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83$$

$$d_{1-14} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1$$

$$d_{1-15} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-16} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-17} = \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{1-18} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.82$$

$$d_{1-19} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1$$

$$d_{1-20} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-21} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-22} = \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{1-23} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.82$$

$$d_{1-24} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1$$

$$d_{1-25} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-26} = \sqrt{(5-8)^2 + (5-8)^2 + (6-9)^2 + (7-9)^2 + (6-9)^2} = 6.32$$

$$d_{1-27} = \sqrt{(6-8)^2 + (7-8)^2 + (7-9)^2 + (7-9)^2 + (6-9)^2} = 4.69$$

$$d_{1-28} = \sqrt{(7-8)^2 + (7-8)^2 + (8-9)^2 + (7-9)^2 + (8-9)^2} = 2.83$$

$$d_{1-29} = \sqrt{(8-8)^2 + (8-8)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2} = 1$$

$$d_{1-30} = \sqrt{(5 - 8)^2 + (5 - 8)^2 + (6 - 9)^2 + (7 - 9)^2 + (6 - 9)^2} = 6.32$$

Sehingga diperoleh iterasi pertama adalah sebagai berikut:

Tabel 15: Cluster Iterasi 1

Calon Pegawai	Kriteria Kelulusan					C1	C2	Jrk Terpendek
	Adminis-trasi	Kesehatan Awal	Psi-kologi	Wawan-cara	Kesehatan Akhir			
1	5	5	6	7	6	6.323	1	1
2	6	7	7	7	6	4.690	3	3
3	7	7	8	7	8	2.828	4.583	2.828
4	8	8	9	9	9	0	6.856	0
5	5	5	6	7	6	6.325	1	1
6	5	5	6	7	6	6.325	1	1
7	6	7	7	7	6	4.690	3	3
8	7	7	8	7	8	2.828	4.583	2.828
9	8	8	9	8	9	1	6.633	1
10	5	5	5	7	6	6.856	0	0
11	5	5	6	7	6	6.325	1	1
12	6	7	7	7	6	4.690	3	3
13	7	7	8	7	8	2.828	4.583	2.828
14	8	8	9	8	9	1	6.633	1
15	5	5	6	7	6	6.325	1	1
16	5	5	6	7	6	6.325	1	1
17	6	7	7	7	6	4.690	3	3
18	7	7	8	7	8	2.828	4.583	2.828
19	8	8	9	8	9	1	6.633	1
20	5	5	6	7	6	6.325	1	1
21	5	5	6	7	6	6.325	1	1
22	6	7	7	7	6	4.690	3	3
23	7	7	8	7	8	2.828	4.583	2.828
24	8	8	9	8	9	1	6.633	1
25	5	5	6	7	6	6.325	1	1

Tabel 15: Cluster Iterasi 1 (Lanjutan)

Calon Pegawai	Kriteria Kelulusan					C1	C2	Jrk Terpendek
	Adminis-trasi	Kesehatan Awal	Psi-kologi	Wawan-cara	Kesehatan Akhir			
26	5	5	6	7	6	6.325	1	1
27	6	7	7	7	6	4.690	3	3
28	7	7	8	7	8	2.828	4.583	2.828
29	8	8	9	8	9	1	6.633	1
30	5	5	6	7	6	6.325	1	1

Sehingga nilai posisi *centroid* yang baru dengan menghitung nilai rata-rata dari suatu data pada centroid yang sama, adalah sebagai berikut:

$$C_{11} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 7.273$$

$$C_{12} = \frac{(7 + 8 + 7 + 8 + 7 + 8 + 7 + 7 + 7 + 7 + 7 + 8 + 7 + 8)}{11} = 7.364$$

$$C_{13} = \frac{(7 + 8 + 7 + 8 + 7 + 7 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 8.273$$

$$C_{14} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 7.455$$

$$C_{15} = \frac{(7 + 8 + 7 + 8 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 8 + 7 + 8)}{11} = 8.182$$

Maka, diperoleh data centroid baru untuk iterasi 1 yang dilanjutkan ke pencarian data iterasi 2.

Tabel 16: Data Centroid Baru Iterasi 1

Data ke-i	Cluster				
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir
4	7.273	7.364	8.273	7.455	8
10	5.579	5.842	6.526	7.105	6.316

Untuk perhitungan jarak pada setiap masing-masing data centroid adalah sebagai berikut:

Tabel 17: Cluster Iterasi 2

Calon Pegawai	Kriteria Kelulusan					C1	C2	Jrk Terpendek
	Administrasi	Kesehatan Awal	Psikologi	Wawancara	Kesehatan Akhir			
1	5	5	6	7	6	4.486	1.1967	1.197
2	6	7	7	7	6	2.753	1.361	1.361
3	7	7	8	7	8	1.220	2.895	1.230
4	8	8	9	9	9	2.203	5.238	2.203
5	5	5	6	7	6	4.486	1.197	1.197
6	5	5	6	7	6	4.486	1.197	1.197
7	6	7	7	7	6	2.753	1.361	1.361
8	7	7	8	7	8	0.698	2.895	0.698
9	8	8	9	8	9	1.661	4.964	1.661
10	5	5	5	7	6	5.066	1.867	1.867
11	5	5	6	7	6	4.486	1.197	1.197
12	6	7	7	7	6	2.753	1.361	1.361
13	7	7	8	7	8	0.698	2.895	0.698
14	8	8	9	8	9	1.661	4.964	1.661
15	5	5	6	7	6	4.486	1.197	1.197
16	5	5	6	7	6	4.486	1.197	1.197
17	6	7	7	7	6	2.753	1.361	1.361
18	7	7	8	7	8	0.698	2.895	0.698
19	8	8	9	8	9	1.661	4.964	1.661
20	5	5	6	7	6	4.486	1.197	1.197
21	5	5	6	7	6	4.486	1.197	1.197
22	6	7	7	7	6	2.753	1.361	1.361

Tabel 17: Cluster Iterasi 2 (Lanjutan)

Calon Pegawai	Kriteria Kelulusan					C1	C2	Jrk Terpendek
	Adminis-trasi	Kesehatan Awal	Psi-kologi	Wawan-cara	Kesehatan Akhir			
23	7	7	8	7	8	0.698	2.895	0.698
24	8	8	9	8	9	1.661	4.964	1.661
25	5	5	6	7	6	4.486	1.197	1.197
26	5	5	6	7	6	4.486	1.197	1.197
27	6	7	7	7	6	2.753	1.361	1.361
28	7	7	8	7	8	0.698	2.895	0.698
29	8	8	9	8	9	1.661	4.964	1.661
30	5	5	6	7	6	4.486	1.197	1.197

Dari penyelesaian kasus di atas, dapat dianalisis bahwa *cluster* dari iterasi 1 dan cluster 2 memiliki kesamaan kelompok cluster, sehingga untuk menghitung *cluster* selanjutnya dihentikan. Dengan menggunakan algoritma K-Means untuk *clustering* penerimaan pegawai di PT Kereta Api Indonesia (Persero), adalah sebagai berikut:

Tabel 18: Hasil *Clustering* Data dari Iterasi 1 dan Iterasi 2

Calon Pegawai	Iterasi 1		Iterasi 2	
	C1	C2	C1	C2
1		1		1
2		1		1
3	1		1	
4	1		1	
5		1		1
6		1		1
7		1		1
8	1		1	
9	1		1	
10		1		1
11		1		1
12		1		1

Tabel 18. a: Hasil *Clustering* Data dari

Iterasi 1 dan Iterasi 2 (Lanjutan)

13	1		1	
14		1	1	
15		1		1
16		1		1
17	1		1	
18	1		1	
19		1		1
20		1		1
21		1		1
22		1		1
23	1		1	
24	1		1	
25		1		1
26		1		1
27		1		1
28	1		1	
29	1		1	
30		1		1

Dari hasil uji coba di atas diperoleh perekrutan pegawai yang lulus berdasarkan kriteria penilaian kelulusan adalah sebagai berikut:

Tabel 19: Hasil Akhir Perekrutan Pegawai

Calon Pegawai	Hasil Akhir
3	LULUS
4	LULUS
8	LULUS
9	LULUS
13	LULUS
17	LULUS
18	LULUS
23	LULUS
24	LULUS
28	LULUS
29	LULUS

Penentuan *centroid* (titik pusat) pada tahap atau langkah awal dalam metode K-Means *clustering* sangat mempengaruhi pada hasil cluster seperti hasil uji coba yang berbeda juga.

2.b. Implementasi Algoritma *K-Means Clustering* dengan *Software*

2.b.1. Form Login

Untuk menggunakan aplikasi ini, maka user diharuskan terlebih dahulu memasukkan username dan password agar dapat mengakses atau mengolah data yang akan dilakukan, adalah sebagai berikut:

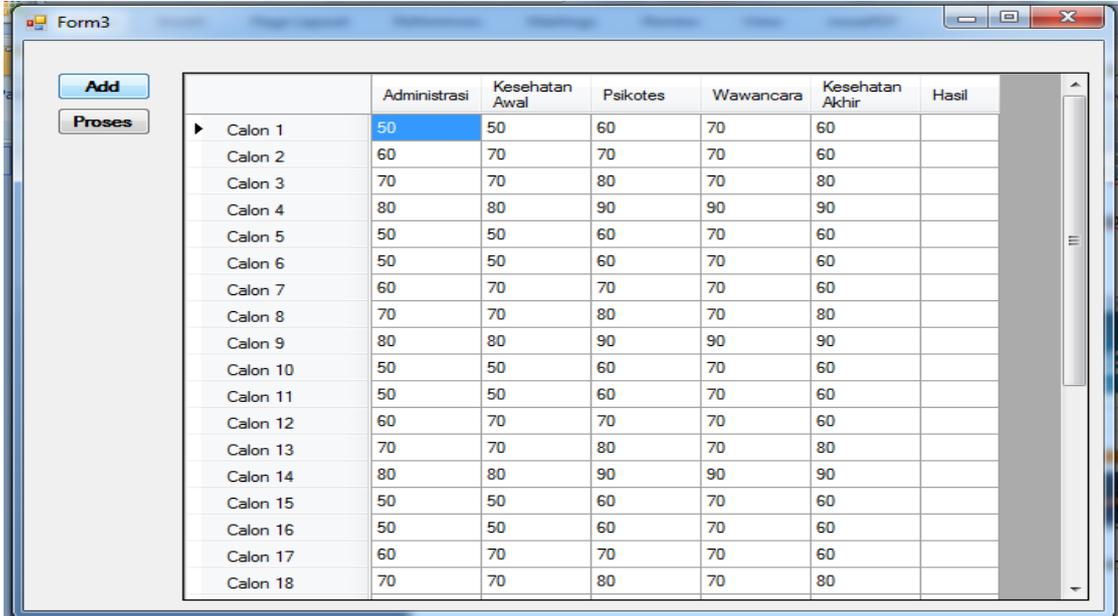


The image shows a screenshot of a login form window titled "Form1". The background is blue with the "KERETA API" logo and tagline "BY KERETA DAN MELAKUKAN PERJALANAN". The form itself is white and contains two input fields: "Username" with a user icon and "Password" with a lock icon. Below the fields is an orange "LOGIN" button.

Gambar 6. Form Login

2.b.2. Form Data Calon Pegawai

Pada form ini ditampilkan data tes atau hasil ujian sementara yang telah dilakukan oleh para calon pegawai, adalah sebagai berikut:

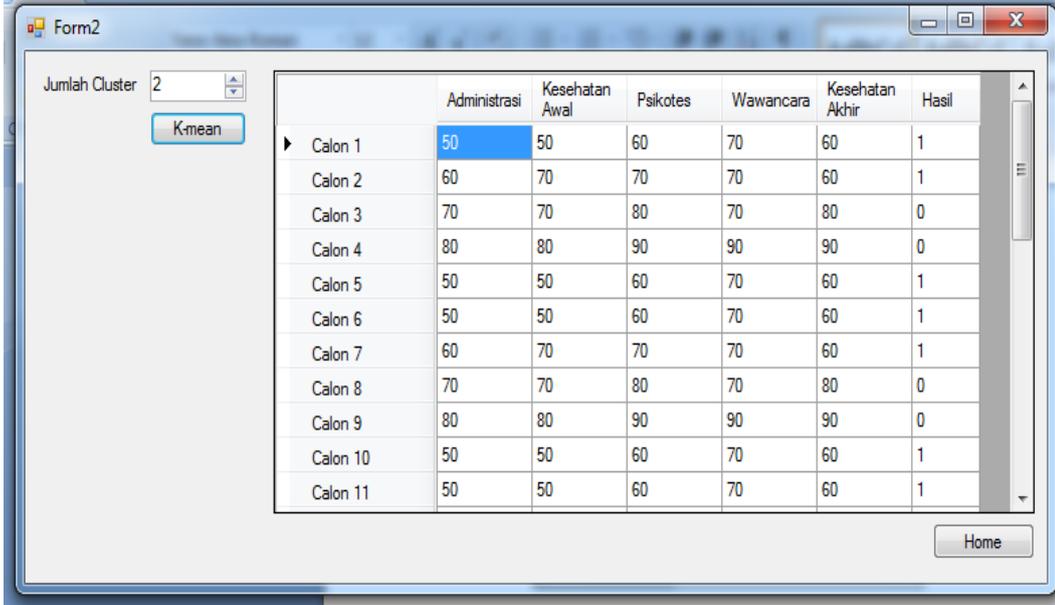


	Administrasi	Kesehatan Awal	Psikotes	Wawancara	Kesehatan Akhir	Hasil
▶ Calon 1	50	50	60	70	60	
Calon 2	60	70	70	70	60	
Calon 3	70	70	80	70	80	
Calon 4	80	80	90	90	90	
Calon 5	50	50	60	70	60	
Calon 6	50	50	60	70	60	
Calon 7	60	70	70	70	60	
Calon 8	70	70	80	70	80	
Calon 9	80	80	90	90	90	
Calon 10	50	50	60	70	60	
Calon 11	50	50	60	70	60	
Calon 12	60	70	70	70	60	
Calon 13	70	70	80	70	80	
Calon 14	80	80	90	90	90	
Calon 15	50	50	60	70	60	
Calon 16	50	50	60	70	60	
Calon 17	60	70	70	70	60	
Calon 18	70	70	80	70	80	

Gambar 7. Form Data Calon Pegawai

2.b.3. Form Hasil Clustering

Form ini merupakan hasil akhir dari proses clustering dengan menggunakan algoritma K-Means, sehingga diperoleh hasil akhir dari hasil tes yang telah dilakukan oleh para calon pegawai:



Form2

Jumlah Cluster 2

Kmean

	Administrasi	Kesehatan Awal	Psikotes	Wawancara	Kesehatan Akhir	Hasil
▶ Calon 1	50	50	60	70	60	1
Calon 2	60	70	70	70	60	1
Calon 3	70	70	80	70	80	0
Calon 4	80	80	90	90	90	0
Calon 5	50	50	60	70	60	1
Calon 6	50	50	60	70	60	1
Calon 7	60	70	70	70	60	1
Calon 8	70	70	80	70	80	0
Calon 9	80	80	90	90	90	0
Calon 10	50	50	60	70	60	1
Calon 11	50	50	60	70	60	1

Home

Gambar 8. Hasil Akhir dari Tes Calon Pegawai

BAB V

PENUTUP

Berdasarkan perancangan, implementasi dan hasil uji coba di atas, dapat ditarik kesimpulan sebagai berikut:

1. Kesimpulan

- a. Di dalam kasus ini dapat diselesaikan dengan menggunakan metode K-Means *clustering* dimana di-*cluster* berdasarkan nilai kriteria kelulusan yang telah ditentukan, yaitu dengan penentuan nilai awal titik pusat (*centroid*) dimana nilai *centroid* awal mempengaruhi nilai *centroid* berikutnya dan penentuan nilai *cluster* berikutnya, apabila *cluster* sebelumnya memiliki pola yang sama dengan pola *cluster* selanjutnya maka perhitungan dihentikan. Sehingga diperoleh hasil *cluster* akhir, tetapi sebaliknya jika berbeda dilakukan perhitungan kembali sampai batas iterasi yang telah ditentukan.
- b. Penerapan metode yang dilakukan dengan menggunakan bahasa pemrograman Visual Basic 2010 mudah diterapkan dalam kasus ini.

2. Saran

Adapun saran yang diperlukan untuk pengembangan aplikasi selanjutnya adalah dapat dilakukan pengembangan dengan jenis data atau kasus yang sama dengan menggunakan metode yang lebih baik seperti metode Fuzzy C-Means.

DAFTAR PUSTAKA

- Agusta, Y. 2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. Denpasar, Bali: Jurnal
- Akbar, A. (2018). Pembangunan Model Electronic Government Pemerintahan Desa Menuju Smart Desa. *Jurnal Teknik dan Informatika*, 5(1), 1-5.
- Budiman, I. 2012. Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. Tesis. Universitas Diponegoro.
- Dhany, H. W., Izhari, F., Fahmi, H., Tulus, M., & Sutarman, M. (2017, October). Encryption and decryption using password based encryption, MD5, and DES. In *International Conference on Public Policy, Social Computing and Development 2017 (ICOPOSDev 2017)* (pp. 278-283). Atlantis Press.
- Hariyanto, E., & Rahim, R. (2016). Arnold's cat map algorithm in digital image encryption. *International Journal of Science and Research (IJSR)*, 5(10), 1363-1365.
- Havena, M., & Marlina, L. (2018). The Technology of Corn Processing as an Effort to Increase The Income of Kelambir V Village. *Journal of Saintech Transfer*, 1(1), 27-32.
- Hendrawan, J. (2018). Rancang Bangun Aplikasi Mobile Learning Tuntunan Shalat. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 44-59.
- Huliman. 2013. Analisis Akurasi Algoritma Pohon Keputusan dan K-Nearest Neighbor (KNN).
John
- Kurnia, D. (2017). Analisis QoS Pada Pembagian Bandwidth Dengan Metode Layer 7 Protocol, PCQ, HTB Dan Hotspot Di SMK Swasta Al-Washliyah Pasar Senen. *CESS (Journal of Computer Engineering, System and Science)*, 2(2), 102-111.
- Kurnia, D., Dafitri, H., & Siahaan, A. P. U. (2017). RSA 32-bit Implementation Technique. *Int. J. Recent Trends Eng. Res*, 3(7), 279-284.
- Larose Daniel,T .2005. *Discovering knowledge in data : an introduction to data*

- Mariance, U. C. (2018). Analisa dan Perancangan Media Promosi dan Pemasaran Berbasis Web Menggunakan Work System Framework (Studi Kasus di Toko Mandiri Prabot Kota Medan). *Jurnal Ilmiah Core IT: Community Research Information Technology*, 6(1).
- Marlina, L., Muslim, M., Siahaan, A. U., & Utama, P. (2016). Data Mining Classification Comparison (Naïve Bayes and C4. 5 Algorithms). *Int. J. Eng. Trends Technol*, 38(7), 380-383.
- Marlina, L., Putera, A., Siahaan, U., Kurniawan, H., & Sulistianingsih, I. (2017). Data Compression Using Elias Delta Code. *Int. J. Recent Trends Eng. Res*, 3(8), 210-217.
- mining,
- Mirza, M. 2008. *Mengenal Diabetes Melitus*. Kata Hati. Yogyakarta.
- Nugraheni, Y. 2011. *Data Mining dengan Metode Fuzzy Untuk Customer Relationship Management (CRM) pada Perusahaan Retail*. Universitas Udayana.
- Nurjayanti B. 2011. *Identifikasi shorea menggunakan K-Nearest Neighbor berdasarkan karakteristik morfologi daun*. Skripsi. Institut Pertanian Bogor.
- Ong, J. O. 2013. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University (12):10-20*.
- Oscar Ong, J .2013. *Implementasi Algoritma k-means Clustering untuk no. 1, Menentukan Strategi Marketing President University*. *JurnalIlmiah Teknik Industri vol. 12,10-13*.
- Paulanda, Z. 2012. *Model Profil Mahasiswa Yang Potensial Drop Out Menggunakan Teknik Kernel-K-Mean Clustering Dan Decision Tree*.Tesis. Universitas Sumatera Utara. 2013.
- Putri, N. A. (2018). *Sistem Pakar untuk Mengidentifikasi Kepribadian Siswa Menggunakan Metode Certainty Factor dalam Mendukung Pendekatan Guru*. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 78-90.
- Rahim, R., Aryza, S., Wibowo, P., Harahap, A. K. Z., Suleman, A. R., Sihombing, E. E., ... & Agustina, I. (2018). *Prototype file transfer protocol application for LAN and Wi-Fi communication*. *Int. J. Eng. Technol.*, 7(2.13), 345-347.
- Rismawan, T & Kusumadewi, S. 2008. *Aplikasi K-Means Untuk Pengelompokkan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) &Ukuran Kerangka*, SNATI.Yogyakarta.

- Ruwaida, D., & Kurnia, D. (2018). Rancang Bangun File Transfer Protocol (FTP) dengan Pengamanan Open SSL pada Jaringan VPN Mikrotik di SMK Dwiwarna. *CESS (Journal of Computer Engineering, System and Science)*, 3(1), 45-49.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Graha Ilmu. Yogyakarta.
- Sarif, M. I. (2017). Penemuan Aturan yang Berkaitan dengan Pola dalam Deret Berkala (Time Series).
- Sarif, M. I. Classification Of Feasibility Of Basic Food Recipients In Kelurahan Tanjung Morawa A, Tanjung Morawa Sub-District Using Naïve Bayes Classifier Algorithm.
- Sistem dan Informatika Vol.3*, pp : 47-60.
- Soegondo, S, dkk, 2004. *Penatalaksanaan Diabetes Mellitus Terpadu*. FKUI. Jakarta.
- Soraya, Y. 2011. *Perbandingan Kinerja Metode Single Linkage, Metode Complete Linkage dan Metode K-Means dalam Analisis Cluster*. Universitas Negeri Semarang.
- Sumartono, I., Siahaan, A. P. U., & Mayasari, N. (2016). An overview of the RC4 algorithm. *IOSR J. Comput. Eng*, 18(6), 67-73.
- Tesis. Universitas Sumatera Utara.
- Wiley & Sons, Inc.