IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK ANALISA PRESTASI SISWA BERDASARKAN DATA SISWA DI SMA NEGERI 1 GROGOL

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Program Studi Teknik Informatika

OLEH :

ADHEN BAGUS PUTRO UTOMO
NPM: 12.1.03.02.0082

FAKULTAS TEKNIK (FT)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI
2016

Skripsi oleh:

ADHEN BAGUS PUTRO UTOMO
NPM: 12.1.03.02.0082

Judul:

"IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK ANALISA PRESTASI SISWA BERDASARKAN DATA SISWA DI SMA NEGERI 1 GROGOL"

Telah Disetujui untuk Diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi
Teknik Informatika Universitas Nusantara PGRI Kediri

Tanggal: 29 Juni 2016

Pembimbing I	Pembimbing II
Ir.Juli Sulaksono, M.M.,M.Kom	Patmi Kasih, M.Kom
NIDN. 0707076505	NIDN. 0701107802

ii

Skripsi oleh:

ADHEN BAGUS PUTRO UTOMO
NPM: 12.1.03.02.0082

Judul:

"IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK ANALISA PRESTASI SISWA BERDASARKAN DATA SISWA DI SMA NEGERI 1 GROGOL"

Telah dipertahankan di Depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri
Pada tanggal: 11 Agustus 2016 .

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji:

1. Ketua : Ir.Juli Sulaksono, M.M.,M.Kom ( )
2. Penguji I : Ahmad Bagus Setiawan, S.T.,M.M.,M.Kom ( )
3. Penguji II : Patmi Kasih, M.Kom ( )

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd
NIP. 19640202 199103 1 002

iii

IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK ANALISA PRESTASI SISWA BERDASARKAN DATA SISWA DI SMA NEGERI 1 GROGOL

Adhen Bagus Putro Utomo
12.1.03.02.0082
Teknik – Teknik Informatika
adhenbagus05@gmail.com
Ir. Juli Sulaksono, M.Kom dan Patmi Kasih, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Dengan kemajuan teknologi informasi sekarang ini, kebutuhan akan informasi yang akurat dalam dunia pendidikan sangat diperlukan. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, seringkali informasi tersebut masih harus digali ulang dari data yang jumlahnya besar. Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar untuk dapat mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait. Dengan diadakannya analisa prestasi siswa, diharapkan dapat diketahui bagaimana mengelompokkan prestasi siswa dan melihat kecenderungannya berdasarkan data siswa. Rumusan masalah penelitian ini adalah : (1) Bagaimana cara mengelompokkan prestasi siswa dengan metode *K-Means Clustering* ? (2) Apakah metode *K-means clustering* dapat membantu guru dalam pengelompokan siswa yang berprestasi dan kurang berprestasi?

Kesimpulan hasil penelitian ini adalah Berdasarkan pembahasan yang telah dilakukan dengan algoritma *K-Means Clustering* Adapun kesimpulan-kesimpulan sementara tersebut adalah sebagai berikut:

(1) Teknik data mining dengan algoritma *K-Means* dapat diimplementasikan pada sistem asalisis perstasi siswa di SMA. Didalam rancangan program analisis ini menggunakan *clustering* adalah salah satu bentuk upaya untuk mengetahui peranan data siswa terhadap kecenderungan prestasi siswa pada tiap kelompok. Jadi penulis melakukan penelitian dengan tujuan mengimplementasikan metode *K-means clustering* untuk analisa prestasi siwa dengan mengelompokan siswa yang berprestasi dan kurang berprestasi (2) membuktikan bahwa metode *K-means clustering* dapat digunakan untuk membantu guru dalam pengelompokan siswa yang berprestasi dan kurang berprestasi.

Kata Kunci : Prestasi Siswa, Data Siswa, Metode *K-Means Clustering.*

I. LATAR BELAKANG

Dengan kemajuan teknologi informasi sekarang ini, kebutuhan akan informasi yang akurat dalam dunia pendidikan sangat diperlukan. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, seringkali informasi tersebut masih harus digali ulang dari data yang jumlahnya besar. Kemampuan tekologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan untuk menganalisis data. Metode tradisional untuk menganalisis data yang ada tidak dapat menangani data dalam jumlah besar. Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar untuk dapat mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait.

Sekolah Menengah Atas (SMA) adalah salah satu bentuk satuan pendidikan formal yang menyelenggarakan pendidikan sebagai jenjang menuju sekolah lanjutan pada perguruan tinggi seperti Universitas. Dimana dipandang perlu adanya suatu analisa yang dapat mendukung dalam mengetahui peranan data siswa terhadap prestasi siswa.

Dengan diadakannya analisa prestasi siswa, diharapkan dapat diketahui bagaimana mengelompokkan prestasi siswa dan melihat perkembangannya dalam hal prestasi siswa dalam mata pelajaran tertentu atau dalam semua mata pelajaran berdasarkan data nilai siswa.

Berdasar uraian di atas, penulis mempunyai gagasan untuk melakukan pengelompokan terhadap siswa yang mampu dan kurang mampu dalam mata pelajaran tertentu, dengan bertujuan agar sekolah bisa memberikan bimbingan kusus kepada siswa-siwa yang kurang mampu dalam mata pelajaran tertentu.

Untuk itu penulis berkeinginan membuat aplikasi sebagai alat bantu yang bisa digunakan untuk melakukan pengelompokan siswa yang mampu dan kurang mampu dalam matapelajaran tertentu yang diwujudkan dalam bentuk Skripsi dengan judul ***"Implementasi Metode K-Means Clustering Untuk Analisa Prestasi Siswa Berdasarkan Data Siswa Di SMA Negeri 1 Grogol".***

II. METODE

A. *K-means*

K-Means adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan hanya bekerja pada atribut numerik. Dalam penerapan metode *K-means* ini ukuran jarak adalah jarak *Encludian* untuk menandai adanya persamaan antar tiap *cluster* dengan jarak minimum dan mempunyai persamaan atau kemiripan yang lebih tinggi. *Encludian*

Matrix antara titik a=(a_1,a_2,a_3,a_4.....,a_n) dan titik b=(b_1,b_2,b_3,b_4.....,b_n) adalah: $(b_i\text{-}a_i)^2$

$$d(a,b) = \sqrt{\sum_{i=l}^{n}(bi - ai)2}$$

adapun dasar rumus *Algoritma K-Means* menurut (Kusrini, 2004) adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah c*luster* yang ingin dibentuk.
2. Bangkitkan k *centroid* (titik pusat *cluster*) awal secara random.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance*.
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya.
5. Tentukan posisi *centroid* baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.
6. Kembali ke langkah no 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Berikut ini adalah data nilai siswa matapelajaran Bahasa Indonesia.

Tabel 2.1. Data Nilai

No.	Nama siswa	nilai
1	A. Rizki Muzaki	80
2	Bety Chariska A. A.	83
3	Byron Khoirul A.	75
4	Diah Palupi P.	78
5	Diky Octaviani N.	77

Tabel data nilai di atas adalah contoh tabel data nilai matapelajaran Bahasa Indonesia, Dari data nilai di atas maka akan dihitung dan diproses dengan metode *K-Means* sebagai berikut:

M1 = 8,0

M2 = 8,3

M3 = 7,5

M4 = 7,8

M5 = 7,7

K = 2

C_1 = (4,5) C_2 = (7,5)

keterangan: M1 adalah data nilai ke 1, M2 adalah data nilai ke 2, dan seterusnya. K adalah jumlah data yang akan di kelompokan, C_1 adalah kelompok data ke 1 dan C_2 adalah kelompok data ke 2.

Setelah data nilai didapati kita langsung masuk dalam perhitungan *K-Means*

Dengan rumus *Encludian Matrix* antara titik a=(a_1,a_2,a_3,a_4.....,a_n) dan titik b=(b_1,b_2,b_3,b_4.....,b_n) adalah : $(b_i\text{-}a_i)^2$

$$d(a,b) = \sqrt{\sum_{i=l}^{n}(bi - ai)2}$$

M1 Sc_1 = $\sqrt{(8-4)^2 + (0-5)^2}$ = $\sqrt{(4)^2 + (-5)^2} = \sqrt{16+25} = \sqrt{41}$

M1 Sc_2 = $\sqrt{(8-7)^2 + (0-5)^2}$ = $\sqrt{(1)^2 + (-5)^2} = \sqrt{2+25} = \sqrt{27}$

M2 Sc_1 = $\sqrt{(8-4)^2 + (3-5)^2}$ = $\sqrt{(4)^2 + (-2)^2} = \sqrt{16+4} = \sqrt{20}$

M2 Sc_2 = $\sqrt{(8-7)^2 + (3-5)^2}$ = $\sqrt{(1)^2 + (-2)^2} = \sqrt{2+4} = \sqrt{6}$

M3 Sc_1 = $\sqrt{(7-4)^2+(5-5)^2}$ = $\sqrt{(3)^2+(0)^2}$ = $\sqrt{6+0}$ = $\sqrt{6}$

M3 Sc_2 = $\sqrt{(7-7)^2+(5-5)^2}$ = $\sqrt{(0)^2+(0)^2}$ = $\sqrt{0+0}$ = 0

M4 Sc_1 = $\sqrt{(7-4)^2+(8-5)^2}$ = $\sqrt{(3)^2+(3)^2}$ = $\sqrt{9+9}$ = $\sqrt{18}$

M4 Sc_2 = $\sqrt{(7-7)^2+(8-5)^2}$ = $\sqrt{(0)^2+(3)^2}$ = $\sqrt{0+9}$ = $\sqrt{9}$

M5 Sc_1 = $\sqrt{(7-4)^2+(7-5)^2}$ = $\sqrt{(3)^2+(2)^2}$ = $\sqrt{9+4}$ = $\sqrt{13}$

M5 Sc_2 = $\sqrt{(7-7)^2+(7-5)^2}$ = $\sqrt{(0)^2+(2)^2}$ = $\sqrt{0+4}$ = $\sqrt{4}$

Dari perhitungan data diatas maka dapat disimpulkan bahwa data nilai siswa M1,M2,M3,M4, dan M5 dinyatakan masuk kelompok ke 2 atau Sc_2 karena dari hasil perhitungan diatas nilai terdekat atau terkecil seluruhnya berada dikelompok ke 2 atau Sc_2. Setelah data awal sudah di ketahui kelompoknya maka proses selanjutnya adalah penjumlahan dan pembagian seperti berikut ini:

M1 = 8,0 masuk kelompok ke 2 atau Sc_2

M2 = 8,3 masuk kelompok ke 2 atau Sc_2

M3 = 7,5 masuk kelompok ke 2 atau Sc_2

M4 = 7,8 masuk kelompok ke 2 atau Sc_2

M5 = 7,7 masuk kelompok ke 2 atau Sc_2

Semua data yang masuk kelompok ke 2 lalu dijumlah, lalu dibagi 5 karena 5 adalah jumlah banyaknya data yang masuk claster ke 2 atau Sc_2, contoh perhitungannya seperti ini.

8,0 + 8,3 + 7,5 + 7,8 + 7,7 = 39,3

39 : 5 = 7,8

3 : 5 = 0,6

Dari perhitungan di atas dapat di ketahui bahwa titik centroid awal telah berubah, dari yang tadinya titik awal centroidnya adalah Sc_1 = (4,5) dan Sc_2 = (7,5) kini telah berubah menjadi Sc_1 = (7,8) dan Sc_2 = (0,6) Karena titik awal centroid telah berubah, lakukan perhitungan lagi sampai titik centroidnya tidak berubah atau bisa juga seperti ini. Jika titik centroid awal berubah maka kita akan melakukan perhitungan yang sama dengan perhitungan yang sebelumnya tadi, akan tetapi dalam perhitungan tahap ke 2 ini kita akan menggunakan titik centroid baru yaitu Sc_1 = 7,8 dan Sc_2 = 0,6 . setelah dalam perhitungan tahap ke 2 dan hasilnya titik centroid tidak berubah maka perhitungannya selesai, dan hasil pengelompokan siswa berprestasi dan kurang berprestasi dapat di ketahui, contoh seperti berikut ini:

M1=8,0 M2=8,3 M3=7,5 M4=7,8 M5=7,7

K=2 Sc_1 =7,8 Sc_2 =0,6

M1 Sc_1 = $\sqrt{(8-7)^2+(0-8)^2}$ = $\sqrt{(1)^2+(-8)^2}$ = $\sqrt{1+64}$ = $\sqrt{65}$

M1 Sc_2 = $\sqrt{(8-0)^2+(0-6)^2}$ = $\sqrt{(8)^2+(-6)^2}$ = $\sqrt{64+36}$ = $\sqrt{100}$

M2 Sc_1 = $\sqrt{(8-7)^2+(3-8)^2}$ = $\sqrt{(1)^2+(-5)^2}$ = $\sqrt{1+25}$ = $\sqrt{26}$

M2 Sc_2 = $\sqrt{(8-0)^2+(3-6)^2}$ = $\sqrt{(8)^2+(-3)^2} = \sqrt{64+9} = \sqrt{73}$

M3 Sc_1 = $\sqrt{(7-7)^2+(5-8)^2}$ = $\sqrt{(0)^2+(-3)^2} = \sqrt{0+9} = \sqrt{9}$

M3 Sc_2 = $\sqrt{(7-0)^2+(5-6)^2}$ = $\sqrt{(7)^2+(-1)^2} = \sqrt{49+1} = \sqrt{50}$

M4 Sc_1 = $\sqrt{(7-7)^2+(8-8)^2}$ = $\sqrt{(0)^2+(0)^2} = \sqrt{0+0} = \sqrt{0}$

M4 Sc_2 = $\sqrt{(7-0)^2+(8-6)^2}$ = $\sqrt{(7)^2+(2)^2} = \sqrt{49+4} = \sqrt{53}$

M5 Sc_1 = $\sqrt{(7-7)^2+(7-8)^2}$ = $\sqrt{(0)^2+(-1)^2} = \sqrt{0+1} = \sqrt{1}$

M5 Sc_2 = $\sqrt{(7-0)^2+(7-6)^2}$ = $\sqrt{(7)^2+(1)^2} = \sqrt{49+1} = \sqrt{50}$

8,0 + 8,3 + 7,5 + 7,8 + 7,7 = 93,3

93 : 5 = 7,8

3 : 5 = 0,6

Setelah dihitung kembali menggunakan titik centroid yang baru dan ternyata hasil klusternya tetap, maka perhitungan selesai.

Tabel 2.2. pengelompokan siswa

No.	Nama siswa	nilai	Kelompok
1	A. Rizki Muzaki	80	Berprestasi
2	Bety Chariska A.	83	Berprestasi
3	Byron Khoirul A.	75	Kurang berprestasi
4	Diah Palupi P.	78	Berprestasi
5	Diky Octaviani	77	Kurang berprestasi

Tabel 2.2. pengelompokan siswa di atas adalah tabel hasil dari perhitungan menggunakan metode *K-Means* yang telah di ketahui hasilnya dan telah di kelompokan.

III. HASIL DAN KESIMPULAN

A. Kesimpulan

Dari uraian yang telah dibuat maka dapat ditarik kesimpulan bahwa:

1. Sistem pendukung keputusan analisa prestasi siswa menggunakan *Metode K-means Clustering* yang digunakan untuk analisa prestasi siswa berdasarkan data siswa di SMA Negeri 1 Grogol dapat di laksanakan, dan untuk asalisa yang sesuai dengan yang diinginkan maka perlu adanya pengisian database yang akurat. Disarankan juga menggunakan metode yang lain seperti *fuzzy Clustering* sebagai pembanding dari metode *K-means Clustering*.
2. Sistem informasi ini dapat membantu pengambilan keputusan untuk analisa prestasi siswa berdasarkan data siswa, dan hasilnya akan dikelompokkan sesuai kelompok yang diinginkan seperti kelompok siswa berprestasi dan kelompok siswa yang kurang berprestasi.

B. Saran

Saran yang dapat diberikan adalah:

1. Dalam membuat ulang sistem yang serupa diharapkan sistem informasi

yang ada dalam program lebih jelas dan lengkap.

2. Untuk menghindari kerangkapan data yang mengakibatkan inkonsistensi data sehingga dapat menyebabkan kesalahan analisis dari objek yang ditinjau maka perlu adanya pengisisan database yang akurat.
3. Mencoba membandingkan menggunakan metode lain seperti metode *Fuzzy Clustering*.

IV. DAFTAR PUSTAKA

Agus, Taufiq, dkk. 2012. *Pendidikan Anak di SD*. Tanggerang: Universitas Terbuka

Ari Muzakir, November 2014. *Analisa Pemanfaatan Algoritma K-means Clustering pada data nilai siswa sebagai penentuan penerimaan beasiswa*. Prosiding Seminar Nasional Aplikasi Sains & Teknologi *(SNAST)* 2014. Yogyakarta, 15 November 2014.

Deka, Makhfuzi, dkk. Juni 2014. *Klasterisasi judul buku dengan menggunakan metode K-means*. Seminar Nasional Aplikasi Teknoligi Informasi *(SNATI)* Yogyakarta, 21 Juni 2014.

Heri, dan Sudiyanto. Juni 2014. *"data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan, dan prestasi masalalu"*. Jurnal Pendidikan Vokasi, Vol 4 Nomor 2 Juni 2014.

Johan Oscar Ong[1]. Juni 2013. *"Implementasi algoritma K-means clustering untuk menentukan strateg marketing President University*. Jurnal Ilmiah Teknik Indrusti, Vol. 12, No.1, Juni Juni 2013.

Mahwanto, Eko. *Sistem Pendukung Keputusan Pembagian Kelas Siswa dari Pengelompokkan K-Means Clustering di SMP Negeri 1 Kepung*. Kediri : Universitas Nusantara PGRI Kediri.

Narwati. *Pengelompokkan Mahasiswa Menggunakan Algoritma K-Means.*

Oyelade, Oladipupo, Obagbuwa, 2010, *Application of K-Menas Clustering Algorithm for Prediction of Students' Acaddemic Performance, International Journal of Computer Science and Information Security*, Volume 7.

Sutrisno, Afriyudi, Widiyanto, 2013, *penerapan Data Mining pada penjualan menggunakan metode Clustering study kasus PT.Indomarco Palembang.*

Syafrianto, Andi. *Perancangan Aplikasi K-Means untuk Pengnelompokkan Mahasiswa STMIK EL-RAHMA Yogyakarta Berdasarkan Frekuensi Kunjungan ke Perpustakaan dan IPK*. Yogyakarta : STMIK EL-RAHMA.