IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DATA SISWA MENGGUNAKAN ALGORITMA K-MEANS *CLUSTERING*

(STUDI KASUS : SMKN 1 KEDIRI)

SKRIPSI

Diajukan Untuk Penulisan Skripsi Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Teknik Informatika

Oleh :

SURYANI RETNO WIDYANINGRUM
NPM: 12.1.03.02.0345

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

Skripsi oleh :

SURYANI RETNO WIDYANINGRUM

NPM: 12.1.03.02.0345

Judul :

IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DATA SISWA MENGGUNAKAN ALGORITMA K-MEANS *CLUSTERING* (STUDI KASUS : SMKN 1 KEDIRI)

Telah disetujui untuk diajukan Kepada

Panitia Ujian/Sidang Skripsi Program Studi Teknik Informatika

FT UN PGRI Kediri

Tanggal : 27 Juli 2016

Pembimbing I	Pembimbing II
Dr. Suryo Widodo, M.Pd.	Risky Aswi Ramadhani, M.Kom
NIDN. 0002026403	NIDN. 0708049001

ii

Skripsi oleh :

SURYANI RETNO WIDYANINGRUM

NPM: 12.1.03.02.0345

Judul :

IMPLEMENTASI DATA MINING PENGELOMPOKKAN DATA SISWA MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING* (STUDI KASUS : SMK N 1 KEDIRI)

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Teknik Informatika FT UN PGRI Kediri

Pada Tanggal : 12 Agustus 2016

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji : Dr.Suryo Widodo, M.Pd. _____
2. Penguji I : Fatkur Rhohman, M.Pd. _____
3. Penguji II : Risky Aswi Ramadhani, M.Kom. _____

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd.

NIP. 196402021991031002

iii

IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DATA SISWA MENGGUNAKAN ALGORITMA K-MEANS *CLUSTERING* (STUDI KASUS : SMKN 1 KEDIRI)

Suryani Retno Widyaningrum
12.1.03.02.0345
Teknik – Teknik Informatika
suryani.retno@yahoo.com
Dr. Suryo Widodo, M.Pd. dan Risky Aswi Ramadhani, M.Kom.
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Data kelas X SMKN 1 Kediri melimpah dan belum dimanfaatkan secara maksimal. Belum adanya pengelompokkan data siswa menyebabkan penumpukan data yang dapat mempengaruhi data itu sendiri. Data siswa bisa dimanfaatkan untuk mengetahui perkembangan siswa selama di sekolah.

Untuk memudahkan pencarian data siswa, data dikelompokkan menggunakan Algoritma K-Means Clustering. Algoritma K-Means merupakan algoritma teknik klastering yang berulang-ulang. Algoritma ini dimulai dengan pemilihan secara acak K, yang merupakan banyaknya klaster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat kluster atau biasa disebut dengan centroid/mean. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus yang sudah disediakan hingga diketemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasi setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut sampai nilai centroid stabil.

Hasil akhir program ini adalah terbentuknya 4 kelompok yang menunjukkan bagaimana nilai siswa selama berada di sekolah yang nantinya informasi tersebut akan digunakan untuk memantau perkembangan proses belajar siswa.

Kata Kunci : *Data Mining, Analisis Data, Clustering, K-Means, Algoritma K-Means Clustering*

I. LATAR BELAKANG

Perkembangan teknologi informasi yang semakin berkembang pesat di segala bidang kehidupan mendorong kita untuk lebih kreatif dan inovatif dalam menyelesaikan masalah yang dihadapi. Dengan adanya berbagai teknologi saat ini memudahkan kita untuk menemukan solusi yang tepat pada suatu permasalahan. Salah satunya adalah pengelompokkan data nilai siswa selama satu semester.

Data siswa SMKN 1 Kediri saat ini sudah terkumpul cukup banyak. Namun data siswa tersebut belum dimanfaatkan secara maksimal sehingga menyulitkan Kepala Sekolah dalam melakukan pemantauan terhadap perkembangan siswa di sekolah. Dengan *Algoritma K-Means Clustering* diharapkan akan diketahui kecenderungan siswa terhadap nilai pada setiap mata pelajaran, serta membantu Kepala Sekolah dan Wali Kelas dalam memantau perkembangan siswa di sekolah.

Algoritma K-Means *Clustering* merupakan salah satu metode data non-*hierarchical clustering* yang dapat mengelompokkan data ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut. Algoritma K-Means merupakan algoritma teknik *cluster* yang berulang-ulang. Algoritma ini dimulai dengan pemilihan secara acak K, yang merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara *random*, untuk sementara nilai tersebut menjadi pusat *cluster* atau biasa disebut dengan *centroid* / *mean*. Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus yang sudah disediakan hingga diketemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasi setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut sampai nilai *centroid* stabil.

Dengan *Algoritma K-Means Clustering* diharapkan dapat membantu Kepala Sekolah dan Guru terkait dalam memantau perkembangan proses pembelajaran siswa di sekolah. Pengelompokkan data siswa dilakukan dengan cara penggalian data yang meliputi beberapa variabel, yaitu : NIS, Nama dan Nilai siswa. Nilai siswa diambil dari jumlah nilai keterampilan dan pengetahuan siswa dari setiap mata pelajaran.

II. METODE PENELITIAN

Sistem pengelompokkan data siswa yang dibangun menggunakan metode K-Means Clustering yang mana metode ini menghitung jarak kedekatan data dengan centroid yang sudah di tentukan lalu di lihat seberapa dekat jarak kemiripannya.

Jumlah Cluster : 4

Jumlah Data : 10

Jumlah Atribut : 10

Tabel 2.1 Data Nilai Siswa

Nama	Nilai									
	A	P	B. I	M	S	B. Ing	Seb	Penj	Prak	Pr
ABD G.	14	13	13	11	11	10	10	11	14	12
ABD ADZIM	20	19	18	19	17	16	18	20	19	18
ACH AJI	18	17	14	17	15	15	14	13	18	16
ACH B.	16	15	16	16	13	15	14	16	18	17
ADE IRFAN	17	16	16	16	16	14	13	14	17	16
ADEK DWI	18	16	15	15	13	13	15	15	16	15
ADIT	18	17	14	17	15	14	13	14	16	15
ADITYA	19	20	17	18	19	18	17	18	19	20
AFIF P	16	17	16	15	14	13	14	16	17	16
AHMAD	10	9	10	8	9	7	8	8	9	9

Iterasi ke-1

1. Penentuan pusat awal cluster

   Nilai *centroid* awal dipilih secara acak dengan melihat bagaimana *range* dari tiap nilai pada data siswa.

Tabel 2.2 Nilai Centroid Awal C1

Mapel	Nilai
Agama	22
Pancasila	20
B. Indo	19
Mat	20
Sejarah	19
B. Ing	17
Sebud	17
Penjas	18
Prakarya	20
Produktif	19

Tabel 2.3 Nilai Centroid Awal C2

Mapel	Nilai
Agama	17
Pancasila	16
B. Indo	15
Mat	16
Sejarah	14
B. Ing	14

Tabel 2.4 Nilai Centroid Awal C2 (Lanjutan)

Mapel	Nilai
Sebud	15
Penjas	15
Prakarya	17
Produktif	16

Tabel 2.5 Nilai Centroid Awal C3

Mapel	Nilai
Agama	15
Pancasila	12
B. Indo	12
Mat	12
Sejarah	10
B. Ing	11
Sebud	11
Penjas	10
Prakarya	13
Produktif	13

Tabel 2.6 Nilai Centroid Awal C4

Mapel	Nilai
Agama	12
Pancasila	10
B. Indo	10
Mat	9
Sejarah	8
B. Ing	8
Sebud	8
Penjas	7
Prakarya	10
Produktif	8

Perhitungan jarak menggunakan Euclidean Distance :

$$d(x,y) = ||x-y||^2 = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Keterangan :

d(x,y) = jarak *Euclid*

x = data *record*

y = data *centroid*

| | = nilai *absolute*

Sebagai cintoh perhitungan jarak data terhadap pusat cluster adalah :

1. $\sqrt{(((14-22)^2)+((13-20)^2)+((13-19)^2)+((11-20)^2+(11-19)^2+(10-17)^2+(10-17)^2+(11-18)^2+(14-20)^2+(12-19)^2)} = \mathbf{22.9}$
2. $\sqrt{(((20-22)^2)+((19-20)^2)+((18-19)^2)+((19-20)^2+(17-19)^2+(16-17)^2+(18-17)^2+(20-18)^2+(19-20)^2+(18-19)^2)} = \mathbf{4.3}$
3. $\sqrt{(((18-22)^2)+((17-20)^2)+((14-19)^2)+((17-20)^2+(15-19)^2+(15-17)^2+(14-17)^2+(13-18)^2+(18-20)^2+(16-19)^2)} = \mathbf{11.2}$

Dan seterusnya dilanjutkan untuk data ke 2-N. Kemudian didapatkan hasil sebagai berikut :

Tabel 2.7 Hasil Perhitungan

Objek	C1	C2	C3	C4
X1	22.9	11.7	3.2	9.5
X2	4.3	9.5	21.2	30.1
X3	11.2	3.4	12.4	21.5
X4	12.1	3	12.3	21.4
X5	11.5	3.2	11.9	20.8
X6	13.1	2.4	10.4	19
X7	12.2	3.4	1.4	20.2
X8	4.4	10.1	21.3	30.4
X9	12.3	2.6	11.8	20.6
X10	33.1	21.7	10.7	3

Setelah didapatkan hasil perhitungan data dengan centroid awal kemudian dicari nilai rata-rata tiap anggota *cluster* yang nantinya akan digunakan untuk menentukan nilai *centroid* baru.

Tabel 2.8 Centroid Baru C1

Mapel	Nilai
Agama	19.75
Pancasila	19.5
B. Indo	17.5
Mat	18.75
Sejarah	18.75
B. Ing	17
Sebud	18.25
Penjas	19
Prakarya	19
Produktif	18.5

Tabel 2.9 Centroid Baru C2

Mapel	Nilai
Agama	17.2
Pancasila	16.2
B. Indo	15.1
Mat	16.2
Sejarah	14.1
B. Ing	14.1
Sebud	14.1
Penjas	14.8
Prakarya	17.1
Produktif	15.9

Tabel 2.10 Centroid Baru C3

Mapel	Nilai
Agama	14.3
Pancasila	13
B. Indo	12.7
Mat	11
Sejarah	10.7
B. Ing	10
Sebud	10
Penjas	10.7
Prakarya	13.7
Produktif	12

Tabel 2.11 Centroid Baru C4

Mapel	Nilai
Agama	10.3
Pancasila	8.6
B. Indo	9.6
Mat	8
Sejarah	8.6
B. Ing	7.3
Sebud	7.6
Penjas	8
Prakarya	9.6
Produktif	9

Selanjutnya dilakukan perhitungan jarak dari data dengan nilai centroid baru. Setelah dilakukan perhitungan dengan 2 kali iterasi didapatkan hasil akhir data siswa.

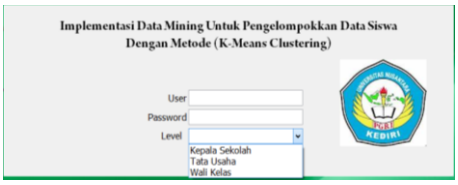
Tabel 2.12 Hasil Perhitungan

Objek	C1	C2	C3	C4
X1	21.5	11.5	0.7	10.3
X2	2.4	9.7	21.2	30.8
X3	10.1	2.9	12.9	22.5
X4	10.5	3.1	12.7	22.1
X5	10.4	2.5	11.9	21.6
X6	11.5	2.7	10.7	20.4
X7	11.1	2.7	11.6	21.2
X8	2.7	10.1	21.6	31.1
X9	10.6	2.6	11.7	21.4
X10	31.4	21.6	10.2	1.1

Tabel 2.5 berisi data siapa saja siswa yang menjadi anggorta C1, C2, C3 dan C4.

III. HASIL DAN KESIMPULAN

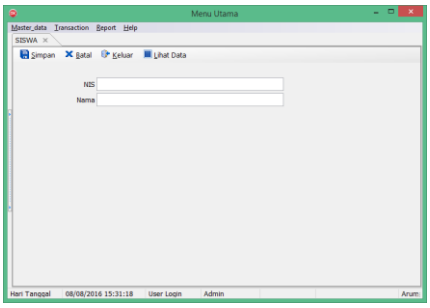
1. Tampilan *Form Login*



Gambar 3.1 *Form Login*

Halaman ini digunakan untuk keamanan data. *User* yang akan menggunakan sistem harus login terlebih dahulu dengan mengisi *username*, *password* dan *level* sesuai dengan data saat mendaftar. Penginputan data dari *form login* harus sesuai. Apabila tidak sesuai maka sistem tidak akan merespon dan tidak dapat dilanjutkan ke menu berikutnya.

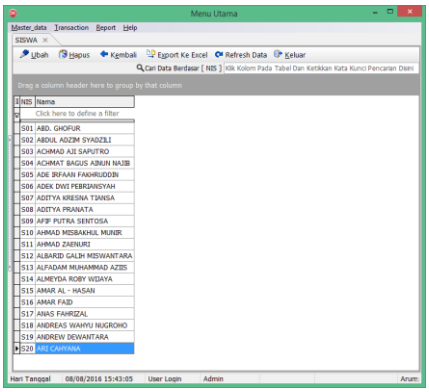
2. Tampilan *Form* Tambah Siswa



Gambar 3.2 *Form* Tambah Siswa

Terdapat 2 *textbox* dan 4 *button*. *Textbox* NIS dan Nama digunakan untuk menginputkan NIS dan Nama siswa. Terdapat *button* Simpan yang digunakan untuk menyimpan hasil inputan, *button* Batal untuk membatalkan inputan, *button* Keluar untuk keluar dari *form* dan *button* Lihat Data untuk melihat hasil data siswa yang sudah di *input*kan.

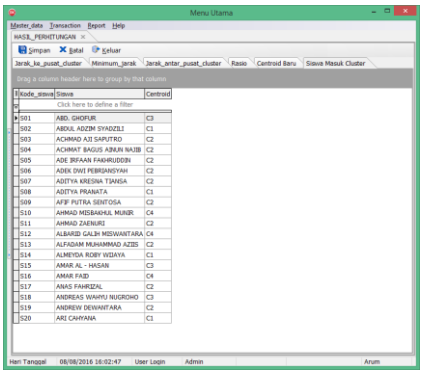
3. **Tampilan *Form* Data Siswa**



Gambar 3.3 *Form* Data Siswa

Halaman ini berisi tentang data NIS dan Nama siswa yang telah diinputkan.

4. **Tampilan Hasil Perhitungan**



Gambar 3.4 Hasil Perhitungan

Halaman ini berisi hasil perhitungan nilai siswa menggunakan *Algoritma K-Means Clustering*. Pada halaman ini dapat dilihat siswa mana saja yang masuk dalam *cluster* yang sudah ditentukan.

5. **Simpulan**
   1. Telah dihasilkan hasil akhir dari perhitungan *Algoritma K-Means Clustering* yaitu pengelompokkan nilai siswa dengan *centroid* yang sudah ditentukan. Jumlah kelompok ada 4. Tiap nilai siswa yang masuk akan dihitung kemiripannya dengan *centroid* sesuai dengan kelompoknya. Setelah itu akan dihasilkan data siswa pada setiap kelompoknya.
   2. Telah dihasilkan program aplikasi untuk mengelompokkan data siswa dengan menggunakan perhitungan *Algoritma K-Means Clustering* yang dimulai dari penentuan nilai centroid awal lalu dihitung jarak data nilai dengan data centroid awal sesuai dengan kelompoknya. Dihitung terus sampai rasio jarak perhitungan terbaru sama dengan rasio jarak perhitungan sebelumnya. Jika sudah sama maka perhitungan di hentikan dan sudah di dapatkan hasilnya.

IV. DAFTAR PUSTAKA

Agustin. 2015. *Implementasi Algoritma K-Means untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus : SMPN*

101 Jakarta). Jurnal Teknik Informatika Vol. 8 No. 1, April 2015.

Arikunto, S. 2010, *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Hasibuan, Zainal A. 2007. *Metodologi Penelitian pada Bidang Ilmu Komputer Dan Teknologi Informasi*. Diunduh 09 November 2015 pukul 13.18 WIB.

Hartanti, Ninik Tri. 2015. *Sistem Pendukung Keputusan untuk Menentukan Program Keahlian di SMK Syubbanul Wathon Magelang*. Makalah disajikan dalam Konferensi Nasional Sistem & Informatika 2015 STMIK STIKOM Bali.

Indrajani. 2015. *Database Design*. Jakarta : PT. Elex Media Komputindo.

Muzakir, Ari. 2014. *Analisa dan Pemanfaatan Algoritma K-Means Clustering pada Data Nilai Siswa sebagai Penentuan Penerima Beasiswa*. Prosiding Seminar N l Aplikasi Sains dan Tel i (SNAST).

Ong, Johan Oscar. 2013. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. Jurnal Ilmiah Teknik Industri, Vol. 12, No. 1, Juni 2013 ISSN 1412-6869.

Pahlevi. 2015. *Penerapan Metode Clustering K- Means Untuk Mengukur Tingkat Kedisplinan Siswa Di SMK Perikanan Nusantara Demak*.

Palupi, Rini Dian. *Metodologi Pengembangan Sistem Informasi*. Diunduh 26 November 2015 pukul 20.00 WIB.

Prasetyo. 2014. *Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab*. Yogyakarta : Andi.

Ramadhani. 2013. *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro.*

Sari. 2014. *Analisis Cluster Menggunakan Algoritma K-Means Untuk Mengelompokan Siswa Kelas Iv Sekolah Dasar Brawijaya Smart School Malang*.

Saputra, Agus. 2015. *Website Toko Online Dengan Smarty Php*. Cirebon : Asfa Solution.

Tampubolon, K., Saragih, H., & Reza, B. 2013. *Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan*. Makalah disajikan dalam Majalah Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI), Volume : I, Nomor : 1, Oktober 2013.