APLIKASI *K-MEANS* UNTUK PENGELOMPOKAN SEKOLAH DASAR (SD) DAN SEDERAJAT BERDASARKAN RATA-RATA HASIL UJIAN NASIONAL SEKABUPATEN JEMBER

[1] Mohamad Turmudi (11 10651036), [2] Lutfi Ali Muharom, S.Si., M.Si, [3] Hardian Oktavianto, S.Si
Jurusan Teknik Informatika Fakultas Teknik Univertas Muhammadiyah Jember
Email : mohamad.turmudi@gmail.com

Abstrak

Pendidikan sekolah dasar (SD) dan sederajatnya merupakan jenjang paling dasar pada pendidikan formal yang mempunyai peran besar bagi keberlangsungan proses pendidikan selanjutnya. Bagi siswa sekolah dasar yang ingin melanjutkan pendidikannya kejenjang yang lebih tinggi, maka siswa harus mengikuti Ujian Nasional (UN) yang dilaksanakan dalam serentak di seluruh Indonesia. Dalam setiap tahunnya setiap sekolah akan menyetorkan laporan individu tentang sekolahnya pada Dinas Pendidikan, yang juga terdapat nilai hasil ujian nasional yang terendah dan tertinggi serta nilai rata-rata dari seluruh siswa yang yang mengikuti ujian tersebut. Karena dari banyaknya jumlah sekolah dasar dan sederajatnya yang ada di Kabupaten Jember, Tentu kita akan kesulitan untuk mencari mana sekolah yang mendapat nilai ujian nasional yang bagus dari ketiga mata pelajaran yang diujikan tersebut. Dari pengujian yang dihasilkan K-Means merupakan salah satu algoritma dasar untuk membentuk cluster dan dapat diterapkan untuk Pengelompokan Sekolah Dasar (SD) Dan Sederajat Berdasarkan Rata-Rata Hasil Ujian Nasional. Pengelompokan dengan menggunakan algortima K-Means memiliki kemampuan yang cukup baik dalam klasifikasi berdasarkan rata-rata hasil ujian sekolah. Dimana dalam pendeteksiannya dipengaruhi oleh parameter-parameter input seperti *threshold*. Merancang sistem yang baru dalam menggelompokkan sekolah dapat membantu pihak instansi dalam mengklasifikasikan sekolah berdasarkan atas rata-rata nilai ujian nasional.

Kata Kunci : *Sekolah Dasar*; *Ujian Nasional; K-means*

1. PENDAHULUAN

1.1 Latar Belakang

Peran komputer semakin banyak di dalam kehidupan masyarakat. Hampir semua bidang kehidupan telah menggunakan komputer sebagai alat bantu. Diharapkan pada perkembangannya, komputer dapat langsung dirasakan manfaatnya oleh masyarakat. Salah satu golongan masyarakat yang banyak menggunakan adalah mahasiswa atau orang-orang yang berada di lingkup pendidikan.

Pendidikan sekolah dasar (SD) dan sederajatnya merupakan jenjang paling dasar pada pendidikan formal yang mempunyai peran besar bagi keberlangsungan proses pendidikan selanjutnya. Bagi siswa sekolah dasar yang ingin melanjutkan pendidikannya kejenjang yang lebih tinggi, maka siswa harus mengikuti Ujian Nasional (UN) yang dilaksanakan dalam serentak di seluruh Indonesia. Dari hasil ujian nasional tersebut akan menentukan lulus dan tidaknya siswa dari sekolah dasar. Dalam setiap tahunnya setiap sekolah akan menyetorkan laporan individu tentang sekolahnya pada Dinas Pendidikan, yang juga terdapat nilai hasil ujian nasional yang terendah dan tertinggi serta nilai rata-rata dari seluruh siswa yang yang mengikuti ujian tersebut. Karena dari banyaknya jumlah sekolah dasar dan sederajatnya yang ada di Kabupaten Jember, Tentu kita akan kesulitan untuk mencari mana sekolah yang mendapat nilai ujian nasional yang bagus dari ketiga mata pelajaran yang diujikan tersebut. Maka dari itu penulis akan membuat sebuah program yang dapat mengelompokkan mana sekolah yang mendapat nilai ujian yang bagus dan yang kurang bagus Berdasarkan nilai rata-rata hasil ujian nasional.

K-means adalah sebuah algoritma untuk mengklasifikasikan atau mengelompokkan objek-objek berdasarkan parameter tertentu ke dalam sejumlah group, sehingga dapat berjalan lebih cepat daripada hierarchical clustering (jika k kecil) dengan jumlah variable yang besar dan menghasilkan cluster yang lebih rapat, menurut Taufik Fuadi Abidin (2009).Maka dari itu, penulis tertarik untuk membuat Tugas Akhir berjudul "APLIKASI *K-MEANS* UNTUK PENGELOMPOKAN SEKOLAH DASAR (SD) DAN SEDERAJAT BERDASARKAN RATA-RATA HASIL UJIAN NASIONAL SEKABUPATEN JEMBER". Dengan adanya aplikasi ini, penulis mengharapkan bisa memudahkan proses pengelompokan sekolah berdasarkan rata-rata hasil ujian nasional.

1.2 Batasan Masalah

1. Parameter data dalam Tugas Akhir ini adalah Mata Pelajaran Matematika, Bhs.Indonesia dan IPA (Ilmu Pengetahuan Alam).
2. Proses pengelompokan berdasarkan data awal tahun 2014.
3. Metode yang digunakan *K-Means.*

1.3 Rumusan Masalah

1. Penerapan *Algoritma K-Means* terhadap *Clustering* data sekolah.

2. Bagaimana mengelompokan Sekolah Dasar dan sederajatnya berdasarkan rata-rata hasil ujian hasional.

2. TINJUAN PUSTAKA

2.1 Pengenalan *K - Means*

2.1.1 Pengertian *K – Means*

"*K-Means* merupakan algoritma yang umum digunakan untuk *clustering* dokumen. Prinsip utama *K-Means* adalah menyusun k *prototype* atau pusat massa (*centroid*) dari sekumpulan data berdimensi n" (Giyanto, Heribertus. 2008). Sebelum diterapkan proses algoritma *K-means*, dokumen akan di *preprocessing* terlebih dahulu. Kemudian dokumen direpresentasikan sebagai vektor yang memiliki *term* dengan nilai tertentu.

Munzir Umran (2009) menyatakan bahwa *K-Means* merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama.

2.1.2 Algortitma *K – Means*

Algoritma *K-means* merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k *cluster* sehingga tingkat kemiripan antar anggota dalam satu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau dapat disebut sebagai *centroid cluster* atau pusat massa (Widyawati, 2010).

Berikut rumus pengukuran jarak menurut (Milde, J. T., 2008) :

$$d_{(x,y)} = \|x-y\|^2 = \sqrt{\sum_{i=1}^{n}(xi - yi)}^{\,2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Keterangan :
d = titik *dokument*
x = data *record*
y = data *centroid*

Jarak yang terpendek antara *centroid* dengan dokumen menentukan posisi *cluster* suatu dokumen. Misalnya dokumen A mempunyai jarak yang paling pendek ke *centroid* 1 dibanding ke yang lain, maka dokumen A masuk ke group 1. Hitung kembali posisi *centroid* baru untuk tiap-tiap *centroid* ($C_{i..j}$) dengan mengambil rata-rata dokumen yang masuk pada *cluster* awal ($G_{i..j}$). Iterasi dilakukan terus hingga posisi group tidak berubah.

Adapun rumus iterasi didefinisikan sebagai berikut :

$$C(i) = \frac{x_1 + x_2 + x_{..} + x_{...}}{\sum x} \quad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

Keterangan :
x1 = nilai data *record* ke-1
x2 = nilai data *record* ke-2
$\sum x$ = jumlah data *record*

K-Means merupakan algoritma *clustering* yang bersifat *partitional* yaitu membagi himpunan objek data ke dalam sub himpunan (*cluster*) yang tidak *overlap*, sehingga setiap objek data berada tepat dalam satu *cluster*. Strategi *partitional-clustering* yang paling sering digunakan adalah berdasarkan kriteria *square error*.Secara umum, tujuan kriteria *square error* adalah untuk memperoleh partisi (jumlah *cluster* tetap) yang meminimalkan total *square error*.

2.1.3 Tahapan Algortima *K – Means*

Berikut adalah langkah-langkah atau tahapan dalam memproses algoritma *K-means* : (Larose, 2005):

1. Langkah pertama : tentukan terlebih dahulu jumlah k *cluster* yang diinginkan.
2. Langkah kedua : lakukan inisialisasi untuk menentukan pusat *cluster*.
3. Langkah ketiga : untuk tiap baris yang ada, temukan pusat *cluster* yang terdekat. Untuk menghitung *distance* atau jarak antara data dengan pusat *cluster* digunakan rumus *Distance Euclidean* :

   $$D(X_i, M_k) = \sqrt{\sum_{j=1}^{P} (X_{ij} - M_{kj})^2}$$

   ……(6)

   Dimana Xi adalah data, Mk adalah *centroid* dari tiap-tiap *cluster* dan p adalah jumlah kriteria yang ada.
4. Langkah keempat : menentukan grup berdasarkan jarak terpendek.
5. Langkah kelima : untuk tiap k *cluster*, temukan *centroid* (*means*) dari *cluster* tersebut dan *update* lokasi dari pusat *cluster* ke dalam nilai *centroid* baru. Dimana M_k adalah *mean* yang baru, N_k adalah jumlah dari pola pada *cluster* k, dan X_{jk} adalah pola nomor 1 urutan k yang menjadi anggota *cluster*.
6. Langkah keenam : ulangi langkah ketiga 3 sampai ke 5 hingga batas nilai iterasi atau nilai toleransi (selisih M lama dan baru yang diperbolehkan untuk menghentikan algoritma) yang ditentukan atau masih ada data yang berpindah.

2.1.4 Kelebihan dan Kekurangan Metode *K – Means*

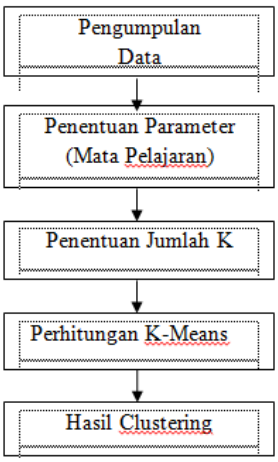
Algoritma *K-means* dinilai cukup efisien, yang ditunjukkan dengan kompleksitasnya **O (tkn),** dengan catatan **n** adalah banyaknya obyek data, **k** adalah jumlah *cluster* yang dibentuk, dan **t** banyaknya iterasi. Biasanya, nilai **k** dan **t** jauh lebih kecil dari pada nilai **n**. Selain itu, dalam iterasinya

algoritma akan berhenti dalam kondisi optimum lokal.

Hal yang dianggap sebagai kelemahan algoritma ini adalah adanya keharusan untuk menentukan banyaknya *cluster* yang akan dibentuk, dan hanya dapat digunakan dalam data yang *mean*-nya dapat ditentukan, dan tidak mampu menangani data yang mempunyai penyimpangan-penyimpangan (*noisy datan* dan *outlier*). Berkhin menyebutkan beberapa kelamahan algoritma *K-means* adalah sebagai berikut :

1. Sangat Bergantung pada pemilihan awal *Centroid.*
2. Tidak jelas berapa banyak *cluster* **k** yang terbaik.
3. Hanya bekerja pada atribut *Numerik.*

3. DESAIN SISTEM



Gambar 3.1 Desain Sistem

Keterangan :

1. Pengumpulan Data Rata-Rata Hasil Ujian.

   Data Rata-Rata Hasil Ujian di ambil dari data setiap sekolah dasar dan sederajatnya yang berada di Kabupaten Jember.
2. Penetuan Parameter

   Penentuan Parameter ini diambil dari mata pelajaran yang di ujikan dalam Ujian Nasional yaitu Matematika, Bhs.Indonesi dan IPA.
3. Penetuan Jumlah K

   Menentukan jumlah kelompok yang telah ditentukan sebelumnya, yaitu pengelompokan sekolah berdasarkan nilai rata-rata hasil ujian nasional dari setiap sekolah.
4. Perhitungan *K-Means*

   Perhitungan *K-means* dimulai dari Menentukan *centroid* awal atau titik tengah dari setiap kelompok. Cara penentuan *centroid* awal atau titik tengah pengelompokan adalah memilih salah satu data dari keseluruhan data set yang digunakan dalam penelitian ini. Dalam penelitian ini, kami membandingkan kelompok *centroid* yang berbeda untuk mengetahui hasil *clustering* yang didapat.
5. Hasil *Clustering*

   Hasil *clustering* merupakan hasil penentuan kelas, dimana setiap data akan masuk ke dalam kelas yang memenuhi persyaratan setiap kelasnya. Penentuan kelas diambil dari hasil jarak terpendek dari pusat *centroid* ke data. Hasil akan di verifikasi dengan pakar untuk mengetahui tingkat kebenaran hasil sistem dengan keadaan sebenarnya.

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

Data yang diolah pada tugas akhir ini adalah data siswa SD Negeri dan sederajat yang diperoleh dari Dinas Pendidikan Kabupaten Jember sejumlah keseluruhan data 1.405 data. Namun, pada penelitian ini, penulis hanya menggunakan 300 data untuk dijadikan sebagai data pengujian. Selengkapnya ditunjukan dalam tabel 4.1 berikut ini.

Tabel 4.1 Data Sekolah

RANGKUMAN KECAMATAN DATA SD/MI TAHUN PELAJARAN 2014/2015

Provinsi : Jawa Timur
Kabupaten/Kota : Kab. Jember
Kecamatan :
Status Sekolah : 1 & 2 (Negeri dan Swasta)

No.	Nomor Statistik Sekolah (NSS)	Nama Sekolah	Alamat	Desa/Kelurahan	Status Akreditasi Sekolah			
					A	B	C	TT
(1)	(2)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	101052412001	SD NEGERI CAKRU 01	JL. DIPONEGORO NO.14	CAKRU	0	1	0	0
2	101052412002	SDN CAKRU 02	Jalan pasar smyar Ipinggir, Cakru	CAKRU	0	1	0	0
3	101052412003	SDN CAKRU 03	Jln. Masjid Dan Gondangrejo	Cakru	1	0	0	0
4	101052412004	SD N CAKRU 04	DUSUN TEMPURAN	CAKRU	0	1	0	0
5	101052412005	SDN KENCONG 01	JL. KRAKATAU NO. 31	KENCONG	1	0	0	0
6	101052412007	SD NEGERI KENCONG 02	JL. KRAKATAU NO. 03	KENCONG	1	0	0	0
7	101052412008	SD N KENCONG 03	KH. KHOLIQ	KENCONG	0	1	0	0
8	101052412010	SD N KENCONG 04	A. YANI No. 11	KENCONG	1	0	0	0
9	101052412009	SD NEGERI KENCONG 05	PONDOK WALUH	KENCONG	1	0	0	0
10	101052412012	SDN KENCONG 06	JL Prang Agung 09 Gumukmas II Kencong	Kencong	0	1	0	0

Tabel 4.2 Data Sekolah dengan Bobot Nilai Kriteria

RANGKUMAN KECAMATAN DATA SD/MI TAHUN PELAJARAN 2014/2015

Provinsi : Jawa Timur
Kabupaten/Kota : Kab. Jember
Kecamatan :
Status Sekolah : 1 & 2 (Negeri dan Swasta)

No.	Nomor Statistik Sekolah (NSS)	Nama Sekolah	Nilai Ujian Sekolah		
			Bhs. Ind.	Mate-matika	IPA
1	101052412001	SD NEGERI CAKRU 01	7,85	8,04	9,2
2	101052412002	SDN CAKRU 02	6,93	5,42	6,43
3	101052412003	SDN CAKRU 03	7,7	6,5	7,15
4	101052412004	SD N CAKRU 04	7,4	6,87	7,5
5	101052412005	SDN KENCONG 01	7	6,5	6,5
6	101052412007	SD NEGERI KENCONG 02	8,35	8,1	7,73
7	101052412008	SD N KENCONG 03	7,31	7,15	7,48
8	101052412010	SD N KENCONG 04	7,39	8,9	8,93
9	101052412009	SD NEGERI KENCONG 05	8,12	8,46	8,63
10	101052412012	SDN KENCONG 06	7,613048478	8,266804848	8,402173913

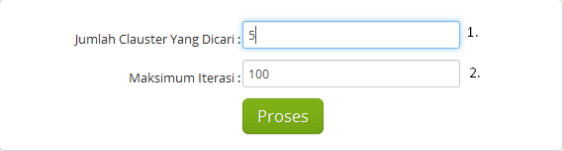
Data kriteria yang digunakan adalah kriteria yang ditentukan oleh pihak Dinas Pendidikan Kabupaten Jember. Kriteria yang ditentukan yaitu nilai rata-rata UN yang diambil dari tiga mata

pelajaran yaitu Bahasa Indonesia, Matematika, dan IPA.

4.2 Data Mining

4.2.1 Proses K-Means

K-Means adalah algoritma *clustering* yang dipilih untuk pengolahan data sehingga informasi yang dibutuhkan dapat terpenuhi. Pada tahap klasterisasi dengan menggunakan *K-Means* ini dimulai dengan pembentukan *cluster*. pembagian *cluster* ini dipilih secara random. Sehingga user tidak perlu menginputkan data secara manual. Gambar 4.1 berikut merupakan implementasi *cluster* random pada aplikasi :



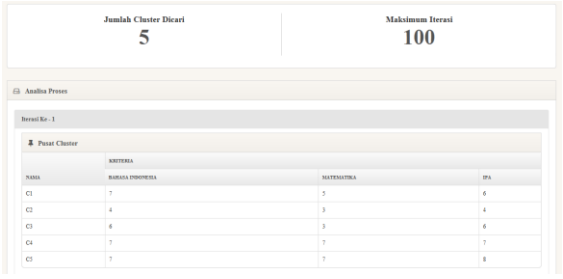
Gambar 4.1 *Inputan Cluster Random*

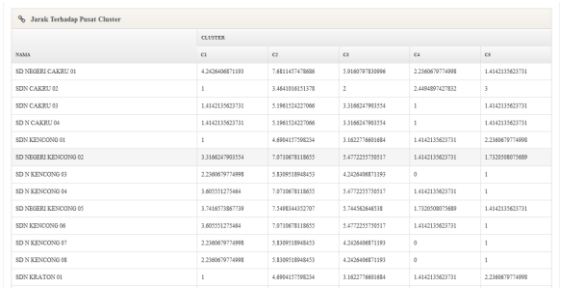
Keterangan :

1. Merupakan inputan untuk jumlah cluster yang akan dicari
2. Maksimum nilai iterasi yang akan dicari

Iterasi pada clustering ini akan berhenti, jika anggota data cluster pada iterasi sebelumnya memiliki nilai yang sama dengan nilai cluster yang dimiliki oleh iterasi berikutnya atau nilai ceteroid pada iterasi awal sama dengan nilai ceteroid selanjutnya.

Langkah selanjutnya adalah melakukan penghitungan untuk menentukan jarak setiap data dengan *centeroid* awal yang telah dibentuk dengan menggunakan rumus *squared distance (jarak)*. hasil dari penghitungan jarak dengan rurnus *squared distance* ini akan berpengaruh pada penempatan setiap data ke *cluster* tertentu. Gambar 4.2 berikut merupakan implementasi dari *squared distance* pada perhitungan k-means yang diambil dari sampel sekolah sebagai berikut :





Gambar 4.2 Implementasi Jarak pada Perhitungan *K-means*

Jumlah keseluruhan iterasi yang dimunculkan adalah 8 iterasi dengan jumlah cluster yang diinputkan adalah 5, yang diperoleh secara acak. Gambar 4.3 berikut ini merupakan jumlah keseluruhan iterasi yang dihasilkan:

Gambar 4.3 Total Iterasi Percobaan

Dari Hasil Iterasi diatas kemudian dilakukan pengelompokkan untuk setiap hasil dari keanggotaannya, berikut merupakan hasil dari pengelompokkan tersebut :

Gambar 4.4 Hasil Akhir

Pada hasil akhir yang ditunjukkan pada gambar 4.4 diatas diketahui bahwa jumlah keanggotaan dari c1 tersdiri dari 98 data/sekolah, c2 12 data/sekolah, c3 1 data/sekolah, c4 181 data/sekolah, dan c5 1 data/sekolah.

5. KESIMPULAN

Setelah melakukan analisis, perancangan, implementasi beserta pengujian, maka diperoleh kesimpulan sebagai berikut :

1. K-Means merupakan salah satu algoritma dasar untuk membentuk cluster dan dapat diterapkan untuk Pengelompokan Sekolah Dasar (SD) Dan Sederajat Berdasarkan Rata-Rata Hasil Ujian Nasional.
2. Pengelompokan dengan menggunakan algortima K-Means memiliki kemampuan yang cukup baik dalam klasifikasi berdasarkan rata-rata hasil ujian sekolah. Dimana dalam pendeteksiannya dipengaruhi oleh parameter-parameter input seperti *threshold*.

3. Merancang sistem yang baru dalam menggelompokkan sekolah dapat membantu pihak instansi dalam mengklasifikasikan sekolah berdasarkan atas rata-rata nilai ujian nasional.