

APLIKASI K-MEANS UNTUK PENGELOMPOKAN SISWA PENDIDIKAN ANAK USIA DINI (PAUD) BERDASARKAN DATA SISWA DI KECAMATAN LEDOKOMBO

Laili Amalia Shofiana (1110651055)¹, Lutfi Ali Muharom, S.Si, M.Si ²,
Bagus Setya Rintyarna, S.T, M.Kom ³, Sistem Bisnis Cerdas,
Jurusan Teknik Informatika, Fakultas Teknik,
Universitas Muhammadiyah Jember
E-mail : lailiamalia037@gmail.com ¹,

ABSTRAK

Pendidikan anak usia dini (PAUD) adalah jenjang pendidikan sebelum jenjang pendidikan dasar yang merupakan suatu upaya pembinaan yang ditujukan bagi anak sejak lahir sampai dengan usia enam tahun. Penelitian ini berusaha untuk mengelompokkan siswa berdasarkan umur dan kemampuan. Pengelompokkan data dilakukan dengan menggunakan metode *Clustering K-Means* yaitu dengan mengelompokkan data berdasarkan umur dan kemampuan (membaca, menulis, menghitung, dan menghafal). 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. Hasil dari pengujian sistem, diperoleh tiga kelompok siswa berdasarkan umur dan kemampuan yaitu, 215 siswa kelompok A, 125 siswa kelompok B, dan 120 siswa kelompok C. Uji coba terhadap 460 data dengan jumlah klaster yang sama, jika suatu klaster tingkat kemiripan yang tinggi, maka rata – rata yang dihasilkan juga memiliki tingkat kemiripan yang tinggi. Hasil perhitungan dataset dengan metode *K-Means*. Perbedaan hasil program *k-means* terjadi karena perbedaan inisialisasi centroid pertama. Metode *k-means* sangat sensitif terhadap inisialisasi centroid awal, sehingga hasil cluster yang dihasilkan berbeda.

Kata Kunci : *Clustering, K-Means, PAUD*

1. PENDAHULUAN

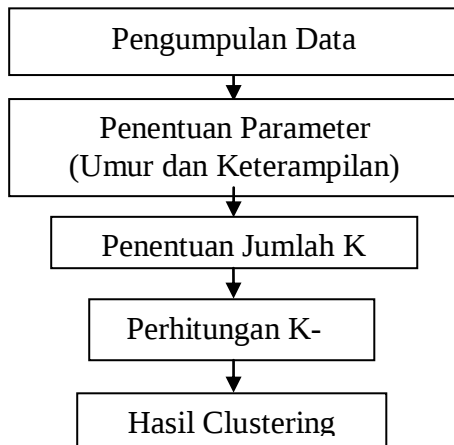
Oleh masyarakat, PAUD diidentikkan pendidikan TK. Tentu saja pendapat ini kurang tepat mengingat pendidikan TK hanya dialami anak selama satu atau dua tahun. Itu pun jika anak sempat mengalami pendidikan TK. Mengingat batasan PAUD adalah usia aktif anak sejak lahir, usia satu tahun sampai enam tahun, sekalipun praktiknya sampai pada kegiatan anak usia sepuluh tahun. Bisa dikatakan PAUD lebih banyak dilaksanakan di dalam keluarga. Dengan demikian, keluarga yang paling bertanggung jawab pada PAUD. Keluarga juga dianggap komponen terdekat dalam kehidupan anak. Kecerdasan anak tidak hanya dilihat dari segi umur saja, beberapa anak pasti ada yang memiliki kecerdasan diluar batas kewajaran. Misalnya pada anak usia 3 tahun, terkadang mereka sudah bisa menghafal al-quran sampai 30 jus. Kecerdasan seperti itulah yang wajib dikelompokkan sesuai umur

dan keterampilan. Di kecamatan Ledokombo, siswa Pendidikan Anak Usia Dini (PAUD) sangatlah banyak, kurang lebih 460 peserta didik. Setiap Kecamatan memiliki Ketua Himpunan Pendidik Anak Usia Dini Indonesia (HIMPAUDI) yang bertugas memimpin lembaga yang ada di kecamatan tersebut. Selain itu, ketua HIMPAUDI juga bertugas mendata siswa setiap ajaran baru berdasarkan kelas.

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)

2. METODOLOGI PENELITIAN

Model yang dikembangkan dalam penelitian dapat dilihat dalam gambar dibawah ini:



Gambar 3.1 Desain Sistem

1. Pengumpulan Data Siswa
Data Siswa atau data peserta didik diambil dari data Siswa PAUD yang berada di kecamatan Ledokombo.
2. Penentuan Parameter
Penentuan Parameter ini diambil dari potensi yang sangat menonjol dalam perkembangan anak yaitu umur dan keterampilan.
3. Penentuan Jumlah K
Menentukan jumlah kelompok yang telah ditentukan sebelumnya, yaitu dikelompokkan ke dalam tiga kelompok. Kelompok tersebut terdiri atas kelompok A, kelompok B, dan kelompok C.
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 dari setiap kelompok adalah memilih salah satu data dari keseluruhan data set yang digunakan dalam penelitian ini. Dalam penelitian ini, kami membandingkan 3 kelompok *centroid* yang berbeda untuk mengetahui hasil yang 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

Data yang digunakan adalah data siswa Pendidikan Anak Usia Dini (PAUD) di kecamatan Ledokombo sebanyak 460 siswa. Siswa tersebut terdiri dari 15 lembaga yang berada di Kecamatan Ledokombo.

Penentuan jumlah cluster dilakukan secara dinamis (sesuai dengan keinginan user), dibawah ini adalah tampilan penentuan jumlah *centroid* (kelas).

Langkah pertama yang dilakukan setelah penentuan jumlah klaster adalah menentukan *centroid* awal atau titik tengah dari setiap klaster dalam penentuan kelas data siswa. Cara penentuan *centroid* awal atau titik tengah dari setiap klaster adalah memilih salah satu dokumen dari keseluruhan data set yang digunakan dalam penelitian yang dipilih secara acak atau *random*.

Centroid	Cluster	Koordinat dari Centroid				
		Umur	Membaca	Menulis	Menghafal	Menghitung
Centroid 1 (M. Farel)	Kelompok A	5	70	70	70	70
Centroid 2 (Anisa R.)	Kelompok B	5	65	60	70	55
Centroid 3 (IzzatulA)	Kelompok C	5	85	80	80	70
Centroid 4 (Sadida GF)	Kelompok D	3	50	50	55	70
Centroid 5 (Moh Isrof)	Kelompok E	4	65	65	65	70

Dari tabel 4.1, peneliti menentukan jumlah klaster sebanyak lima klaster. Klaster tersebut terdiri dari klasterkelompok A, klaster kelompok B, klaster kelompok C, klaster kelompok D, klaster kelompok E. Untuk klaster kelompok A *centroid* awal yang digunakan adalah M. Farel, sedangkan klaster kelompok B *centroid* awal yang digunakan adalah Anisa R, klaster kelompok C menggunakan Izzatul A sebagai *centroid* awalnya, klaster kelompok D menggunakan Sadida GF, dan klaster kelompok E menggunakan Moh Isrof.

Melakukan proses perhitungan untuk mencari jarak terdekat dengan menggunakan rumus :

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

Proses perhitungan jarak terdekat akan terus dilakukan (sampai iterasi ke-n) dimana, iterasi akan berhenti jika tidak ada lagi perubahan nilai pada iterasi ke n-1 dan iterasi ke n

Jumlah Klaster	Jumlah Anggota	Rata-Rata				
		Umur	Membaca	Menulis	Menghitung	Menghafal
5	5.1 = 168 siswa	4,77	69,76	72,38	71,58	70,80
	5.2 = 84 siswa	4,53	61,79	61,36	64,50	57,86
	5.3 = 75 siswa	5,19	81,01	81,28	80,74	74,26
	5.4 = 33 siswa	3,80	53,00	51,00	52,17	56,50
	5.5 = 100 siswa	4,47	63,73	64,70	64,11	68,81

Uji coba terhadap jumlah klaster yang sama.

Parameter	cluster ke-	Uji Coba									
		1	2	3	4	5	6	7	8	9	10
Jumlah Cluster	1	168 siswa	128 siswa	79 siswa	64 siswa	90 siswa	36 siswa	180 siswa	104 siswa	126 siswa	62 siswa
	2	84 siswa	38 siswa	167 siswa	72 siswa	53 siswa	161 siswa	82 siswa	112 siswa	85 siswa	90 siswa
	3	75 siswa	83 siswa	58 siswa	123 siswa	136 siswa	71 siswa	57 siswa	127 siswa	29 siswa	59 siswa
	4	33 siswa	152 siswa	70 siswa	143 siswa	163 siswa	109 siswa	102 siswa	100 siswa	144 siswa	144 siswa
	5	100 siswa	59 siswa	77 siswa	111 siswa	38 siswa	29 siswa	32 siswa	17 siswa	120 siswa	115 siswa
Rata - rata umur	1	4,77	4,64	4,17	5,00	5,17	4,24	4,54	4,92	4,94	5,23
	2	4,53	4,00	5,03	4,79	4,76	4,82	4,80	4,36	5,01	4,85
	3	5,19	5,18	4,53	4,48	4,58	5,24	5,00	4,97	5,13	4,58
	4	3,8	4,49	4,48	4,79	4,49	4,38	4,53	4,39	4,34	4,31
	5	4,47	4,75	4,42	4,30	4,08	5,13	5,38	5,57	4,42	4,88
Rata - rata Membaca	1	69,76	69,06	57,85	75,97	78,47	54,26	64,44	71,29	71,73	76,50
	2	61,79	53,61	75,08	67,92	61,73	69,49	67,62	63,97	75,79	71,46
	3	81,01	78,53	68,88	64,11	68,65	81,27	81,32	74,83	87,00	69,86
	4	53,00	65,48	60,83	69,44	64,78	64,25	67,01	59,44	60,97	59,91
	5	63,73	61,96	65,13	63,24	53,65	63,75	81,96	68,57	62,28	72,50
Rata - rata Menulis	1	72,38	74,35	58,31	70,65	79,50	53,82	63,26	73,79	68,78	78,92
	2	61,36	52,22	75,48	75,69	66,75	72,46	71,38	69,99	78,79	63,54
	3	81,28	79,02	62,14	65,39	72,74	78,90	80,15	76,66	87,65	68,33
	4	51,00	64,70	66,33	70,00	61,84	63,95	72,34	65,79	67,58	62,90
	5	64,70	65,49	70,64	61,59	52,97	62,50	84,64	60,00	59,07	74,65
Rata - rata Menghitung	1	71,58	72,41	58,43	73,31	79,90	54,12	66,97	75,08	69,83	80,50
	2	64,50	53,19	75,35	70,74	71,12	71,63	69,31	63,68	76,88	67,32
	3	80,74	77,55	63,65	69,56	67,99	81,02	80,15	74,97	81,00	69,35
	4	52,17	64,81	67,89	72,24	64,56	64,75	64,46	61,63	65,19	61,51
	5	64,11	70,78	66,97	58,41	53,92	69,38	83,75	81,43	63,25	74,08
Rata - rata Menghafal	1	70,80	68,71	56,69	77,58	73,27	51,32	64,40	64,44	68,83	80,00
	2	57,86	54,17	71,70	74,90	77,14	74,55	78,54	65,86	71,95	74,02
	3	74,26	76,03	71,22	69,83	68,36	70,25	69,71	75,13	76,00	60,65
	4	56,50	63,49	70,89	62,41	64,19	64,39	65,22	61,18	69,95	64,06
	5	68,81	76,47	61,58	58,85	53,51	55,00	80,71	75,00	60,89	69,82

Perbedaan hasil program *k-means* terhadap jumlah klaster yang sama terjadi karena perbedaan inisialisasi centroid pertama. Metode *k-means* sangat sensitif terhadap inisialisasi centroid awal, sehingga hasil cluster yang dihasilkan berbeda.

5. KESIMPULAN DAN SARAN

a Kesimpulan

1. Program yang telah dibangun dapat digunakan untuk mengelompokkan siswa berdasarkan umur dan kemampuan (membaca, menulis, menghitung, dan menghafal). Program hanya dapat digunakan untuk data numerik.
2. Penggunaan metode *K-means* dapat mempermudah pengelompokan data, terutama data dalam jumlah banyak.
3. Data siswa dikelompokkan sesuai jumlah kelas yang diinginkan pengguna, jumlah kelas dapat berubah sesuai dengan keinginan pengguna.
4. Hasil perhitungan dataset dengan metode *K-Means*. Perbedaan hasil program *k-means* terjadi karena perbedaan inisialisasi centroid pertama. Metode *k-means* sangat sensitif terhadap inisialisasi centroid awal, sehingga hasil cluster yang dihasilkan berbeda.

b. Saran

Beberapa saran yang mungkin ingin mengembangkan Tugas Akhir ini, sarannya adalah :

Untuk peneliti selanjutnya dapat menggunakan data lainnya dan bisa juga jika dijadikan Sistem Informasi sebuah instansi yang membutuhkan.

REFERENSI

- [1] Septioko, Tinus. 2012. Aplikasi K-Means Untuk Pengelompokan Rumah Tangga Di Salatiga Berdasarkan Data SUSENAS 2011.
- [2] Febrian, J. dan F. Andayani. 2002. Kamus komputer dan istilah teknologi informasi. Bandung : Informatika
- [3] Giyanto, Heribertus. 2008. Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada. .
- [4] Prasetyo, Resardi Esa. 2013. Penggunaan Metode K – Means Sebagai Alat Bantu Temu Kembali Informasi Di Perpustakaan Universitas Muhammadiyah Jember.
- [5] Agustina, Silvi, Dhimas Yudho, Hadi Santoso, Nofia diarn asusanto, Arif Tirtana, dan Fakhri Khusnu. Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means