

## Klasterisasi Data Untuk Sistem Rekomendasi Pemilihan Sekolah Dasar di Wilayah Surakarta Dengan Metode *Hierarchy Clustering*

Siti Rokhmah<sup>1\*)</sup>, Moch Bagoes Pakarti<sup>2)</sup>

<sup>1\*)</sup> Informatika, Institut Teknologi Bisnis AAS Indonesia

<sup>2)</sup> Informatika, Institut Teknologi Bisnis AAS Indonesia

<sup>1\*)</sup> [sitirokhmah@itbaas.ac.id](mailto:sitirokhmah@itbaas.ac.id), <sup>2)</sup> [mobagoes@gmail.com](mailto:mobagoes@gmail.com)

### ABSTRACT

*One of the key public health service facilities is Public Health Center (Puskesmas). For new residents, it is very important to have information about the location of the puskesmas or hospital as well as information about the shortest path to the puskesmas or hospital. The size of the region and the volume of traffic present a challenge for people who are unfamiliar with the Karanganyar Regency's routes. The aim of this research is to enable the community to easily locate the health center and to provide access for numerous individuals who require assistance in locating the health center in Karanganyar Regency. Floyd-Warshall research methodology was used in this study to create this system while data collection from Karanganyar District Health Office is then required in this study. The results of this study are in the form of the closest route that is connected to Google Maps, which is information about the location of the health center in Karanganyar Regency which is beneficial for people outside Karanganyar Regency. This study's test findings, which used Floyd Warshall's results approach, further ensure its success in identifying the closest path. The test scenario and the available observations matched 91.6% of the 12 test components, according to the system functionality test. Based on the validity test, it has an average error in predicting the distance from the starting point to the search location of 98%.*

**Keywords** : *Floyd Warshall algorithm, geographic information systems, public health centers.*

### I. PENDAHULUAN

Pendidikan merupakan hak bagi setiap orang. Karena pentingnya pendidikan, pendidikan mendapatkan perhatian yang besar dari negara. Pendidikan formal paling awal umumnya diawali dengan pendidikan sekolah dasar antara umur 6-12 tahun. Dalam pemilihan sekolah dasar bagi anak orang tua memiliki peran yang besar, karena usia anak sekolah masih tergolong usia anak-anak sehingga membutuhkan arahan dari orang tua untuk menentukan sekolah yang tepat sesuai karakter dan perkembangan anak (Islam et al., 2022),(Yuliana & Aini, 2018). Selain sekolah dasar negeri, Berbagai lembaga kini banyak yang menyelenggarakan sekolah dasar swasta yang menawarkan berbagai metode dan fasilitas pembelajaran. Penyelenggara pendidikan berlomba-lomba menawarkan berbagai layanan agar dapat menarik minat para orang tua calon siswa (Atmojo et al., 2021). Beberapa faktor yang menjadi dasar dalam pemilihan sekolah dasar bagi orang tua diantaranya adalah model pembelajaran, biaya, fasilitas sekolah, budaya sekolah, lokasi sekolah, prestasi sekolah dan keamanan serta fasilitas lingkungan sekolah (Khasanah, 2012).

Berdasarkan data badan statistik surakarta tahun 2020, jumlah penduduk usia sekolah dasar adalah 45.170 anak, sedangkan jumlah sekolah dasar swasta dan negeri dibawah kementerian pendidikan dan kebudayaan pada tahun 2020 sebanyak 250 sekolah dan *Madrasah Ibtidaiyah* (MI) dibawah kementerian pendidikan dan kebudayaan sebanyak 9 sekolah (Surakarta, 2020). Salah satu permasalahan yang dihadapi para orang tua dalam menentukan pemilihan sekolah dasar adalah terdapatnya banyak pilihan sekolah yang membuat orang tua calon siswa bingung dalam menentukan pilihan, sehingga dibutuhkan

sebuah sistem rekomendasi yang dapat membantu mempermudah orang tua dalam menentukan sekolah dasar untuk anak-anaknya (Khasbulloh, 2022).

Tujuan dari penelitian ini mengembangkan konsep klasterisasi data berbasis data mining untuk sistem rekomendasi sekolah dasar di kota surakarta, sehingga dapat membantu orang tua dalam menentukan sekolah dasar bagi anaknya. Penelitian ini dibatasi pada sistem rekomendasi SD di wilayah surakarta, data sekolah dasar berdasarkan faktor-faktor yang mendasari pemilihan sekolah dasar oleh orang tua, data-data tersebut diolah dengan metode klasterisasi berbasis *hierarchy* sehingga menghasilkan rekomendasi pemilihan sekolah dasar di wilayah Surakarta.

## II. TINJAUAN PUSTAKA

Beberapa penelitian yang melatar belakangi penelitian ini diantaranya adalah penelitian dengan judul faktor-faktor yang mempengaruhi pemilihan sekolah dasar dalam mendukung kota surakarta layak anak (Saputri et al., 2018). Pada penelitian tersebut dikemukakan faktor-faktor yang menjadi pertimbangan orang tua dalam menentukan pemilihan sekolah dasar untuk anak-anak. Penelitian lain adalah penelitian dengan judul *Review of Clustering-Based Recommender Systems*. Dalam penelitian tersebut diuraikan pendekatan untuk desain sistem rekomendasi menggunakan sistem klasterisasi. Pada penelitian ini dilakukan *review* terhadap metode pendekatan berbasis konten dan metode klasterisasi pada sistem rekomendasi. Selain itu ada penelitian dengan judul *An Efficient Recommender System using Hierarchy clustering algorithm* yang membahas tentang efektifitas metode *clustering* dalam sistem rekomendasi dengan menganalisis algoritma baru untuk mengembangkan metode rekomendasi yang populer yaitu metode *collaborative filtering* dengan metode yang memanfaatkan data mining sebagai pendekatan pada sistem rekomendasi.

Berdasarkan penelitian tersebut maka dikembangkan penelitian dengan mengembangkan klasterisasi data dengan metode *Hierarchy clustering* untuk sistem rekomendasi pemilihan sekolah dasar di wilayah Surakarta. Penelitian ini memiliki beberapa perbedaan dari penelitian sebelumnya yaitu pemanfaatan data mining dengan klasterisasi data untuk sistem rekomendasi dan penggunaan algoritma *Hierarchy Clustering* untuk pengelompokan data yang dijadikan dasar dalam sistem rekomendasi.

## III. METODE PENELITIAN

Ada beberapa tahapan dalam penelitian ini, yaitu

### 1. Studi pustaka

Dalam penelitian ini mempelajari referensi berupa jurnal, buku maupun artikel lain yang terkait dengan penelitian. Adapun jurnal yang dijadikan referensi adalah jurnal yang membahas tentang Data mining, klasterisasi data, sistem rekomendasi, *Hierarchy Clustering*, dan faktor yang mempengaruhi pemilihan sekolah dasar. Selain itu juga buku dan artikel tentang klasterisasi dan metode *Hierarchy Clustering*.

### 2. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data sekolah dasar di wilayah surakarta dan data kuisioner faktor yang mempengaruhi pemilihan sekolah dasar di wilayah surakarta. Data SD di wilayah surakarta diperoleh dari data di kementerian pendidikan dan hasil observasi ke SD di wilayah surakarta, sedangkan data faktor yang mempengaruhi pemilihan SD di wilayah di surakarta didapatkan dari data kuisioner yang disebarkan ke orang tua yang memiliki anak yang bersekolah di wilayah surakarta. Selain itu juga dilakukan wawancara kepada pihak sekolah terkait kondisi sekolah dan melakukan observasi secara langsung ke sekolah di wilayah Surakarta.

3. Pra Pemrosesan Data

Pra pemrosesan data dilakukan untuk mempermudah memahami data sehingga mempermudah dalam pemilihan metode data mining, selain itu pra pemrosesan juga dapat meningkatkan kualitas data dan efisiensi pemrosesan penambangan data (Edy Irwansyah, 2015b). Pemrosesan data meliputi pembersihan data dan seleksi atribut.

4. Klasterisasi data dengan metode *Hierarchy Clustering*

Klasterisasi adalah suatu proses mengelompokkan sejumlah data menjadi beberapa kelompok sehingga object dapat dikelompokkan berdasarkan kecenderungan kesamaan dan perbedaan. Istilah lain untuk klasterisasi adalah *unsupervised classification* karena klastering bersifat dipelajari dengan memperhatikan object (Pratikto & Damastuti, 2021). Metode ini membangun klaster dengan mempartisi data secara rekursif baik dengan menggunakan urutan dari data atas ke bawah ataupun bawah ke atas (Fadliana & Rozi, 2015), (Nisha & Kaur, 2016).

1. Jarak antar obyek

Untuk mengukur jarak antar obyek menggunakan rumus euclidean distance. *Euclidean distance* merupakan salah satu metode berbasis *geometris* (Chandan K. Reddy, 2018). *Euclidean distance* mengukur jarak terdekat dari satu titik ke titik lainnya dengan menggunakan persamaan 1.

$$dist(p, q) = \sqrt{\sum_1^m (p - q)^2} \quad (1)$$

Dimana p dan q merupakan titik yang akan dicari jaraknya. Rentang *similarity* pada euclidean distance adalah antara 0 hingga 1. 0 menyatakan bahwa obyek tidak memiliki kemiripan dan 1 jika obyek tersebut identik.

2. Pengelompokan dengan *Hierarchy Clustering*

Pada penelitian menggunakan metode *agglomerative clustering*, alasan penggunaan metode ini karena metode ini efisien digunakan dalam pengelompokan (Dewi & Ahadiyah, 2022). Untuk menentukan titik-titik dalam sebuah klaster menggunakan algoritma pada *agglomerative clustering*, yaitu

a. *Single linkage*

*Single linkage* adalah pengelompokan *agglomerative* berdasarkan jarak terkecil antar obyek. Algoritma *Single linkage* dapat dihitung dengan persamaan 2

$$d_{(uv)w} = \min\{d_{uw}, d_{vw}\} \quad (2)$$

Dimana  $d_{uw}$  adalah jarak terdekat antara u dan w, dan  $d_{vw}$  merupakan jarak terdekat antara v dan w.

b. *Complete linkage*

*Complete linkage* adalah pengelompokan klaster berdasarkan jarak terjauh, atau yang memiliki kemiripan terkecil. Algoritma *Complete linkage* dapat dihitung dengan persamaan 3,

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\} \quad (3)$$

c. *Average linkage*

*Average linkage* adalah metode dalam analisis klaster yang mengelompokkan klaster berdasarkan rata-rata jarak antar obyek. Untuk menghitung algoritma antar obyek dapat menggunakan rumus persamaan 4.

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{uv} N_w} \quad (4)$$

Dimana  $d_{ik}$  adalah jarak antara obyek i pada klaster (uv) dan k pada klaster dan k pada klaster w, sedangkan  $N_{uv} N_w$  adalah banyaknya klaster item pada klaster uv dan w

d. *Ward linkage*

*Ward linkage* adalah metode analisis klaster yang meminimumkan variasi antar obyek yang berada pada klaster dengan menggunakan perhitungan *error sum of*

*square* (ESS) dalam satu kelompok. Untuk menghitung analisis klater dengan *ward linkage*, menggunakan persamaan 5.

$$d_{(uv)w} = \frac{[(n_w+n_u)d_{uw}+(n_w+n_v)d_{vw}]-n_w d_{(uv)}}{n_w+n_{uv}} \quad (5)$$

Langkah akan terus terulang sampai menemukan klaster yang dikehendaki.

### 3. Evaluasi klaster

Pengujian dilakukan untuk menguji seberapa dekat hubungan antar obyek dalam klaster, sehingga diketahui seberapa tepat data dikelompokkan. Koefisiensi *silhouette* merupakan gabungan dari metode *cohesion* yang berfungsi mengukur seberapa dekat relasi dan metode *separation* untuk mengukur seberapa jauh kluster terpisah dengan klaster lain (Hidayati et al., 2021). Koefisiensi *silhouette* nilai persamaan 6.

$$S_i = \frac{b_i - a_i}{\max(a_i - b_i)} \quad (6)$$

$S_i$  adalah nilai *silhouette*,  $a_i$  adalah rata-rata jarak obyek ke- $i$  dengan semua obyek dalam klaster dan  $b_i$  adalah rata-rata jarak obyek ke- $i$  dengan klaster yang berbeda. Nilai *Silhouette* berkisar antara 0 sampai 1, dimana jika nilai *silhouette* lebih dari 0,6 dianggap struktur klaster yang dihasilkan adalah baik, sedangkan jika nilainya dibawah 0,5 maka struktur dianggap lemah.

### 4. Pemodelan Preferensi Pengguna

Setelah proses pengklasteran selesai, selanjutnya dapat membangun model preferensi pengguna untuk setiap klaster. Model ini merupakan model prediktif yang memberikan prediksi preferensi pengguna terhadap terhadap item berdasarkan pengklasteran (Gupta & Patil, 2015). Model preferensi pengguna dibangun berdasarkan data dari kuisioner yang didapatkan dari data yang mempengaruhi wali murid siswa yang memiliki anak usia SD atau pra SD dalam menentukan pemilihan sekolah dasar anak. Dari hasil kuisioner didapatkan data presentase tertinggi dari faktor yang mempengaruhi pemilihan sekolah dasar di kota surakarta yaitu:

1. Biaya pendidikan
2. Unsur Religius sekolah
3. Fasilitas Sekolah
4. Lama pembelajaran di sekolah
5. Status sekolah

### 5. Evaluasi model sistem rekomendasi

Evaluasi dilakukan untuk mengetahui kelayakan hasil klasterisasi untuk dijadikan dasar dalam rekomendasi sistem. Evaluasi pada sistem rekomendasi dilakukan dengan menggunakan model *Mean absolute Error* (MAE) yaitu menghitung rata-rata selisih nilai prediksi terhadap nilai aktual. Metode ini digunakan untuk mengukur keakuratan terhadap prediksi (Fadlil & Mahmudy, 2007).

## III. HASIL DAN PEMBAHASAN

### 1. Prapemrosesan

#### a. Data asli

Data yang digunakan dalam penelitian ini berjumlah 245 sesuai dengan data sekolah dasar di wilayah kota Surakarta yang terdaftar di Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Data awal penelitian dapat dilihat pada tabel 1.

Tabel 1. Data asli Sekolah Dasar di Wilayah Surakarta.

No	Nama SD	Ak red	Religi	Status	Jam	PD	Guru	kelas	perpus	lab	Uang masuk	SPP
1	SDN Banyuagung	A	Umu m	Negeri	Reg	124	8	6	1	1	0	0
2	SDN Banyuanyar2	A	Umu m	Negeri	Reg	111	8	6	1	1	0	0
....	.....											
245	SD Muhammadiyah Makam BERGOLA	C	islam	Swasta	Reg	47	6	6	0	1	800.000	125.000

b. Penentuan variabel data

Penentuan variabel pada penelitian ini didasarkan pada data faktor yang mendasari orang tua dalam menentukan pemilihan sekolah dasar yang diambil dari data kuisioner yang disebarkan kepada wali murid di wilayah surakarta dengan responden sebanyak 83 responden. Dari hasil pengolahan data kuisioner digunakan presentase tertinggi faktor yang mempengaruhi orang tua dalam memilih sekolah dasar untuk anak. Berikut variabel yang digunakan sebagai data klaster,

1. Biaya pendidikan (Uang Masuk dan SPP)
2. Unsur Religius sekolah (Religi)
3. Fasilitas Sekolah (Fasilitas)
4. Lama pembelajaran di sekolah (Jam belajar)
5. Status sekolah (status)

c. Transformasi data

Setelah menentukan variabel data, langkah berikutnya adalah melakukan transformasi data. Transformasi data merupakan proses merubah skala data kedalam bentuk lain sehingga data memiliki distribusi yang diharapkan. Pada transformasi data ini juga dilakukan pengkodean untuk atribut karakter menjadi atribut berbentuk numerik (Rahmatullah et al., 2020). Hasil transformasi data dapat dilihat pada tabel 2.

Tabel 2. Data hasil transformasi data

No	Nama SD	Religi	Status	Jam sekolah	Fasilitas Sekolah	Uang masuk	SPP
1	SDN Banyuagung	1	1	1	3	1	1
2	SDN Banyuanyar2	1	1	1	3	1	1
...	...						
23	Sd Muhammadiyah 19 Kemlayan	2	2	2	2	2	2

2. Analisis Jumlah klaster

a. Jarak antar obyek

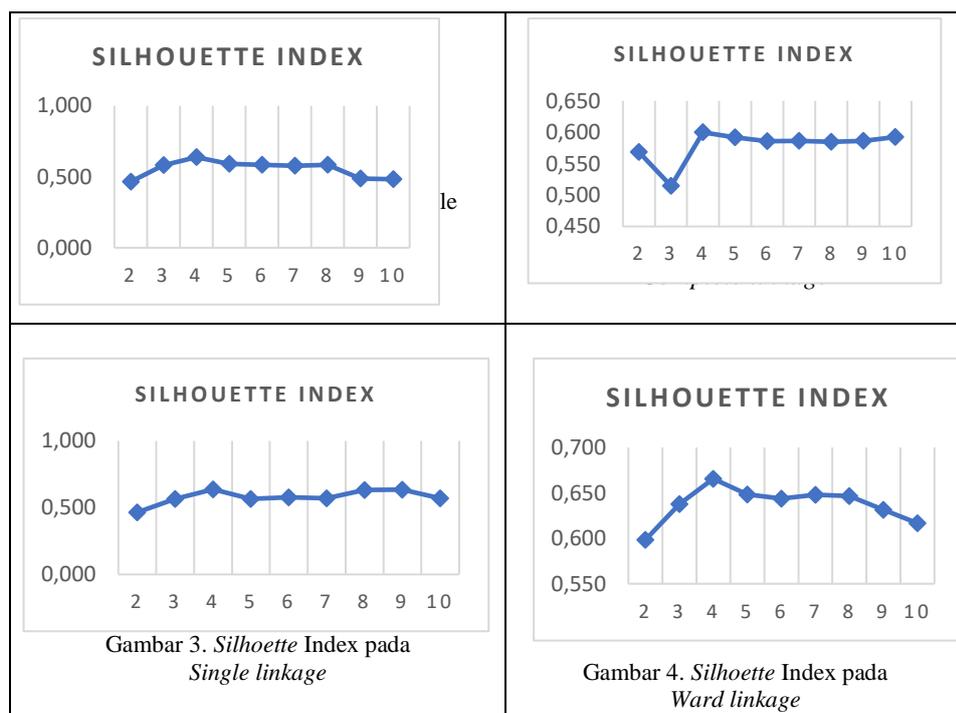
Jarak antar obyek diukur dengan menggunakan rumus euclidean distance, yang dapat dilihat pada persamaan 1. Matriks jarak antar obyek dapat dilihat pada tabel 3.

Tabel 3. Matriks jarak antar obyek

	SD 1	SD 2	....	SD 244	SD 245
SD 1	0	0,000	...	3,664	3,109
SD 2	0,000	0	....	3,664	3,109
....	...	...	.....	....	...
SD 244	3,664	3,664	.....	0	1,120
SD 245	3,109	3,109	.....	1,120	0

b. Penentuan jumlah kluster

Untuk menentukan jumlah kluster, digunakan grafik index *silhouette*, untuk melihat grafik silhoete dapat dilihat pada gambar 1 berikut.



Gambar 1. grafik silhoete

Gambar diatas adalah grafik index *Silhouette* untuk masing-masing metode. Dari grafik indeks silhoutte diatas menunjukkan kecenderungan jumlah kluster optimal yang digunakan adalah 4 kluster, sehingga pada penelitian menggunakan jumlah kluster = 4. Untuk nilai

c. Hasil *clustering* menggunakan *Hierarchy Clustering*

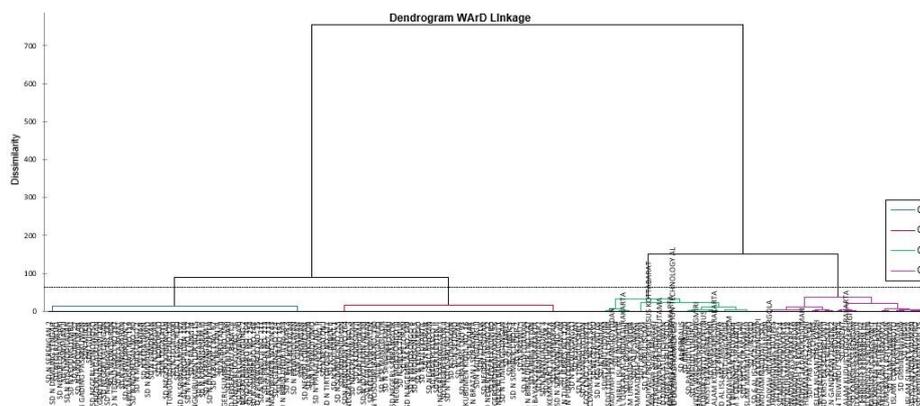
Clusetreing disini digunakan untuk mengelompokkan data sekolah di wilayah Surakarta berdasarkan variabel yang sudah terbentuk. Pengelompokanm didasarkan pada jarak antar obyek yang sudah dihitung dnegan menggunakan euclidean distance. Pembentukan kluster dilakukan dengan menggunakan 4 algoritma yaitu *Single linkage*, *Complete linkage*, fleksible linkage dan *ward linkage*. Untuk melihat perbandingan nilai *silhouette* pada msaing-masing metode dapat dilihat pada tabel 4.

Tabel 4. Nilai *Silhouette* untuk masing-masing metode

metode	Nilai Silhoutte klaster=4
<i>Single linkage</i>	0,638
Fleksible linkage	0,638
Complate linkage	0,600
<i>Ward linkage</i>	0,666

Dari hasil perbandingan nilai *Silhouette* metode-metode tersebut diperoleh nilai *silhouette* terbesar untuk klaster = 4 adalah metode *ward linkage*, sehingga untuk selanjutnya metode penelitian ini menggunakan metode *ward linkage*.

Hasil analisa klaster menggunakan metode ward dapat dilihat pada dendrogram gambar ..



Gambar 2. Dendrogram 4 klaster dengan metode *ward linkage*

Gambar 2. merupakan dendrogram visualisasi *cluster optimal* dari metode *ward linkage*. Pada dendrogram tersebut dapat diketahui anggota dari masing-masing klaster, dimana klaster 1 ditandai dengan garis merah dengan jumlah anggota 72 sekolah, klaster 2 ditandai dengan garis biru dengan anggota klaster 80 sekolah, klaster 3 ditandai dengan garis hijau dan kaslter 4 ditandai dengan garis ungu dengan anggota 45 sekolah. sedangkan untuk sentral obyek adalah klaster 1 dengan sentral obyek SD N Banyuwangung 3, klaster 2 dengan sentrak obyek SDN banyuanyar 3, kalster 3 dengan sentral obyek SD Muhammadiyah 3 Surakarta, dan klaster 4 dengan sentral obyerk SD Muhammadiyah 15 Surakarta. Untuk melihat lebih rinci adat sentral obyek dapat dilihat pada tabel 5.

Tabel 5. *Obyek Central* data

Cluster	Status	Religi	Fasilitas	Baiya Masuk	SPP
1 (Sd N Banyuwangung 3)	1,000	1,000	3,000	1,000	1,000
2 (Sd N Banyuanyar 3)	1,000	1,000	2,000	1,000	1,000
3 (Sd Muhammadiyah 3 Ska)	2,000	2,000	3,000	3,000	3,000
4 (Sd Muhammadiyah 15)	2,000	2,000	3,000	2,000	2,000

Masing-masing klaster memiiki claster centroid yang dijadikan dasar dalam menghitung jarak dengan data *testing* yang akan dijadikan dasar dalam sistem rekomendasi untuk melihat *cluster centroid* pada masing-masing klaster dapat dilihat pada tabel 6.

Tabel 6. Pusat klaster (*Cluster Centroid*)

Cluster	status	religi	fasilitas	Baiya Masuk	SPP	Sum of weights	Within-cluster variance
1	1,000	1,000	3,208	1,000	1,000	72,000	0,167
2	1,000	1,000	1,838	1,000	1,000	80,000	0,138
3	2,000	2,200	3,244	3,467	3,467	45,000	1,543
4	2,000	2,340	2,596	1,766	1,766	47,000	1,624

Pada tabel diatas menjelaskan rician pusat klaster pada masing-masing variabel. Pada tabel tersebut juga memuat jumlah bobot klaster dan varian dalam masing-masing klaster.

3. Evaluasi klaster

Evaluasi klaster dilakukan dengan menggunakan koefisiensi *silhouette*. Koefisiensi *silhouette* menghitung sleuruh obyek yang terdapat pada *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap obyek setiap klaster dengan klaster terdekatnya. Nilai keseluruhan *silhouette* untuk jumlah klaster k, dapat didefinisikan sebagai  $sil(k)$  yang dihitung dengan persamaan 7.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \tag{7}$$

Dimana  $sil(k)$  adalah nilai silhoutte untuk seluruh klaster,  $|k|$  adalah banyaknya klaster k dan  $sil(c_i)$  adalah rata-rata nilai silhoutte. Rentang rata-rata nilai *silhouette* adalah 0 sampai 1, semakin mendekati angka 1 maka dapat dikatakan obyek memiliki kecocokan yang kuat pada klaster, sedangkan jika mendekati 0 maka obyek memiliki kecocokan yang lemah pada klaster. Untuk melihat tabel nilai *silhouette* dapat dilihat pada tabel 7.

Tabel 7. Kriteria nilai *silhouette*

Nilai <i>Silhouette</i>	penafsiran
0,71-1,00	Klaster kuat
0,51-0,70	Klaster baik
0,26-0,50	Klaster lemah
0,00-0,25	Klaster buruk

Pada penelitian ini nilai *Silhouette* untuk optimal cluster dengan metode *Ward linkage* adalah **0,666**, sehingga pengklasteran dapat dikatakan baik, dan layak untuk digunakan sebagai dasar dalam sistem rekomendasi.

4. Preferensi pengguna

Preferensi pengguna merupakan model prediktif untuk memprediksi terhadap item berdasarkan hasil pengklasteran. Dari hasil tabel 1. Diperoleh nilai tertinggi bahwa faktor utama dalam memilih sekolah dasar anak adalah ditentukan oleh beberapa faktor yaitu biaya sekolah, fasilitas sekolah, status sekolah, unsur religius sekolah dan jam sekolah. dari hasil kuisisioner didapatkan hasil analisis yang dapat dilihat pada tabel 8.

Tabel 8. Tabel hasil preferensi pengguna dari kuisisioner

Data ke	status	fasilitas	Unsur religius	Jam sekolah	Biaya pendidikan
1	2	2	2	1	2
2	1	2	4	2	4
3	1	3	2	1	3
4	2	2	4	2	4
5	1	4	4	2	1
6	2	3	3	2	2
....	...	...	...	...	...
75	2	4	4	4	1

Keterangan: 1 = tidak pernah menjadikan variabel sebagai pertimbangan.  
 2 = kadang menjadikan variabel sebagai pertimbangan.  
 3 = sering menjadikan variabel sebagai pertimbangan  
 4 = selalu dijadikan sebagai pertimbangan

Dari data hasil kuisioner tersebut dapat diambil kesimpulan preferensi pengguna adalah sebagai berikut

1. Pengguna yang selalu mempertimbangkan jam sekolah *fullday*, fasilitas yang baik, sekolah dengan status swasta dan unsur religius yang spesifik sebagai faktor utama dalam memilih sekolah dasar anak cenderung tidak pernah mempertimbangkan biaya.
2. Pengguna yang sering mempertimbangkan jam sekolah *fullday*, fasilitas yang baik, sekolah dengan status swasta dan unsur religius yang spesifik sebagai faktor utama dalam memilih sekolah dasar anak cenderung kadang mempertimbangkan biaya.
3. Pengguna yang kadang mempertimbangkan jam sekolah *fullday*, fasilitas yang baik, sekolah dengan status swasta dan unsur religius yang spesifik sebagai faktor utama dalam memilih sekolah dasar anak cenderung kadang mempertimbangkan biaya.
4. Pengguna yang tidak pernah mempertimbangkan jam sekolah *fullday*, fasilitas yang baik, sekolah dengan status swasta dan unsur religius yang spesifik sebagai faktor utama dalam memilih sekolah dasar anak cenderung selalu mempertimbangkan biaya
5. Penerapan sistem rekomendasi

Sistem rekomndasi didasarkan pada hasil klasterisasi yang dihasilkan, untuk menentukan rekomendasi sekolah maka digunakan perhitungan data testing terhadap claster centroid, untuk menghitung kedekatan data testing dengan centroid digunakan rumus euclidean yaitu menggunakan rumus persamaan 1 salah satu contoh data sampling data yang digunakan adalah

1. Status sekolah: swasta (1)
2. Fasilitas Sekolah: sangat baik (1)
3. Unsur religius: islam (2)
4. Jam sekolah: *fullday*
5. Spp: diatas 800.000 (4)
6. Biaya masuk: diatas 10 juta (4)

Untuk melihat perhitungan contoh sistem rekomendasi data sampling terhadap pusat klaster dapat dilihat pada tabel 9.

Tabel 9. Jarak sampel 1 dengan pusat klaster

	claster 1	claster 2	claster 3	claster 4
sample1	4,634	5,016	<b>1,043</b>	3,141

Dari tabel diatas dapat terlihat bahwa jarak terdekat sampel 1 adalah pada klaster 3 yaitu dengan jarak rata-rata **1,043**, sehingga kriteria pada sampel 1 akan direkomendasikan sekolah yang terdapat pada klaster 3.

6. Evaluasi sistem rekomendasi

Model sistem rekomendasi yang dihasilkan dapat diukur tingkat keakuratannya dapat menggunakan metode *Mean Average Error* (MAE). MAE menghitung error absolut antara rating yang sebenarnya dengan rating prediksi. Hasil perhitungan akurasi dengan metode MAE dapat dilihat pada persamaan 10.

$$MAE = \frac{1}{n} \sum_{j=1}^n |Y' - Y| \tag{10}$$

Hasil perhitungan nilai MAE dapat dilihat pada tabel 10.

Tabel 10. Hasil akurasi sistem rekomendasi dengan perhitungan MAE

	claste r 1	claster 2	claster 3	claster 4	hasil rekomendasi	kenyataan	$ Y'-Y $
sampel 1	4,634	5,016	<b>1,043</b>	3,141	3	3	0
sampel 2	0,199	<b>1,108</b>	4,229	3,263	2	2	0
sampel 3	3,638	3,797	<b>0,721</b>	1,695	3	4	1
sampel 4	2,193	2,449	2,701	<b>2,097</b>	4	4	0
sampel 5	2,106	<b>0,799</b>	4,734	3,579	2	2	0
sampel 6	2,581	<b>2,315</b>	3,062	2,492	2	1	1
sampel 7	2,106	<b>0,799</b>	4,734	3,579	2	2	0
sampel 8	4,634	5,016	<b>1,043</b>	3,141	3	3	0
sampel 9	3,710	4,177	<b>0,992</b>	2,125	3	3	0
sampel 10	3,022	3,580	1,992	<b>1,500</b>	4	4	0
Jumlah							2
MAE							<b>0,20</b>

Dari tabel diatas diperoleh nilai rata-rata kesalahan dengan menggunakan rumus MAE adalah sebesar 0,20. Hasil tersebut menunjukkan bahwa akurasi sistem rekomendasi baik dan layak untuk digunakan.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa sistem rekomendasi dengan menggunakan data hasil klusterisasi layak untuk digunakan. Berdasarkan hasil evaluasi sistem rekomendasi dengan menggunakan perhitungan statistik *Mean Average Error* (MAE), diperoleh nilai MAE sebesar 0,2. Jumlah klaster yang digunakan pada penelitian ini adalah 4, hal tersebut didasarkan pada grafik nilai *silhouette* dari metode *Single linkage*, *Complete linkage*, *Ward linkage* dan fleksible *lnkage* diperoleh nilai *silhouette* tertinggi adalah pada klaster 4, dan metode yang digunakan adalah *Ward linkage* yang memiliki nilai *silhouette* tertinggi. Evaluasi klaster dilakukan dengan menghitung nilai *silhouette*, dan diperoleh nilai 0,666 yang menunjukkan bahwa pengelompokan klaster tersebut memiliki kedekatan yang baik antar obyek dalam satu klaster.

### 5.2 Saran

Penelitian ini dapat dikembangkan dengan mengembangkan aplikasi sistem rekomendasi yang dapat dimanfaatkan oleh orang tua dalam menentukan pilihan sekolah dasar bagi anaknya. Penelitian selanjutnya dapat dikembangkan penelitian terkait aplikasi sistem rekomendasi pemilihan sekolah dasar yang dapat langsung di implementasikan oleh orang tua murid.

## DAFTAR PUSTAKA

- Alam Jusia, P., Muhammad Irfan, F., & Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, S. (2019). Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means. *Jurnal IKRA-ITH Informatika*, 3(3), 75.
- Atmojo, S., Dewi, S., Widhiyanta, N., & Utami, R. (2021). Sistem Informasi Rekomendasi

- Pemilihan Sekolah Dasar. *Jurnal Ilmiah Edutic*, 7(2), 85–93.
- Beregovskaya, I., & Koroteev, M. (2021). *Review of Clustering-Based Recommender Systems*. September. <https://doi.org/10.13140/RG.2.2.34745.90725>
- Chandan K. Reddy, C. C. A. (2018). *Data Clustering Algorithms and Applications*. CRC Press.
- Chang, Y., Zhao, Y., & Feng, S. (2012). Comparisons of clustering SSCI journals by emerging hierarchical clustering, hierarchical clustering and minimum spanning tree. *Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012, Fskd*, 2898–2901. <https://doi.org/10.1109/FSKD.2012.6233748>
- Das, J., Mukherjee, P., Majumder, S., & Gupta, P. (2014). Clustering-based recommender system using principles of voting theory. *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, 230–235. <https://doi.org/10.1109/IC3I.2014.7019655>
- Dewi, A. F., & Ahadiyah, K. (2022). Agglomerative *Hierarchy Clustering* Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan. *Zeta - Math Journal*, 7(2), 57–63. <https://doi.org/10.31102/zeta.2022.7.2.57-63>
- Edy Irwansyah, M. F. (2015a). *Clustering Teori dan Aplikasi*.
- Edy Irwansyah, M. F. (2015b). *Clustering Teori dan Aplikasi*. DeePublish.
- Fadliana, A., & Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, 4(1), 35–40. <https://doi.org/10.18860/ca.v4i1.3172>
- Fadlil, J., & Mahmudy, W. F. (2007). Pembuatan Aplikasi Rekomendasi Menggunakan Decision Tree dan Clustering. *Ilmiah KURSOR*, 3(Kursor), 45–46.
- Gellysa Urva, Desyanti, dkk. (2020). *Penerapan Data Mining Di Berbagai Bidang : Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia.
- Gupta, U., & Patil, N. (2015). Recommender system based on Hierarchical Clustering algorithm Chameleon. *Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015*, 1006–1010. <https://doi.org/10.1109/IADCC.2015.7154856>
- Hidayati, R., Zubair, A., Pratama, A. H., & Indana, L. (2021). Analisis *Silhouette Coefficient* pada 6 Perhitungan Jarak K-Means Clustering *Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering*. *Techno.COM*, 20(2), 186–197.
- Islam, J. K., Sekolah, D. I., Islam, D., Sdit, T., Rusadi, M. A., & Fauzi, A. M. (2022). *AL-QALAM AL-QALAM*. 14(1).
- Khasanah, N. (2012). FAKTOR-FAKTOR YANG MEMPENGARUHI ORANG TUA DALAM PENGAMBILAN KEPUTUSAN MEMILIH SEKOLAH DASAR SWASTA (SD Virgo Maria 2 dan SDIP. H. Soebandi Kecamatan Bawen Kabupaten Semarang). *Satya Widya*, 28(2), 137. <https://doi.org/10.24246/j.sw.2012.v28.i2.p137-146>
- Khasbulloh, M. N. (2022). Preferensi Masyarakat dalam Pemilihan Lembaga Pendidikan Islam: Studi Pada SD NU Insan Cendekia Kediri. *JoIEM (Journal of Islamic Education Management)*, 1(2), 51–66. <https://doi.org/10.30762/joiem.v1i2.99>
- Kusrini, E. taufiq luthfi. (2009). *Algoritma Data Mining*. Penerbit Andi.
- Nisha, & Kaur, P. J. (2016). Cluster quality based performance evaluation of hierarchical clustering method. *Proceedings on 2015 1st International Conference on Next Generation Computing Technologies, NGCT 2015, September*, 649–653. <https://doi.org/10.1109/NGCT.2015.7375201>
- Pratikto, R. O., & Damastuti, N. (2021). Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir. *JOINTECS (Journal of*

- Information Technology and Computer Science*), 6(1), 13.  
<https://doi.org/10.31328/jointecs.v6i1.1473>
- Rafles Ginting. (2022). *Analisis Big Data*. CV. Penerbit Lakeisha.
- Rahmatullah, S., Wahyuni, S., Chaining, M. F., & Method, F. C. (2020). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Furniture Terlaris Menggunakan Metode Knearest Neighbor*. 2, 75–86.
- Rokhmah, S., & Rozaq Rais, N. A. (2022). Application of Data Mining for Prediction of Long Covid on Covid-19 Survival With Feature Selection and Naïve Bayes Method. *Jurnal Teknik Informatika (Jutif)*, 3(5), 1397–1405.  
<https://doi.org/10.20884/1.jutif.2022.3.5.561>
- Saputri, R. D., Putri, R. A., & Rini, E. F. (2018). Faktor-faktor yang Mempengaruhi Pemilihan Sekolah Dasar dalam Mendukung Kota Surakarta Layak Anak. *Region: Jurnal Pembangunan Wilayah Dan Perencanaan Partisipatif*, 13(2), 152.  
<https://doi.org/10.20961/region.v13i2.21172>
- Soltani, A., & Akbarzadeh-T, M. R. (2013). Confabulation based recomender system. *Proceedings of the 3rd International Conference on Computer and Knowledge Engineering, ICCKE 2013, Ickce*, 82–86.  
<https://doi.org/10.1109/ICCKE.2013.6682822>
- Starczewski, A. (2012). A new hierarchical clustering algorithm. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7268 LNAI(PART 2), 175–180. [https://doi.org/10.1007/978-3-642-29350-4\\_21](https://doi.org/10.1007/978-3-642-29350-4_21)
- Surakarta, B. (2020). *Data sekolah dasar di kota surakarta*.  
<https://surakartakota.bps.go.id/statictable/2020/03/30/167/jumlah-sekolah-guru-dan-murid-sekolah-dasar-sd-di-bawah-kementrian-pendidikan-dan-kebudayaan-menurut-kecamatan-di-kota-surakarta-2018-2019-dan-2019-2020.html>
- Yuliana, K., & Aini, A. (2018). *Sistem Pendukung Keputusan Pemilihan Sekolah Dasar Menggunakan Metode Analytic Hierarchy Process (Ahp) Pada Gugus Sungai Jindah Kecamatan Banjarmasin Utara*. 1(12), 205–220.  
<http://amikjtc.com/jurnal/index.php/jurnal/article/view/143>