

# Penentuan Kriteria Dalam Memilih Sekolah Dasar Dengan Menerapkan K-Means Clustering (Studi Kasus : Wilayah Kecamatan Mampang)

*Determination of Criteria in the Selection of Primary Schools*

*By applying K-Means Clustering*

*(Case Study: Mampang District Area)*

**Dwi Wulandari Sari<sup>1</sup> dan Kurnia Gusti Ayu<sup>2</sup>**

<sup>1</sup>Universitas Mercu Buana

E-mail : dwi.wulandari@mercubuana.ac.id

<sup>2</sup>Universitas Mercu Buana

E-mail : kurnia.gusti@mercubuana.ac.id

**Abstrak---** Memilih sekolah adalah sebuah hal yang menjadi rutinitas orangtua ketika sudah waktunya bagi anak-anak untuk bersekolah. Semakin banyaknya sekolah dasar baik negeri ataupun swasta menjadi banyaknya pula pilihan bagi orangtua sebelum mendaftarkan putra/putri mereka. Sayangnya terdapat permasalahan ketika orangtua ingin memilih sekolah dasar yang sesuai dengan kebutuhan dan kemampuan orangtua dan anak-anak. Beberapa penelitian telah dilakukan untuk permasalahan di atas, seperti Clustering K-Means (Studi Kasus : Perpustakaan). Tulisan ini bertujuan menganalisa untuk menentukan kriteria yang umumnya dijadikan dasar dalam pemilihan sekolah dasar dengan menggunakan Algoritma K-Means Clustering. Dengan menggunakan software analisis data yaitu XLSTAT, dilakukan perhitungan terhadap 54 sekolah dasar baik negeri dan swasta. Didapatkan hasil penelitian berupa rata-rata orangtua di wilayah Mampang Prapatan memilih sekolah berdasarkan kriteria Lokasi sekolah, sejarah keluarga yang pernah bersekolah ditempat yang sama dan terakhir pendidikan orangtua yang ikut menentukan sekolah anak. Diharapkan dengan hasil penelitian ini dapat berguna bagi pihak sekolah sehingga dapat memperhitungkan perkiraan jumlah siswa di tahun ajaran berikutnya.

**Kata Kunci---** K-Means, Clustering, metode CRIPS-DM

**Abstract---** It has become a routine for parents to choose a suitable school when the time comes for their children to enter school. The increasing number of elementary schools, both public and private ones, means that parents have a lot of choices for their children. Unfortunately, there is a certain issue when picking elementary schools that suit the needs and capabilities of both parents and children. A number of researches have been done to address this issue, such as K-Means Clustering (Case Study: Library). This article aims to determine the criteria that are most commonly used as points of consideration by parents when choosing elementary schools using K-Means Clustering algorithm. By using XLSTAT data analysis software, a calculation has been performed on 54 elementary schools, both public and private ones. The result of this research is that the average parents in Mampang Parapatan use the criteria of location, history of family members who have gone to the same school, and parents's education level in determining which schools are suitable for their children. It is expected that this research can be useful for schools in predicting the number of students who enroll in the following academic year.

**Keywords---** K-Means, Clustering, CRIPS-DM method

## 1. PENDAHULUAN

Sekolah adalah tempat kita menimba ilmu, di tempat inilah anak-anak kita mengenal berbagai bidang ilmu sekaligus bersosialisasi dengan guru dan teman-temannya. Sekolah bisa menjadi tempat yang nyaman ketika anak-anak kita merasa bersemangat untuk bisa meraih sesuatu. Untuk

mendapatkan tempat yang nyaman saat anak-anak kita menimba ilmu, kita sebagai orangtua harus mampu menempatkan mereka sesuai kemampuan atau bakat serta lingkungan yang mendukung sehingga mereka kelak bisa tumbuh dan berkembang secara maksimal.

Oleh karena itu, memilih sekolah menjadi hal yang penting untuk dipikirkan orangtua. Sebenarnya banyak sekali kendala yang dihadapi orangtua ketika ingin memilih sekolah baik sekolah dasar ataupun tingkat menengah. Hal ini terjadi karena, belum adanya informasi tentang sekolah itu sendiri dan belum ada dukungan data/informasi tentang kriteria apa saja yang menjadi alasan bagi orangtua memilih sekolah-sekolah tersebut.

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

Salah satu metode yang terdapat dalam data mining yang digunakan dalam penelitian ini adalah pengelompokan (*Clustering*). dimana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai “vektor karakteristik” atau “centroid”[1].

Nurul Rohmawati, dkk, mengangkat judul “Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa”. Penelitian ini menjelaskan bahwa Pengelompokan data pelamar beasiswa Bantuan Belajar Mahasiswa (BBM) dikelompokkan menjadi 3 kelompok yaitu berhak menerima, dipertimbangkan dan tidak berhak menerima beasiswa. Pengukuran ini di lihat dari hasil cluster dengan menghitung nilai kemurnian (*purity measure*) dari masing – masing cluster yang dihasilkan. Data yang digunakan dalam penelitian ini adalah data mahasiswa yang mengajukan beasiswa kepada Fakultas Ilmu Komputer UNSIKA sebanyak 36 mahasiswa. Data akan diubah menjadi 3 dataset dengan format yang berbeda-beda, yakni data atribut kodifikasi sebagian, atribut kodifikasi keseluruhan dan atribut data asli. Nilai *purity* pada dataset data kodifikasi sebagian untuk hasil cluster algoritma kmeans sebesar 61.11%. Pada dataset kodifikasi keseluruhan nilai *purity* hasil cluster algoritma kmeans sebesar 80.56%. Dan untuk dataset data asli nilai *purity* hasil cluster algoritma k-means sebesar 75%. Maka dapat disimpulkan bahwa algoritma kmeans lebih cocok digunakan pada dataset dengan format atribut yang dikodifikasi keseluruhan.[2].

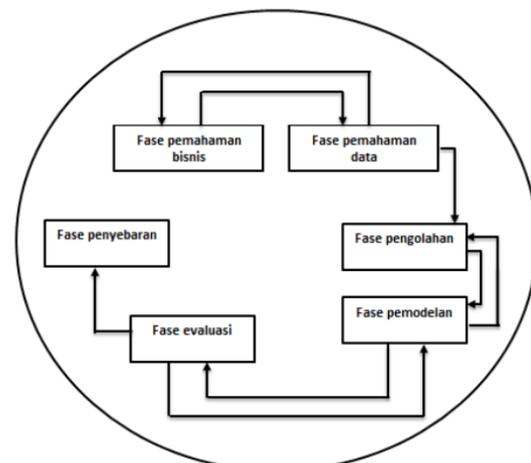
Windha Mega P.D, mengangkat judul “Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita”. Penelitian ini menjelaskan bahwa Pengelompokan status gizi

balita di Desa Kembang Songo menggunakan metode K-Means dilakukan melalui beberapa tahapan yaitu : penentuan tujuan bisnis, pengumpulan data 50 balita di Desa Karang Songo, pengelompokan status gizi balitake dalam 5 cluster yaitu cluster 1 - gizi buruk; cluster 2 - gizi kurang; cluster 3 - gizi baik; cluster 4 - gizi lebih; cluster 5 - obesitas, perhitungan cluster menggunakan software SPSS, analisa hasil data output, pengelompokan status gizi balita menggunakan tabel Growth Chart, dan yang terakhir melakukan pengujian dengan membandingkan hasil pengelompokan algoritma K-means dan tabel Growth Chart. Dengan membandingkan hasil pengelompokan menggunakan tabel Growth Chart dan algoritma K-Means didapat 17 data yang memiliki kelompok yang sama. Dari angka ini dapat disimpulkan bahwa algoritma K-Means hanya memiliki nilai akurasi 34% benar[3]

Pada penelitian ini, penulis menggunakan data orangtua murid kelas 1 di Sekolah Dasar diantaranya lokasi tempat tinggal, lokasi sekolah, sejarah keluarga yang pernah bersekolah ditempat yang sama, pendapatan orangtua, pendidikan terakhir orangtua. Diharapkan penelitian ini dapat diketahui kriteria yang biasa digunakan untuk memilih sekolah dasar dengan menggunakan Algoritma K-Means Clustering serta mencari nilai akurasi penggunaan K-Means Clustering untuk mengetahui kecocokan penggunaan algoritma ini dalam penentuan kriteria.

## 2. METODA

Metoda penelitian yang digunakan adalah Metodologi data mining CRISP-DM yang diperkenalkan pertengahan tahun 1990 oleh sebuah perusahaan konsorsium Eropa[4]. Dalam CRISP\_DM, sebuah proyek Data Mining memiliki siklus hidup yang terbagi dalam 6 fase seperti terlihat pada gambar 1[5].



Gambar 1. Siklus hidup CRISP-DM

### 3. HASIL DAN PEMBAHASAN

Adapun langkah-langkah yang dilakukan pada penelitian ini menurut pada gambar 1 adalah sebagai berikut :

#### 1. Fase Pemahaman Bisnis (*Business Understanding*)

meliputi penentuan tujuan bisnis, menilai situasi saat ini, menetapkan tujuan data mining dan mengembangkan rencana proyek. Tujuan bisnis yang akan dilakukan pada penelitian ini adalah bagaimana melakukan pengelompokan kriteria pemilihan sekolah dasar di kecamatan Mampang Prapatan menggunakan K-Means.

#### 2. Fase Pemahaman Data (*Data Understanding Phase*)

Setelah tujuan bisnis dan rencana proyek diterapkan, langkah selanjutnya melakukan pengumpulan data awal, deskripsi data, eksplorasi data, dan verifikasi kualitas data. Penelitian yang diusulkan ini menggunakan data primer, dengan respondennya adalah siswa dengan usia sekolah dasar khususnya kelas 1 yang ada di masing-masing sekolah di kecamatan Mampang Prapatan Jakarta Selatan. Dimana jumlah sekolah dasar negeri dan swasta berjumlah 54 sekolah. Dan setiap sekolah dasar akan diambil sample responden sebanyak 2 orang. Sehingga total data seluruhnya 108 responden. Perhitungan yang akan digunakan untuk menentukan kriteria pemilihan sekolah dasar. Dimana data yang akan diambil terkait : lokasi sekolah, lokasi rumah, memiliki sejarah keluarga pernah bersekolah di tempat yang sama, pendapatan orangtua, pendidikan orangtua. Menjadikan 5 cluster yaitu C1, C2, C3, C4 dan C5

#### 3. Fase Pengolahan Data (*Data Preparation Phase*)

Pada tahap ini diidentifikasi dan pembangunan jawaban dari data yang telah dikumpulkan untuk bisa melakukan pengelompokan dan pemilahan ke dalam kelompok-kelompok yang telah ditentukan. Terdapat beberapa *missing value* pada kriteria penghasilan orang tua, kemudian *missing value* ini akan diisi dengan menggunakan teknik *mean imputation* atau diisi dengan nilai rata-rata dari kriteria penghasilan orang tua.

$$\bar{x} = \frac{\text{Jumlah total atribut penghasilan orang tua}}{\text{Jumlah data}}$$

Pengkategorian kriteria penghasilan orang tua dibagi dengan jumlah tanggungan orang tua (dalam penelitian ini disingkat JT) dengan cara mencari nilai *standar deviasi* dan *mean* dari masing-masing kriteria kemudian dikategorikan berdasarkan tabel 1.

**Tabel 1. Pengkategorian JT<sub>[6]</sub>**

Kategori	Kualifikasi	Kodifikasi
Kategori 4	$JT \leq \bar{x} - S$	4
Kategori 3	$\bar{x} - S < JT < \bar{x}$	3
Kategori 2	$\bar{x} \leq JT < \bar{x} + S$	2
Kategori 1	$JT \geq \bar{x} + S$	1

Dari data yang akan diolah dengan menggunakan metode *k-means clustering*, maka data yang berjenis data nominal seperti pendidikan orang tua, lokasi rumah, dan lokasi sekolah harus diinisialisasikan terlebih dahulu dalam bentuk angka.

Untuk transformasi data lokasi sekolah, penulis merubah data tak terdefiniskan menjadi data nominal sebagai berikut :

**Tabel 2. Data Nominal Lokasi Sekolah**

Lokasi Sekolah	Angka
Pela Mampang	1
Bangka	2
Mampang Prapatan	3
Tegal Parang	4
Kuningan Barat	5

Untuk transformasi data lokasi rumah, penulis merubah data tak terdefiniskan menjadi data nominal sebagai berikut :

**Tabel 3. Data Nominal Lokasi Rumah**

Lokasi Rumah	Angka
Pela Mampang	1
Bangka	2
Kuningan Barat	3
Mampang Prapatan	4
Tegal Parang	5
Pancoran	6
Pasar Minggu	7
Cilandak	8
Kemang	9

Untuk transformasi data pendidikan orangtua, penulis merubah data tak terdefiniskan menjadi data nominal sebagai berikut :

**Tabel 4. Data Nominal Pendidikan Orangtua**

Pendidikan Orangtua	Angka
SD	1
SMP	2
SMA	3
S1	4
S2	5

## 4. Pemodelan

Pemodelan data mining dalam penelitian ini dibuat dengan menggunakan perangkat lunak XLSTAT[13]. Pada aplikasi ini telah tersedia algoritma clustering berupa algoritma k-means dapat dijalankan dengan menggunakan aplikasi tersebut.

Dari perhitungan menggunakan software XLSTAT yang memiliki menu analysis data K-means clustering, diperoleh hasil sebagai berikut :

Tabel 5. Jarak antara class centroid

Distances between the class centroids:					
	1	2	3	4	5
1	0	2,426	4,335	6,818	1,779
2	2,426	0	3,290	4,551	2,549
3	4,335	3,290	0	4,156	2,965
4	6,818	4,551	4,156	0	6,210
5	1,779	2,549	2,965	6,210	0

Tabel 6. Data Central Object

Central objects:					
Class	Lokasi Sekolah	Lokasi Rumah	Jumlah Kel.Pernah Sekolah ditempat yang sama	Pendapatan Orangtua	Pendidikan Orangtua
1 (Siswa 1)	1,000	1,000	1,000	4,000	4,000
2 (Siswa 17)	1,000	4,000	1,000	3,000	3,000
3 (Siswa 54)	4,000	4,000	1,000	4,000	3,000
4 (Siswa 89)	3,000	8,000	1,000	4,000	3,000
5 (Siswa 85)	3,000	2,000	1,000	3,000	3,000

Tabel 7. Jarak antara central object

Distances between the central objects:					
	1 (Siswa 1)	2 (Siswa 17)	3 (Siswa 54)	4 (Siswa 89)	5 (Siswa 85)
1 (Siswa 1)	0	3,317	4,359	7,348	2,646
2 (Siswa 17)	3,317	0	3,162	4,583	2,828
3 (Siswa 54)	4,359	3,162	0	4,123	2,449
4 (Siswa 89)	7,348	4,583	4,123	0	6,083
5 (Siswa 85)	2,646	2,828	2,449	6,083	0

Setelah dilakukan berbagai perhitungan dengan menggunakan software ini, didapatkan hasil akhir sebagai berikut :

Tabel 8. Hasil Akhir

Results by class:					
Class	1	2	3	4	5
Objects	29	15	25	17	22
Sum of weights	29	15	25	17	22
Within-class variance	1,845	1,552	2,990	2,618	1,320
Minimum distance to centroid	0,773	0,646	0,830	0,771	0,534
Average distance to centroid	1,270	1,110	1,622	1,504	1,044
Maximum distance to centroid	2,473	2,342	2,602	2,390	2,218

## 5. Evaluasi

Berdasarkan perhitungan dan hasil akhir yang tertera pada tabel 8, didapatkan bahwa C1 = 29, C2 = 15, C3 = 25, C4 = 17, C5 = 22. Dimana masing-masing C merupakan kriteria yang biasa dipilih orangtua. Yaitu C1 = Lokasi sekolah, C2 = Lokasi Rumah, C3 = jumlah keluarga yang pernah bersekolah ditempat yang sama, C4 = Pendapatan Orangtua, dan C5 = Pendidikan Orangtua. Sehingga jika diurutkan maka

Tabel 9. Urutan Centroid

Urutan	Centroid	Jumlah
1	C1	29
2	C3	25
3	C5	22
4	C4	17
5	C2	15

## 4. SIMPULAN

Penelitian ini dilakukan untuk mencari kriteria yang biasanya dijadikan dasar dalam memilih sekolah dasar bagi anak. Penyebaran kuisisioner terhadap 54 sekolah dasar baik negeri dan swasta, dengan mengambil 2 responden dari masing-masing sekolah. Data yang diolah dalam software analisis data berjumlah 108 responden. Dengan perhitungan K-Means clustering berdasarkan class yang digunakan. Maka hasil penelitian menunjukkan bahwa rata-rata orangtua di wilayah kecamatan Mampang Prapatan memilih sekolah berdasarkan 1. Lokasi sekolah, 2. Sejarah keluarga yang pernah bersekolah ditempat yang sama, 3. Pendidikan orangtua, 4. Pendapatan Orangtua dan 5. Lokasi rumah.

Dengan hasil penelitian seperti yang tertera diatas, diharapkan informasi yang didapat bisa dijadikan dasar dalam memperkirakan jumlah calon siswa pada tahun ajaran baru.

## 5. SARAN

1. Penelitian ini bisa dijadikan dasar dalam penelitian selanjutnya apabila ingin digunakan untuk mencari kriteria pemilihan sekolah dasar pada kecamatan atau wilayah yang lain.
2. Dengan informasi ini bisa berguna bagi pihak sekolah khususnya sekolah swasta untuk memperbaiki fasilitas sarana dan prasarana sehingga dapat lebih menarik minat peserta didik di tahun ajaran baru.
3. Untuk pengembangan penelitian berikutnya, bisa dibuatkan informasi berbasis web yang mampu menyajikan informasi yang terkait dengan pemilihan sekolah dasar sehingga membantu orangtua yang baru menyekolahkan anaknya.

## UCAPAN TERIMA KASIH

Ucapan terima kasih peneliti berikan kepada :

1. Pihak sekolah yang menjadi tempat penelitian dan menyebarkan responden kepada orangtua siswa.
2. Pihak Universitas Mercu Buana, atas kesempatan dan dana yang diberikan kepada peneliti sehingga bisa menjalankan penelitian
3. Orangtua dan keluarga peneliti atas waktu dan bantuannya sehingga memiliki kesempatan untuk menyelesaikan penelitian ini tepat pada waktunya.

## DAFTAR PUSTAKA

- [1] Tanjung Muhammad Rusdi, Erlinda Ningsih. (2014). “Perbandingan Penggunaan Data Real dan Data Hasil Normalisasi Pada Pengelompokan Data Dengan Menggunakan Metode Clustering (Studi Kasus PT. PHP Motor)”, Prosiding KeTIK UIN.Universitas Islam Negeri Medan – Sumatera Utara.
- [2] W.R. Rohmawati, dkk. (2015). “Implementasi Algoritma K-Means dalam Pengklasteran Mahasiswa Pelamar Beasiswa”, JITTER, ISSN : 2407 – 3911, Volume 1, No.2, 30 April.
- [3] Atthina, N., Iswari, I. (2015). *Klasterisasi Data Kesehatan Penduduk untuk enentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means*, Jurnal SNASTI, B52B-59.
- [4] Duhita, W.M.Pradnya. (2015). “Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita”, Jurnal Informatika, Vol. 15, No. 2, 2 Desember.
- [5] Hastuti, N. F., Saptono, R., & Suryani, E. (2012). “Pemanfaatan Metode K-Means Clustering Dalam Penentuan Penerima Beasiswa”. Jurnal Informatika.
- [6] Kusriani dan E.T Luthfi, (2009), “Algoritma Data Mining”, Andi Yogyakarta.