p-ISSN 2085-7829
e-ISSN 2798-3455

Penerapan Algoritma K-Medoids pada Pengelompokan Wilayah Provinsi di Indonesia Berdasarkan Indikator Pendidikan

Application of the K-Medoids Algorithm to Grouping Provinces in Indonesia Based on Education Indicators

Rama Septian[1, a)], Syaripuddin[1], dan Darnah Andi Nohe[2]
[1]Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman
[2] Laboratorium Statistika Terapan FMIPA Universitas Mulawarman

[a)]Corresponding author: ramaseptian53@gmail.com

ABSTRACT

Cluster analysis aims to group data that has the same characteristics into the same cluster, while data that has different properties will be placed in different clusters. The K-Medoids method is used in this clustering process by using representative objects as center points (medoids). The K-Medoids method was developed to overcome the weakness of the K-Means method which is sensitive to outliers, because objects with large values can deviate from the distribution of data based on their size. This study aims to obtain optimal clusters for clustering provincial regions in Indonesia based on education indicators, using 2020 education indicator data. The K-Medoids clustering results are validated using the Silhouette Coefficient (SC) which aims to assess the quality and strength of the cluster, by combining the cohesion and separation methods. Based on the results of the study, it was found that the optimal cluster consisted of 2 clusters with an SC value of 0.464. Cluster 1 consists of 14 provinces, while cluster 2 consists of 20 provinces.

Keywords: *K-Medoids, Silhouette Coefficient, Educational Indicators*

1. Pendahuluan

Data *mining* adalah suatu proses untuk mendapatkan informasi yang berguna dari gudang basis data berskala besar, yang membantu dalam proses pengambilan keputusan. Data mining merupakan gabungan dari beberapa disiplin ilmu, yang didefinisikan sebagai proses penemuan pola baru dari kumpulan data berskala besar, melibatkan kecerdasan buatan, pembelajaran mesin, statistika, dan teknologi basis data (Prasetyo, 2012). Data mining juga dapat digunakan untuk melakukan proses pengelompokan atau klasterisasi dengan tujuan untuk mengetahui pola universal dari data yang ada (Prasetyo, 2012). Menurut Suyanto (2017), klasterisasi adalah proses pengelompokan himpunan data ke dalam beberapa kelompok atau klaster sedemikian rupa sehingga objek-objek dalam suatu klaster memiliki kemiripan yang tinggi, namun sangat berbeda dengan objek-objek di klaster lainnya. Kemiripan dan ketidakmiripan dihitung berdasarkan nilai-nilai atribut yang menggambarkan objek tersebut.

Analisis klaster terdiri dari dua metode, yaitu metode hierarki dan metode non-hierarki. Menurut Anderberg (1973), metode non-hierarki dimulai dengan asumsi jumlah kelompok yang akan dibentuk sebelumnya, dan umumnya digunakan jika jumlah unit dalam objek pengamatan besar. Salah satu metode non-hierarki adalah K-medoids. K-medoids merupakan salah satu metode berbasis partisi yang menggunakan objek perwakilan (representatif) yang disebut medoids sebagai titik pusat atau controid. K-medoids melakukan partisi dengan cara meminimalkan ketidakmiripan atau dissimilarity antar setiap objek i dan objek representatif terdekat (Suyanto, 2017). Metode K-medoids didasarkan pada penggunaan medoids, dengan tujuan dapat mengurangi sensitivitas dari partisi yang dihasilkan sehubung dengan nilai-nilai ekstrim yang terdapat dalam dataset (Triyanto, 2015).

Pada penelitian ini, menggunaka metode pengelompokan berbasis partisi yaitu menggunakan K-medoids yang diaplikasikan dalam penentuan karakteristik provinsi di indonesia berdasarkan indikator Pendidikan.Adapun variabel yang digunakan adalah angka partisipasi kasar, angka partisipasi murni, angka partisipasi sekolah, persentase guru layak, dan angka putus sekolah. Berdasarkan data Survei Sosial Ekonomi pada bulan Maret 2020 yang dilakukan Badan Pusat Statistik (BPS) dalam Statistik Pendidikan 2020 pada jenjang SMA/MA sederajat, Indonesia belum mencapai kriteria Tuntas Paripurna dalam penuntasan wajib belajar karena nilai angka partisipasi kasar belum mencapai 95%. Sedangkan dilihat dari angka putus sekolahnya masih terdapat 1.13% penduduk yang putus sekolah dalam menempuh pendidikan jenjang SMA/MA sederajat.

2. Tinjauan Pustaka

2.1 *Data Mining*

Data mining merupakan sebuah langkah dalam proses *Knowledge Discovery in Database* (KDD) yang terdiri dari penerapan analisis data dan penemuan algoritma yang menghasilkan enumerasi tertentu terhadap

p-ISSN 2085-7829
e-ISSN 2798-3455

pola pada data. *Data mining* juga diartikan sebagai sebuah proses ekstrasi informasi atau pengetahuan baru dari sejumlah besar data yang dapat berguna dalam proses pengambilan keputusan. Pengetahuan bisa berupa pola data atau relasi antar data yang valid. *Data mining* ditujukan untuk mengekstrak pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan managemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan *online updating* (Suyanto, 2017).

Menurut Defiyanti (2017), *data mining* adalah proses penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Karakteristik *data mining* sebagai berikut:

a. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
b. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
c. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

2.2 Fungsi *Data Mining*

Menurut Dunham (2003), *data mining* melibatkan banyak algoritma yang berbeda untuk menyelesaikan tugas yang berbeda. Semua dari algoritma ini mencoba untuk menyesuaikan model dengan data. Algoritma memeriksa data dan menentukan model yang paling dekat dengan karakteristik data yang diperiksa. Algoritma data mining dapat dikarakteristikkan terdiri dari tiga bagian:

i. Model: Tujuan dari algoritma ini adalah untuk menyesuaikan model dengan data
ii. Preferensi: Beberapa kriteria harus digunakan agar sesuai dengan satu model diatas yang lain
iii. Pencarian: Semua algoritma memerlukan beberapa teknik untuk mencari data.

2.3 Operasi *Data Mining*

Menurut Fayyad (1996), istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut berbeda, *data mining* merupakan bagian dari proses *Knowledge Discovery in Database* (*KDD*) yang sering melibatkan aplikasi berulang dari metode *data mining* tertentu. *Data mining* menggunakan model yang cocok untuk menentukan pola dari data yang diamati. Model tersebut berperan menyimpulkan, apakah model tersebut mencerminkan pengetahuan berguna atau menarik.

2.4 Analisis Klaster

Analisis klaster adalah salah satu alat bantu pada proses *data mining* yang bertujuan untuk mengelompokkan objek-objek kedalam suatu klaster. Klaster adalah sekelompok atau sekumpulan objek data yang memiliki kemiripan satu sama lain dalam klaster yang sama dan ketidak miripan terhadap objek pada klaster yang berbeda. Objek-objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan setiap objek pada klaster yang sama dan memaksimalkan ketidaksamaan setiap objek pada klaster yang berbeda (Defiyanti, 2017). Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain dan berbeda dengan objek dalam kelompok lain. Lebih besar kemiripannya dalam kelompok dan lebih besar perbedaannya di antara kelompok lainnya (Prasetyo, 2012).

Asumsi dalam analisis kelompok yaitu sampel yang diambil harus mewakili populasi (representatif) dan tidak adanya variabel penelitian yang memiliki hubungan linier yang besar dengan variabel lainnya (nonmultikolinieritas). Menurut Gujarati (2003), multikolinieritas adalah terjadinya hubungan linier yang kuat (hampir sempurna) antara satu variabel dengan variabel yang lainnya. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinieritas adalah dengan melihat nilai korelasi antar variabel penelitian. Menurut Gujarati & Porter (2010), jika pada variabel penelitian tersebut terdapat korelasi yang cukup tinggi yaitu di atas 0,8 maka dapat dikatakan adanya gejala multikolinearitas. Perhitungan koefisien korelasi menggunakan korelasi *pearson* adalah sebagai berikut

$$r_{x_j x_l} = \frac{n\left(\sum_{i=1}^{n} x_{ij} x_{il}\right) - \left(\sum_{i=1}^{n} x_{ij}\right) \cdot \left(\sum_{i=1}^{n} x_{il}\right)}{\sqrt{n\left(\sum_{i=1}^{n} x_{ij}^{\,2}\right) - \left(\sum_{i=1}^{n} x_{ij}\right)^2} \sqrt{n\left(\sum_{i=1}^{n} x_{il}^{\,2}\right) - \left(\sum_{i=1}^{n} x_{il}\right)^2}}, i = 1, 2, ..., n \quad (1)$$

dengan

$r_{x_i x_j}$ = Nilai koefisien korelasi antara variabel x ke-*j* dan variabel x ke-*l*

n = Banyaknya data

p-ISSN 2085-7829
e-ISSN 2798-3455

2.5 *K-Medoids*

K-medoids merupakan metode berbasis partisi yang menggunakan objek representatif yang disebut *medoids* sebagai titik pusat atau *centroid.* Algoritma *k-medoids* melakukan partisi dengan cara meminimalkan jumlah ketidak miripan antara setiap objek *i* dan objek representatif terdekat. Setiap objek yang tersisa dikelompokan dengan objek representatif yang paling mirip dan perhitungan jarak dihitung dari jarak antar masing-masing data (Suyanto, 2017).

Algoritma *k-medoids* mencoba untuk menentukan partisi sebanyak *K* untuk *q* objek. Setelah pemilihan nilai awal *k-medoids* tersebut, dilakukan proses berulang untuk membuat pilihan yang lebih baik dari *medoids* sebelumnya dengan menganalisis semua kemungkinan pasangan objek, sedemikian sehingga satu objek adalah *medoids* dan yang lainnya tidak. Ukuran kualitas pengelompokan terbaik dihitung untuk setiap kombinasi tersebut, pilihan terbaik dari titik dalam satu iterasi dipilih sebagai *medoids* untuk iterasi berikutnya. Adapun tahapan-tahapan dari algoritma *k-medoids* adalah sebagai berikut:

i. Memilih secara acak objek sebanyak *K* sebagai objek representatif *om* (*medoids*).
ii. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *medoids* seperti dinyatakan oleh Persamaan (2) sebagai berikut:

$$d\left(x_{ij},o_{mj}\right)=\sqrt{\left(x_{i1}-o_{m1}\right)^2+\left(x_{i2}-o_{m2}\right)^2+\ldots+\left(x_{iq}-o_{mq}\right)^2}, \tag{2}$$

dengan $d\left(x_{ij},o_{mj}\right)$ adalah jarak dari data ke-*i* pada variable ke-*j* terhadap *medoids* ke-*m* pada variable ke-*j* dimana *m* =1,2,…*K* serta *j* = 1,2,…,q.

iii. yang merupakan jumlah ketidakmiripan dari semua objek ke *medoids* terdekat berdasarkan jarak antara objek terhadap setiap *medoids* yang paling minimum.
iv. Memilih secara acak objek yang tidak representatif o_h (*non-medoids*).
v. Menghitung jarak *euclidean* untuk setiap objek terhadap masing-masing *non- medoids* seperti dinyatakan oleh Persamaan (3) sebagai berikut:

$$d\left(x_{ij},o_{hj}\right)=\sqrt{\left(x_{i1}-o_{h1}\right)^2+\left(x_{i2}-o_{h2}\right)^2+\ldots+\left(x_{iq}-o_{hq}\right)^2}, \tag{3}$$

dengan $d\left(x_{ij},o_{hj}\right)$ adalah jarak dari data ke-*i* pada variabel ke-*j* terhadadap *non-medoids* ke-*h* pada variabel ke-*j* dimana *h*= 1,2..,*K*. (Han & Kamber, 2006)

vi. Menetapkan setiap objek ke gugus yang sesuai dengan *non-medoids* terdekat dan menghitung fungsi objektif yang merupakan jumlah *dissimilarity* dari semua objek ke *non-medoids* terdekat berdasarkan jarak antara objek terhadap setiap *medoids* yang paling minimum.
vii. Menghitung selisih dari fungsi objektif dengan cara mengurangkan fungsi objektif *non-medoids* dengan fungsi objektif *medoids.*
viii. Mengganti *medoids* o_{m} dengan *non-medoids* o_{h} apabila pertukaran semacam mengurangi fungsi objektif.
ix. Mengulangi langkah (4-8) sampai tidak ada lagi perubahan objek representatif.
x. Analisis selesai jika sudah tidak terdapat perubahan objek representatif.

2.6 *Validasi Data Hasil Klasterisasi*

Salah satu metode evaluasi yang dapat digunakan untuk melihat kualitas dan kekuatan klaster adalah metode *silhouette coefficient.* Metode ini merupakan metode validasi klaster yang menggabungkan metode *cohesion* dan *separation*. Tahapan perhitungan *silhouette coefficient* adalah sebagai berikut:

i. Menghitung rata-rata jarak dari suatu data ke-*i* dengan semua data yang berada pada satu klaster yang sama dengan menggunakan Persamaan (4).

$$a_i=\frac{1}{n_p-1}\sum_{r=1}^{n_p-1}d_{i,r}, r\neq i, \tag{4}$$

dengan p merupakan anggota klaster, *p*= 1,2,…,*K*.

ii. Menghitung rata-rata jarak suatu data ke-*i* dengan semua data yang berada pada klaster yang berbeda dengan menggunakan Persamaan (6), kemudian ambil nilai terkecilnya berdasarkan Persamaan (5)

$$b_i=\min\left\{d_i\left(p\right)\right\}, r\neq i, \tag{5}$$

dengan rumus jarak suatu data ke-*i* dengan semua data pada klaster yang berbeda adalah

$$d_i\left(p\right)=\frac{1}{n_p}\sum_{r=1}^{n_p}d_{i,r}, \tag{6}$$

dengan *p*=1,2,..,*K*.

p-ISSN 2085-7829
e-ISSN 2798-3455

iii. Menghitung nilai s*ilhouette coefficient* untuk setiap data ke-*i*

$$SC_1(i) = \frac{b_i - a_i}{\max\{a_i b_i\}}, i = 1, 2, ..., n. \quad (7)$$

nilai *SC* dari sebuah klaster $SC_2(p)$ diperoleh dengan menghitung rata-rata nilai $SC_1(i)$ semua data yang bergabung dalam klaster tersebut dengan menggunakan persamaan (8).

$$SC_2(p) = \frac{1}{n}\sum_{x_i \in C_p}^{n_p} SC_1(i) \quad (8)$$

setelah itu nilai *SC* global diperoleh dengan menghitung rata-rata nilai $SC_2(p)$ dari semua klaster dengan menggunakan persamaan (9).

$$SC = \frac{\sum_{p=1}^{k}\left(n_p \times SC_2(p)\right)}{\sum_{p=1}^{k} n_p} \quad (9)$$

dengan:

a_i : Rata-rata jarak data ke-*i* dengan semua data pada klaster yang sama
b_i : Rata-rata jarak data ke-*i* dengan semua data pada klaster yang berbeda
$SC_1(i)$: Nilai *silhouette coefficient* pada data ke-*i*
$SC_2(p)$: Nilai *silhouette coefficient* pada klaster ke-*p*
SC : Nilai *silhouette coefficient* global
x_i : Data pengamatan ke-*i*
C_p : Klaster ke-*p*
n_p : Jumlah data dalam klaster ke-*p*
K : Banyaknya klaster

Nilai *silhouette coefficient* berdasarkan Kauffman dan Rousseuw (1990) adalah sebagai berikut:

Tabel 1. Nilai *Silhouette Coefficient*

No.	Rentang Nilai *SC*	Keterangan
1	$0{,}7 < SC \leq 1$	*Strong Structure*
2	$0{,}5 < SC \leq 0{,}7$	*Medium Structure*
3	$0{,}25 < SC \leq 0{,}5$	*Weak Strukture*
4	$SC \leq 0{,}25$	*No Structure*

2.7 Pendidikan

Pendidikan adalah segala pengalaman belajar yang berlangsung dalam segala lingkungan dan sepanjang hidup, serta pendidikan dapat diartikan sebagai pengajaran yang diselenggarakan di sekolah sebagai lembaga pendidikan formal (Mudyaharjo, 2001). Wajib belajar adalah program pendidikan minimal yang harus diikuti oleh warga negara Indonesia atas tanggung jawab pemerintah pusat dan pemerintah (Andini, 2017). Salah satu upaya yang dilakukan untuk meningkatkan kualitas pendidikan di Indonesia adalah dengan menjalankan program wajib belajar 12 tahun.

Menurut Peraturan Pemerintah Republik Indonesia Nomor 47 Tahun 2008 tentang Wajib Belajar, Pasal 12 ayat (1) yaitu setiap warga negara Indonesia usia wajib belajar wajib mengikuti program wajib belajar. Program wajib belajar mencangkup pendidikan dasar dan menengah. Menurut Undang Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 17 ayat (2) pendidikan dasar berbentuk Sekolah Dasar (SD) dan Madrasah Ibtidaiyah (MI) atau bentuk lain yang sederajat serta Sekolah Menengah Pertama (SMP) dan Madrasah Tsanawiyah (MTs), atau bentuk lain yang sederajat, lalu Pasal 18 ayat (3) yaitu pendidikan menengah berbentuk Sekolah Menengah Atas (SMA), Madrasah Aliyah (MA), Sekolah Menengah Kejuruan (SMK), dan Madrasah Aliyah Kejuruan (MAK), atau bentuk lain yang sederajat.

3. Metodologi Penelitian

3.1 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

X_1 : Angka Partisipasi Kasar
X_2 : Angka Partisipasi Murni

p-ISSN 2085-7829
e-ISSN 2798-3455

X_3 : Angka Patrisipasi Sekolah
X_4 : Persentase Guru Layak
X_5 : Angka Putus Sekolah

3.2 Tahapan Analisis Data

Langkah-langkah analisis data dalam penelitian ini adalah sebagai berikut:

i. Melakukan analisis statistika deskriptif.
ii. Mengidentifikasi multikolinieritas dengan menggunakan metode korelasi pearson.
iii. Melakukan pengelompokan data dengan menerapkan algoritma k-medoids dengan tahapan berikut:
   a. Memilih objek representative (o_m) secara acak sebanyak K sebagai *medoids* (pusat klaster).
   b. Menghitung jarak *euclidean* untuk setiap objek pengamatan terhadap masing-masing *medoids* berdasarkan persamaan (2).
   c. Menetapkan setiap objek ke gugus yang sesuai dengan *medoids* terdekat dan menghitung fungsi objektif yang merupakan jumlah ketidak miripan dari semua objek ke *medoids* terdekat berdasarkan jarak antara objek terhadap *medoids* yang paling minimum.
   d. Mengganti objek representatif o_m dengan objek yang tidak representatif o_h.
   e. Menghitung jarak *euclidean* untuk setiap objek terhadap masing-masing *non medoids* berdasarkan persamaan (3).
   f. Menetapkan setiap objek ke gugus yang sesuai dengan *non-medoids* terdekat dan menghitung fungsi objektif.
   g. Menghitung selisih dari fungsi objektif dengan cara mengurangkan fungsi objektif *non-medoids* dengan fungsi objektif *medoids*.
   h. Mengganti *medoids* o_m dengan *non-medoids* o_h apabila pertukaran semacam mengurang fungsi objektif.
   i. Mengulangi langkah (d-h) sampai tidak ada lagi perubahan objek representatif.
   j. Menginterpretasikan hasil klaster yang terbentuk
iv. Menghitung nilai SC (*silhouette coefficient*) untuk mengetahui kualitas dari hasil pengelompokan dengan tahapan sebagai berikut:
   a. Menghitung rata-rata jarak dari suatu data ke-i dengan semua data yang berada pada suatu klaster yang sama dengan menggunakan persamaan (4).
   b. Menghitung rata-rata jarak suatu data ke-i dengan semua data yang berada pada klaster yang berbeda degan menggunakan persamaan (6).
   c. Menghitung rata-rata nilai $SC_1(i)$ untuk setiap data ke-i dengan menggunakan persamaan (8).
   d. Menghitung rata-rata nilai $SC_2(p)$ dengan menggunakan persaman (9).
   e. Mengunakan nilai SC global dengan menggunakan persamaan (10).
   f. Menentukkan nilai K optimal berdasarkan nilai *silhouette coefficient* terbesar.

4. Hasil dan Pembahasan

Hasil analisis statistika deskriptif data Indikator Pendidikan di 34 provinsi ada di Indonesia pada tahun 2020 dapat dilihat pada Tabel 2.

Tabel 2. Statistika Deskriptif

Variabel	Banyaknya Data	Minimum	Maksimum	Rata-rata	Simpangan Baku
X_1	34	73,35	98,31	87,08	6,103
X_2	34	44,73	73,45	62,32	6,162
X_3	34	64,83	88,95	74,97	5,953
X_4	34	84,2	97,86	90	3,166
X_5	34	0,18	3,53	1,234	0,715

Berdasarkan Tabel 2 dapat dilihat bahwa banyaknya data pada masing-masing variabel berjumlah 34 data pengamatan. Dimana nilai minimum dari variabel X_1 yaitu 73,35, nilai maksimum 98,31, dengan nilai rata-rata 87,08, dan simpangan baku 6,107. Pada variabel X_2 nilai minimum yaitu 44,73, nilai maksimum 73,45, dengan nilai rata-rata 62,32, dan simpangan baku 6,162. Pada variabel X_3 nilai minimum yaitu 64,83, nilai maksimum 88,95, dengan nilai rata-rata 74,97dan simpangan baku 5,953. Pada variabel X_4 nilai minimum yaitu 84,2, nilai maksimum 97,86, dengan nilai rata-rata 90, dan simpangan baku 3,166. Dan pada variabel X_5 nilai minimum yaitu 0,18, nilai maksimum 3,53, dengan nilai rata-rata 1,234, dan simpangan baku 0,715.

Dari hasil pada analisis 2 klaster, 3 klaster dan 4 klaster didapatkan nilai SC masing-masing dapat dilihat pada Tabel 3.

p-ISSN 2085-7829
e-ISSN 2798-3455

Tabel 3. Perbandingan Hasil Validasi Klaster Berdasarkan nilai *SC Global*

Jumlah Klaster	Klaster	Jumlah Anggota	SC
2	1	14	0,464
	2	20	
3	1	6	0,437
	2	7	
	3	21	
4	1	6	0,383
	2	7	
	3	14	
	4	7	

Berdasarkan Tabel 3 terlihat bahwa penerapan algoritma *k-medoids* pada masing masing klaster menghasilkan jumlah klaster optimal dengan di bentuk menjadi 2 klaster karena memiliki nilai *SC* lebih besar dibandingkan dengan hasil nilai *SC* pada pembentukan klaster 3 dan klaster 4, yang menunjukkan bahwa nilai *SC* yang didapatkan dapat digunakan sebagai pendukung keputusan untuk nilai jumlah klaster paling cocok digunakan.

5. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang diperoleh yaitu klaster optimal yang terbentuk pada pengelompokan provinsi di Indonesia berdasarkan Indikator Pendidikan dengan menggunakan metode *k-medoids* adalah sebanyak 2 klaster dengan nilai *silhouette coefficient* sebesar 0,464. Klaster 1 beranggotakan 14 provinsi dan klaster 2 beranggotakan 20 provinsi.

Referensi

Andini, P. D. (2017). Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pendidikan Formal Wajib Belajar 12 Tahun Menggunakan *Cluster Hierarchy*.

BPS. (2020). *Statistik Pendidikan 2020*. Badan Pusat Statistik, Jakarta.

Defiyanti. (2017). Optimalisasi *K-Medoid* dalam Pengklasteran Mahasiswa Beasiswa dengan *Cubic Clastering Criterion. Jurnal TEKNOSI, 3(1).*

Dunham, M. H. (2003). *Data mining Introductory and Advance Topics*. New Jersey: Prentice Hall.

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceeding of Second International Conference on Knowledge Discovery*, Portland.

Gujarati, D.N. & D.C. Porter. 2010. *Dasar-Dasar Ekonometrika*, Edisi 5. Jakarta: Salemba Empat

Han, J., & Kamber, M. (2006). *Data Mining: Concept and Techniques*. Waltham: Morgan Kauffman Publisher.

Kaufman, L. & Rousseeuw, P. J. (1990). *Finding Group in Data*. New York: John Willey & Sons.

Mudyahardjo. 2001. *Filsafat Ilmu Pendidikan*. Bandung: Remaja Rosdakarya.

Prasetyo, E. (2012). *Data mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset.

Suryabrata, S. (2002). *Metodologi Penelitian*. Yogyakarta: Universitas Gadjah Mada

Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.

Triyanto, W. A. (2015). Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk. Jurnal SIMETRIS. 6(1), 183–188.