

PENGELOMPOKKAN KABUPATEN/KOTA DI JAWA TIMUR BERDASARKAN BANYAKNYA SEKOLAH MENGGUNAKAN ALGORITMA *PARTITIONING AROUND MEDOIDS* (PAM) DAN METODE *DAVIES BOULDIN INDEX*

Muhammad Fikri¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email : muhammadfikri2672@gmail.com

ABSTRAK

Pendidikan merupakan proses kegiatan belajar mengajar pengetahuan, keterampilan, dan kebiasaan kelompok orang dari generasi ke generasi melalui pengajaran, pelatihan, dan penelitian. Di Indonesia jenjang pendidikan meliputi SD, SMP, SMA, dan Perguruan Tinggi. Pada penelitian ini membahas penyebaran sekolah negeri yang tidak merata, salah satu contohnya seperti peserta didik yang tempat tinggalnya masih tidak terjangkau zona, sementara banyak daerah yang pembagian zonasi pada awalnya, didasarkan pada wilayah administrasi kecamatan. Jadi permasalahan ini juga membuat pemerataan sekolah belum maksimal sehingga pemerataan pendidikan juga masih menjadi kendala besar di dunia pendidikan Indonesia. Pada penelitian ini membahas tentang penerapan algoritma PAM untuk mengelompokkan kabupaten/kota berdasarkan banyaknya sekolah di Jawa Timur, data yang digunakan yaitu banyaknya desa/kelurahan yang mempunyai sekolah pada tahun 2018. Hasil dari penerapan algoritma PAM dalam mengelompokkan banyaknya sekolah kabupaten/kota di Jawa Timur diperoleh *Cluster* optimum dengan hasil 2 *Cluster* berdasarkan indeks nilai *Davies bouldin* sebesar 0,553 dengan skenario 2 sampai 10 *Cluster*. Sedangkan jumlah anggota pada masing-masing *Cluster* 1 terdapat 15 kabupaten/kota & *Cluster* 2 terdapat 23 kabupaten/kota.

Kata kunci : Pendidikan, *Clustering*, *Partitioning Around Medoids* (PAM), *Davies Bouldin Index*

ABSTRACT

Education is a teaching and learning activities process of knowledge, skills, and groups of people habits from generation to generation through teaching, training, and research. We are educated to be useful people for the country, homeland and nation. In Indonesia, the level of education includes SD, SMP, SMP, SMA and Higher education. This research discusses the uneven distribution of public shools in each sub-district, while many areas are intially zoned based on the adminitrative area of the sub-district. So, this problem also makes the distribution of schools not optimal so that even education distribution is still a big obstacle in Indonesian education. In this study, it discusses the application of the PAM algorithm to group world/cities based on the number of schools in East Java, the data used is the number of villages/sub-districts that have schools in 2018. The results of the application of the PAM algorithm in grouping the number of district/city schools in East Java obtained the optimum cluster with the results of 2 clusters based on the Davies Bouldin value index of 0.553 with a scenario of 2 to 10 clusters. While the number of members in each cluster 1 consists of 15 district / cities & cluster 2 there are 23 districts/cites.

Keywords : Education, Clustering, Partitioning Around Medoids (PAM), Davies Bouldin Index.

1. Pendahuluan

Pendidikan adalah proses kegiatan belajar mengajar pengetahuan, keterampilan, dan kebiasaan kelompok orang dari generasi ke generasi melalui pengajaran, pelatihan, dan penelitian. Pendidikan dilakukan sebagai usaha sadar diri dan terencana untuk mengembangkan kemampuan diri, Pada saat ini penerimaan murid baru dengan menggunakan sistem zonasi.

Menurut Retno (2019). Penyebaran sekolah negeri yang tidak merata di setiap kecamatan dan kelurahan, sementara banyak daerah yang pembagian zonasi Permasalahan ini membuat pemerataan sekolah belum maksimal sehingga pemerataan pendidikan juga masih menjadi kendala besar di dunia pendidikan Indonesia.

Menurut (Milya, 2018) Menteri Pendidikan dan Kebudayaan juga mengklaim pendidikan di berbagai daerah Indonesia sudah cukup maksimal akan tetapi kualitas setiap sekolah masih sangat minim dan belum merata seperti peserta didik yang tempat tinggalnya masih tidak terjangkau zona sekolah, permasalahan pemerataan kualitas pendidikan yang masih terjadi di Indonesia meliputi kurangnya sarana, prasarana dan penunjang pendidikan salah satu contohnya adalah bangunan sekolah.

Menurut data BPS di Jawa Timur jumlah desa/kelurahan adalah 8.501 di tahun 2018, terdapat 8.443 desa/kelurahan memiliki sekolah SD/MI dan sisanya 58 desa/kelurahan tidak memiliki sekolah SD/MI, 4.696 desa/kelurahan memiliki sekolah SMP dan sisanya 3.805 desa/kelurahan tidak memiliki sekolah SMP, 2.385 desa/kelurahan memiliki sekolah SMA/MA dan sisanya 6.116 tidak memiliki sekolah SMA/MA, 1.531 desa/kelurahan memiliki sekolah SMK dan 6.970 tidak memiliki sekolah SMK.

Karena kurangnya pemerataan sekolah pada kabupaten/kota maka penulis melakukan penelitian terhadap data banyak kabupaten/kota yang mempunyai sekolah berdasarkan 4 variabel yaitu SD/MI, SMP/MTS, SMA, SMK di Jawa Timur pada tahun 2018. Penulis menggunakan algoritma *K-medoids* dengan metode *Davies Bouldin Index (DBI)* untuk menemukan nilai optimum, maka penulis membuat penelitian dengan judul algoritma *Partitioning Around Medoids (PAM)* untuk mengelompokkan kabupaten/kota berdasarkan banyaknya desa/kelurahan di Jawa Timur yang memiliki sekolah dengan metode *Davies Bouldin Index (DBI)*.

2. PENELITIAN TERKAIT

A. Pendidikan

Menurut Melmambessy(2012) Pendidikan adalah proses pengalihan pengetahuan secara sistematis dari seseorang kepada orang lain sesuai standar yang telah ditetapkan oleh para ahli. Dengan adanya pengetahuan tersebut diharapkan dapat merubah sikap tingkah laku, kedewasaan berpikir dan kedewasaan kepribadian ke dalam pendidikan formal dan pendidikan informal.

B. *Data Mining*

Data Mining adalah serangkaian suatu proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006).

C. Clustering

Menurut Rui Xu & Donald (2009) (dalam Hasugian, 2018). *Clustering* (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok.

D. Partitioning Around Medoids (PAM)

Menurut Leonard Kaufman dan Peter J. Rousseeuw (1987) (dalam Santoso, Februariyanti, & Henry, 2016) Algoritma *Partitioning Around Medoids* (PAM) atau dikenal juga dengan *K-Medoids* adalah algoritma pengelompokan yang berkaitan dengan algoritma *K-Means*.

Algoritma *Partitioning Around Medoids* (PAM) menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan n objek menjadi sejumlah k cluster.

Langkah-langkah algoritma *Partitioning Around Medoids* (PAM) menurut Han and Kamber (2006) adalah:

1. Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*.
2. Menempatkan objek *non-medoids* ke dalam *cluster* yang paling dekat dengan *medoids* menggunakan *Euclidian Distance*.
3. Secara acak memilih o_{random} : sebuah objek *non-medoids*.
4. Menghitung total *cost*, S (selisih), dari pertukaran *medoids* o_j dengan o_{random} .
5. Jika $S < 0$ maka menukar o_j dengan o_{random} untuk membentuk sekumpulan k objek baru sebagai *medoids*.
6. Mengulangi langkah 3 hingga 5 sampai tidak ada perubahan *medoids*, sehingga diperoleh *cluster* serta anggota *cluster* masing-masing.

Nilai *cost* dinyatakan dengan persamaan:

$$Total\ Cost = \sum \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \dots \dots \dots (2.1)$$

Dengan:

n = jumlah suatu data

k = indeks data

x_k = nilai atribut ke- k dari x

y_k = nilai atribut ke- k dari y

Nilai S dinyatakan dalam persamaan:

$$S = Total\ Cost\ baru - Total\ Cost\ Lama \dots \dots \dots (2.2)$$

Dengan:

S = Selisih

Total Cost Baru = Jumlah *cost non-medoids*

Total Cost Lama = Jumlah *cost medoids*.

E. Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster* pada suatu metode *clustering*. Dalam penelitian ini DBI digunakan untuk melakukan validasi data pada setiap *cluster*. Adapun langkah-langkah untuk menghitung *Davies Bouldin Index* (Sujacka, 2019), yaitu: Menghitung *Sum of Square Within cluster* (SSW).

Sum of Square Within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah *cluster* ke- i yang dirumuskan sebagai berikut:

$$SSWi = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j c_i) \dots \dots \dots (2.3)$$

Dengan:

m_i = jumlah data dalam *cluster* ke- i

c_i = centroid *cluster* ke- i

$d(x_j C_j)$ = merupakan jarak setiap data terhadap *centroid*

1. Menghitung *Sum of Square Between cluster* (SSB).

Sum of Square Between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j) \dots \dots \dots (2.4)$$

2. Menghitung jumlah *ratio* (rasio).

Setelah nilai SSW dan SSB diperoleh, kemudian melakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{ij} = \frac{SSWi+SSWj}{SSBij} \dots \dots \dots (2.5)$$

3. Menghitung nilai *Davies Bouldin Index*(DBI).

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Indeks* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \dots \dots \dots (2.6)$$

Dengan:

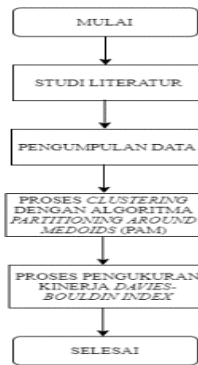
k = jumlah *cluster* yang ditentukan.

Skema *clustering* yang optimal menurut indeks pengukuran *Davies Bouldin* adalah *cluster* yang memiliki nilai indeks terkecil atau minimal (Salazar, dkk. 2002).

F. Rapidminer

Menurut Dennis, dkk. (2013), Perangkat lunak yang bersifat terbuka (*open source*). *Rapid Miner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

G. Diagram Alur Penelitian



Gambar 1. Diagram Alur Penelitian

H. Pengumpulan Data

Pengumpulan data dilakukan berdasarkan studi literatur dari situs resmi Badan Pusat Statistik Jawa Timur yakni <https://jatim.bps.go.id> untuk melengkapi data yang diperlukan. Data yang digunakan dalam penelitian ini adalah data banyaknya kabupaten/kota di Jawa Timur yang memiliki sekolah pada tahun 2018 sebanyak 152 data.

3. HASIL DAN PEMBAHASAN

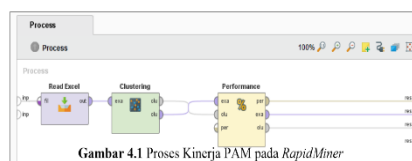
A. Data pengujian

Pada bab ini membahas mengenai hasil yang akan diperoleh dari pengujian data, data tersebut akan diolah menggunakan algoritma *Partitioning Around Medoids* (PAM). Berikut hasil dan pembahasan dari Algoritma *Partitioning Around Medoids* (PAM) untuk mengelompokkan kabupaten/kota di Jawa Timur yang mempunyai sekolah berdasarkan 4 variabel yaitu SD/MI, SMP/MTs, SMA, SMK. Data di peroleh dari Badan Pusat Statistik (BPS) di Jawa Timur pada tahun 2018.

Tabel 4.1 Banyaknya Kabupaten/Kota yang mempunyai sekolah,

| NO | Kabupaten/Kota | SD/MI | SMP/MTs | SMA | SMK |
|-----|----------------|-------|---------|-----|-----|
| 1 | Pacitan | 171 | 96 | 29 | 24 |
| 2 | Ponorogo | 304 | 130 | 69 | 38 |
| 3 | Trenggalek | 157 | 77 | 33 | 28 |
| 4 | Tulungagung | 270 | 98 | 37 | 25 |
| 5 | Blitar | 248 | 114 | 39 | 27 |
| 6 | Kediri | 343 | 130 | 54 | 40 |
| 7 | Malang | 390 | 297 | 108 | 101 |
| 8 | Lumajang | 205 | 148 | 73 | 31 |
| 9 | Jember | 248 | 222 | 127 | 114 |
| 10 | Banyuwangi | 217 | 153 | 97 | 63 |
| ... | ... | ... | ... | ... | ... |
| 38 | Kota Batu | 24 | 17 | 9 | 7 |

B. Rapid Miner



Gambar 4.1 Proses Kinerja PAM pada *RapidMiner*

terdapat beberapa operator yang digunakan. Berikut fungsi dari operator-operator tersebut:

- Read Excel*: Operator ini dapat digunakan untuk memuat data dari *sheet* pada *Microsoft Excel*.
- Clustering*: Operator ini melakukan pengelompokkan menggunakan metode *clustering k-medoids*. Pada penelitian ini jumlah *cluster* yang akan digunakan sebagai pengujian yaitu 2 sampai 10 *cluster*.
- Performance*: Operator *performance* yang digunakan yaitu *cluster distance performance* dimana operator ini digunakan untuk evaluasi kinerja metode *k-medoids* berdasarkan nilai *davies bouldin index*.

C. Penentuan Jumlah Cluster Optimum

Setelah melalui proses *cluster* dengan menggunakan algoritma PAM, kemudian dilakukan proses dengan menggunakan metode DBI untuk penentuan *cluster optimum*.

Tabel 1. Nilai *Davies Bouldin Indeks*

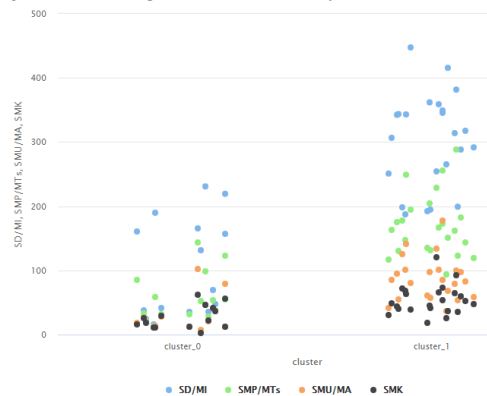
| Cluster | Nilai DBI | Anggota Cluster |
|---------|-----------|---|
| 2 | 0,553 | Cluster 1 : 15 dan Cluster 2 : 23 |
| 3 | 0,741 | Cluster 1 : 15, Cluster 2 : 15 , dan Cluster 3 : 8 |
| 4 | 0,994 | Cluster 1 : 8, Cluster 2 : 15, Cluster 3 : 6, dan Cluster 4 : 9 |
| 5 | 0,872 | Cluster 1 : 8, Cluster 2 : 7, Cluster 3 : 5, Cluster 4 : 12, dan Cluster 5 : 6 |
| 6 | 0,741 | Cluster 1 : 7, Cluster 2 : 8, Cluster 3 : 2, Cluster 4 : 12, Cluster 5 : 4, dan Cluster 6 : 5 |
| 7 | 0,814 | Cluster 1 : 7, Cluster 2 : 8, Cluster 3 : 5, Cluster 4 : 2, Cluster 5 : 5, Cluster 6 : 3, dan Cluster 7 : 8 |
| 8 | 0,904 | Cluster 1 : 8, Cluster 2 : 2, Cluster 3 : 7, Cluster 4 : 5, Cluster 5 : 5, Cluster 6 : 3, Cluster 7 : 2, dan Cluster 8 : 6 |
| 9 | 0,838 | Cluster 1 : 5, Cluster 2 : 8, Cluster 3 : 5, Cluster 4 : 4, Cluster 5 : 2, Cluster 6 : 4, Cluster 7 : 3, Cluster 8 : 4, dan Cluster 9 : 3 |
| 10 | 0,764 | Cluster 1 : 5, Cluster 2 : 8, Cluster 3 : 5, Cluster 4 : 4, Cluster 5 : 1, Cluster 6 : 4, Cluster 7 : 1, Cluster 8 : 3, Cluster 9 : 4, dan Cluster 10 : 3 |

Dari Tabel 1 menunjukkan nilai *Davies Bouldin index* dengan pengujian skenario *cluster* 2 sampai dengan 10 *cluster*, maka *cluster* yang paling optimum berdasarkan nilai *Davies Bouldin indeks* terkecil yaitu 2 *cluster* dengan nilai *Davies Bouldin indeks* yaitu 0,553.

D. Hasil Cluster Profiling

Didapat *cluster* optimum berada pada *cluster* 2. Untuk mengetahui karakteristik dari masing-masing *cluster*, maka dilakukan *profiling Cluster* pada data banyaknya jumlah sekolah di setiap kabupaten/kota yang masuk dalam kelompok *cluster* 1, dan *cluster* 2 Berdasarkan perhitungan yang telah dilakukan sebelumnya. Dari hasil penyebaran data pemerataan

sekolah di dapatkan hasil fitur SD, SMP, SMA, SMK pada *Cluster 2* memiliki jumlah bangunan sekolah yang tinggi dibandingkan dengan *Cluster 1* yang jumlah bangunan sekolahnya rendah.



Gambar 4.13 Diagram Hasil *Cluster Profiling* pada 2 *Cluster*.

Dari hasil penyebaran data pemerataan sekolah di dapatkan hasil fitur SD, SMP, SMA, SMK pada *Cluster 2* memiliki jumlah bangunan sekolah yang tinggi dibandingkan dengan *Cluster 1* yang jumlah bangunan sekolahnya rendah.

4. Kesimpulan

1. Penggunaan algoritma PAM untuk proses *clustering* data jumlah banyaknya sekolah di Jawa Timur dengan 38 kabupaten/kota berdasarkan 4 variabel dari validasi *cluster* menggunakan metode DBI didapat nilai *cluster* yang paling optimum adalah *cluster 2* yang memiliki nilai terendah pada DBI yaitu 0.553.
2. Karakteristik pada *Cluster 1* di dominasi oleh kabupaten/kota yang tergolong rendah yang beranggotakan 15 kabupaten/kota yaitu Pacitan, Trenggalek, Situbondo, Madiun, Magetan, Kediri, Blitar, Malang, Probolinggo, Pasuruan, Mojokerto, Madiun, Surabaya, dan Batu. sedangkan *Cluster 2* didominasi oleh kabupaten/kota yang tergolong tinggi dengan beranggotakan 23 kabupaten/kota yaitu Ponorogo, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, dan Sumenep.

5. Saran

1. Dalam proses validasi *cluster* untuk mencari *cluster optimum* dapat menggunakan alternatif lain seperti *Silhouette Coefficient*, *Calinski-Harabasz index* dan *Elbow*.
2. Pada penelitian ini dapat dikembangkan dengan menambahkan attribute dan variabel pada data yang digunakan, dan perhitungan manual *cluster* pada algoritma *Partitioning Around Medoids (PAM)* pada Microsoft Excel dapat digunakan sebagai alternative pembanding hasil *cluster* pada *tools RapidMiner*.

DAFTAR PUSTAKA

- Anonim. <https://indonesia.go.id/province/jawa-timur>. 2017. Diakses pada 6 April 2020.
- Anonim. <http://ilmuteknologyindustri.blogspot.com/2017/03/pengertian-data-mining-dan-jenis-fase.html>. Diakses pada 12 April 2020.
- Asri, M.F., 2018 “*PENGARUH SISTEM ZONASI TERHADAP KUALITAS PENDIDIKAN DI INDONESIA*”. Jurnal DDIP Program Studi Pendidikan Bahasa Indonesia Jurusan Bahasa dan Sastra Indonesia dan Daerah Fakultas Bahasa dan Seni, Universitas Negeri Padang.
- Christie, A.D., Baskoro, D.A., Ambarwati, L., & Wicaksana, I.W.S. 2013 “*Belajar Data Mining Dengan Rapid Miner*”. Jakarta: Gramedia Pustaka.
- Hasugian, P. 2018. “Penerapan data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Means”. *Jurnal Mantik Penusa. Volume 2, No. 2*. Teknik Informatika STMIK Pelita Nusantara, Medan.
- Hermawati, F. 2013. *Data Mining*. Yogyakarta: Andi Offset. Tersedia di <https://ejournal.bsi.ac.id/ejurnal/index.php/jtk/article/view/2560/1918> Diakses pada tanggal 05 April 2020.
- Han, J. dan Kamber, M. 2001. “*Data Mining: Concepts and Techniques*”. Simon Fraser University. USA: Academic Press.
- Moses, M. 2012. “Analisis Pengaruh Pendidikan, Pelatihan, Dan Pengalaman Kerja Terhadap Produktivitas Kerja Pegawai Dinas Pertambangan Dan Energi Provinsi Papua”. *Media Riset Bisnis & Manajemen Vol. 12 No. 01 April 2012*. STIE Port Numbay, Jayapura.
- Meilani, B.D., Asadulloh M. 2015. “*Data Mining Untuk Menggali Pola Mahasiswa Baru Menggunakan Metode Frequent Pattern Growth. Surabaya: (STUDI KASUS : INSTITUT TEKNOLOGI ADHI TAMA SURABAYA)*”. Institut Teknologi Adhi Tama Surabaya, Surabaya.
- Pramesti, D. F, Furqon, M., dan Dewi, C. 2017. “Implementasi Metode KMedoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)”. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 1, No. 9*. Fakultas Ilmu Komputer. Universitas Brawijaya, Malang. Tersedia di <https://virtue.dewadg.id/jdih-bulungan-dev-files/f7f44049-1106-11ea8b12-02420aff01b3.pdf> Diakses pada 05 April 2020.
- Sujacka, Retno. 2019. Peningkatan Akurasi Algoritma K-Means dengan Clustering Purity sebagai Titik Pusat Cluster Awal (Centroid).[pdf] repository.usu.ac.id.
- Prasetyo, E., Asroni., Fitri H. 2018. “*Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta. (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)*”. *Semeta Tenika* Vol. 21. No. 1, 60-64, Mei 2018. Universitas Muhammadiyah Yogyakarta, Yogyakarta.

Santoso, D.B., Februariyanti, & Henry. 2016. “*Algoritma Partitioning Around Medoids (PAM) Clustering untuk Melihat Gambaran Umum Skripsi Mahasiswa*”. *Jurnal Teknologi Informasi DINAMIK* Vol. 21, No. 1, Januari 2016. Universitas Stikubank, Semarang.

Sujacka, Retno. 2019. Peningkatan Akurasi Algoritma K-Means dengan Clustering Purity sebagai Titik Pusat Cluster Awal (Centroid).[pdf] repository.usu.ac.id.

Wanto A., Damanik, I.I.P., Solikhun., Saragih, I.S., Parlina, I., Suhendro, D. 2019. “*Algoritma K-medoids untuk mengelompokkan desa yang memiliki infrastruktur di Indonesia*”. *Seminar Nasional Riset Information Science (SENARIS)*. Pematangsiantar, Indonesia.

Waworuntu, M.N.V., & Amin, M.F. 2018. “*Penerapan Metode K-means pemetaan calon penerima jamkesda*”. *Jurnal Ilmu, Ilmu Komputer*. Vol. 05 No. 02 September 2018. Banjarbaru.

