



Perbandingan Algoritma K-Means dan Fuzzy C-Means untuk Clustering Puskesmas Berdasarkan Gizi Balita di Surabaya

Alvin Kurnia Rahmansyah¹, Ariiq Thufeil Shidqul Aziz², Nur Novianto³, Dwi Rolliawati⁴.

¹²³⁴Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Islam Negeri Sunan Ampel, Jl. Ahmad Yani No.117, Jemur Wonosari, Kec. Wonocolo, Kota SBY, Jawa Timur 60237.

ABSTRACT

Diseased environments and low public awareness of environmental hygiene in several cities in Indonesia make many people, especially young children, susceptible to various diseases. The 63 Puskesmas in Surabaya City have weighed 148,720 toddlers. There are 4.2% of toddlers with malnutrition, 0.1% of toddlers with malnutrition and 4.5% of toddlers with stunting. Clustering is a technique in data mining that aims to group objects (data) into several clusters or groups so that similar objects are united into the same cluster. This method is to implement, test, and evaluate the K-Means and Fuzzy C-Means algorithms in clustering. so the results of this study are to classify health centers based on toddler nutrition in the city of Surabaya. The best silhouette coefficient results from K-Means with the Normalization process is 0.51833215835383. While the best result on the Fuzzy C-Means algorithm with the Normalization process is 0.49666983222478. So it can be concluded that the K-Means Algorithm is better at clustering Puskesmas based on toddler nutrition in Surabaya City.

Keywords: Clustering, Algoritma, K-Means, Fuzzy, C-Means

ABSTRAK

Lingkungan yang berpenyakit dan rendahnya kesadaran masyarakat akan kebersihan lingkungan di beberapa kota di Indonesia membuat banyak orang terutama anak kecil rentan terhadap berbagai penyakit. Pada 63 Puskesmas yang ada di Kota Surabaya telah menimbang 148720 balita. Terdapat 4,2% balita dengan gizi kurang, 0,1% balita gizi buruk dan 4,5% balita gizi stunting. Clustering adalah suatu teknik dalam data mining yang bertujuan untuk mengelompokkan objek (data) ke dalam beberapa cluster atau kelompok sehingga objek yang serupa disatukan ke dalam cluster yang sama. metode ini adalah untuk melakukan penerapan, pengujian, serta evaluasi terhadap algoritma K-Means dan Fuzzy C-Means dalam melakukan clustering. sehingga hasil dari penelitian ini adalah untuk mengelompokkan Puskesmas berdasarkan gizi balita yang ada di Kota Surabaya. Hasil terbaik silhouette coefficient dari K-Means dengan proses Normalization adalah 0.51833215835383. Sedangkan hasil terbaik pada algoritma Fuzzy C-Means dengan proses Normalization adalah 0.49666983222478. Sehingga dapat disimpulkan bahwa pada clustering Puskesmas berdasarkan gizi balita Kota Surabaya algoritma K-Means lebih baik.

Kata Kunci: Clustering, Algoritma, K-Means, Fuzzy, C-Means

1. PENDAHULUAN

Indikator kesehatan anak umur lima tahun merupakan masa penting saat anak memerlukan nutrisi yang cukup sebagai penunjang pertumbuhan fisiknya. Salah satunya adalah nilai gizi, nilai gizi balita dapat dihitung dengan indikator tinggi badan berdasarkan umur, berat badan berdasarkan umur, serta berat badan berdasarkan tinggi badan dengan memasukkan nilai tinggi badan, berat badan dan umur. Lingkungan yang berpenyakit dan rendahnya kesadaran masyarakat akan kebersihan lingkungan di beberapa kota di Indonesia membuat banyak orang terutama anak kecil rentan terhadap berbagai penyakit. Balita yang rawan sakit cenderung kurang makan dibandingkan dengan balita yang jarang sakit.

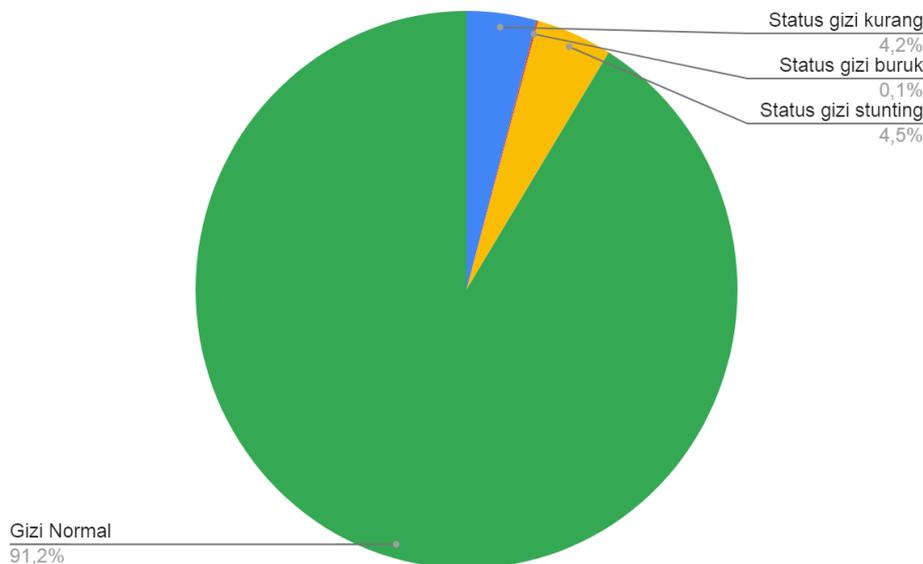
Malnutrisi pada anak kecil berdampak negatif pada pertumbuhan fisik dan mental, yang pada gilirannya mempersulit beberapa proses pembelajaran yang dilakukan anak kecil, seperti berbicara, berjalan, makan dan lain-lain. Balita yang kekurangan gizi cenderung memiliki kecerdasan intelektual (IQ) yang lebih rendah daripada balita yang sehat. Hal ini karena malnutrisi pada anak menghambat sintesis protein DNA, yang menghambat pembentukan sel otak, yang selanjutnya menghambat perkembangan otak. Informasi yang cepat dan akurat tentang nilai gizi anak kecil membantu mencegah kekurangan gizi. Sehingga dapat dijadikan acuan pengolahan yang tepat dalam klasifikasi dataset gizi bayi Puskesmas Kota Surabaya di opendata.surabaya.go.id sebagai rekomendasi.

Kabupaten Surabaya sebagai salah satu kota di negara Indonesia juga tidak luput dari masalah gizi balita. Berdasarkan data pada opendata.surabaya.go.id dari 63 Puskesmas yang ada di Kota Surabaya telah menimbang 148720 balita. Terdapat 4,2% balita dengan gizi kurang, 0,1% balita gizi buruk dan 4,5% balita gizi stunting. Berikut grafik persentase gizi balita pada Puskesmas di Kota Surabaya.

Pada penelitian terdahulu mengenai klustering karyawan berdasarkan nilai kinerja menggunakan algoritma K-Means dan algoritma Fuzzy C-Means. Penelitian tersebut menunjukkan bahwa Algoritma Fuzzy C-Means lebih baik digunakan untuk melakukan

pengelompokan data nilai kinerja karyawan di PT. XYZ karena nilai akurasi bernilai mendekati 100% yaitu sebesar 76% dibandingkan dengan nilai akurasi algoritma K-Means sebesar 44%

Pada penelitian lainnya untuk mengelompokkan daerah penyebaran covid-19 di Indonesia menggunakan algoritma K-Means dan algoritma Fuzzy C-Means. Hasil penelitian tersebut menunjukkan bahwa pengujian clustering k-Means menghasilkan nilai akurasi yang lebih besar sebesar 1,165219 dibandingkan Fuzzy c-Means. Sehingga clustering k-Means diambil untuk menentukan pengelompokan daerah penyebaran Covid-19 per provinsi Indonesia.



Gambar 1. Grafik Persentase gizi balita Kota Surabaya

2. TINJAUAN PUSTAKA

Clustering adalah suatu teknik dalam data mining yang bertujuan untuk mengelompokkan objek (data) ke dalam beberapa cluster atau kelompok sehingga objek yang serupa disatukan ke dalam cluster yang sama, dan yang berbeda harus menjadi bagian dari cluster yang berbeda. Berdasarkan perspektif optimasi, tujuan utamanya untuk memaksimalkan homogenitas (internal) dalam sebuah cluster dan heterogenitas (eksternal) di antara cluster yang berbeda [1].

Proses pengelompokan mencoba untuk meminimalkan pembentukan fungsi tujuan yang telah ditentukan sebelumnya selama proses pengelompokan, terutama untuk meminimalkan perubahan dalam suatu kluster dan memaksimalkan perubahan antar kluster. Dengan menggunakan pengelompokan, akan mempermudah mengklasifikasikan bagian yang padat, menemukan pola distribusi yang umum, dan menemukan hubungan serupa antar atribut data.

Salah satu metode clustering yaitu K-Means, yaitu metode analisis cluster yang bertujuan untuk memecah objek menjadi k cluster kemudian diamati dimana setiap objek cluster diperoleh melalui rata-rata terdekat. Algoritma ini merupakan salah satu pembelajaran yang terkenal sederhana dan mudah dipelajari sebagai pemecahan masalah pengelompokan dari sebuah dataset [2]. Algoritma K-Means adalah algoritma evolusi yang mana metode operasinya memiliki artian yang serupa dengan nama algoritmanya. K-Means ini mengelompokkan pengamatan ke dalam kelompok k, di mana k merupakan parameter input. Tiap data kemudian ditetapkan pada setiap pengamatan cluster berdasarkan kedekatan pengamatan nilai rata-rata cluster. Nilai rata-rata pada cluster kemudian dihitung secara berulang pada proses awal [3].

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \dots \dots \dots (1)$$

Metode Clustering lain selain K-Means adalah Fuzzy C-Means yaitu teknik klasterisasi yang kerap diimplementasikan dalam metode clustering [9]. Fuzzy C-Means menggunakan penggabungan Fuzzy, dimana setiap data bisa dipunyai oleh sebagian cluster, dan setiap cluster memiliki derajat keanggotaan yang berbeda. Algoritma Fuzzy C-Means adalah algoritma berulang yang mengiterasi proses clustering data. Penggunaan algoritma fuzzy C-Means bertujuan untuk mencari pusat titik cluster, kemudian digunakan sebagai pencarian data yang masuk kedalam cluster [10].

$$C_j = \frac{\sum u_{ij}^m x_j}{\sum u_{ij}^m} \dots \dots \dots (2)$$

$$u_{ij}^m = \frac{1}{\sum \left(\frac{|x_i - c_j|}{|x_i - c_k|} \right)^{\frac{2}{m-1}}} \dots \dots \dots (3)$$

KNIME memiliki platform integrasi terbaik untuk pelaporan dan analisis data. Perangkat lunak sumber terbuka untuk analisis data, pelaporan, dan platform integrasi yang dapat menghubungkan berbagai komponen dalam pembelajaran mesin dan penambangan data menggunakan konsep grafik data. Keuntungan menggunakan perangkat lunak KNIME adalah dapat memodelkan semua pembelajaran mesin dalam bahasa pemrograman yang berbeda.

Preprocessing adalah teknik ekstraksi data pertama untuk mengubah data mentah atau umumnya raw data yang dikumpulkan dari berbagai sumber menjadi data yang lebih bersih yang dapat digunakan untuk diproses lebih lanjut[5]. Tujuan utama pemrosesan data adalah pembersihan data, yaitu, mengisi nilai-nilai yang hilang, memperhalus noise data, mengidentifikasi dan menghapus outlier, dan menyelesaikan ketidakkonsistenan. Selain transformasi data, normalisasi dan penggabungan. Akhirnya ada reduksi data, yang menawarkan volume representatif yang lebih kecil tetapi memberikan hasil analisis yang sama atau serupa, dan terakhir diskretisasi data, reduksi data tetapi sangat penting, terutama untuk data numerik [6].

3. METODOLOGI PENELITIAN

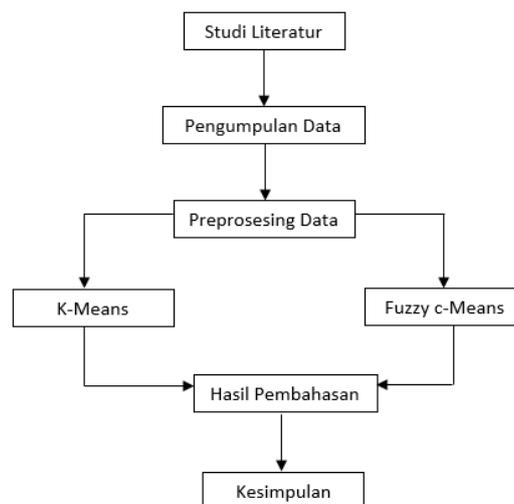
3.1. Sumber Data Penelitian

Sumber data yang digunakan dalam penelitian ini adalah data primer dari website *opendata.surabaya.go.id*. Data ini berasal dari 63 Puskesmas Surabaya mengenai gizi balita yang ada di Kota Surabaya. Data ini diambil pada tahun 2021 pada 63 Puskesmas yang ada di Kota Surabaya.

3.2. Metodologi Penelitian

Metode yang digunakan pada penelitian ini adalah terapan deskriptif. Tujuan dari metode ini adalah untuk melakukan penerapan, pengujian, serta evaluasi terhadap algoritma K-Means dan Fuzzy C-Means dalam melakukan clustering. sehingga hasil dari penelitian ini adalah untuk mengelompokkan Puskesmas berdasarkan gizi balita yang ada di Kota Surabaya.

3.3. Tahapan Penelitian



Gambar 2. Tahapan penelitian

Pada tahapan Pengumpulan Data, data yang diambil adalah Data Puskesmas tentang jumlah gizi balita yang ada di Kota Surabaya pada tahun 2021. Atribut yang dipilih dalam penelitian ini diantaranya seperti ditunjukkan pada tabel 1 berikut.

Tabel 1. Atribut data

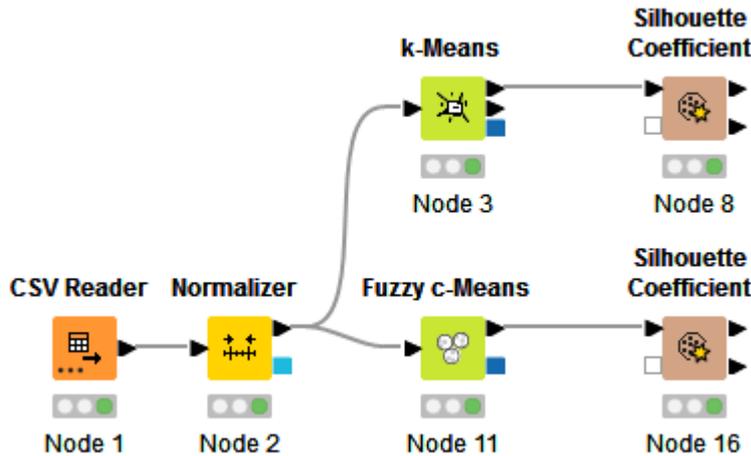
No	Parameter
1	Nama Sarana Kesehatan
2	Status Gizi Kurang
3	Status Gizi Buruk
4	Status Gizi Stunting

Pada Preprocessing dilakukan manipulasi atau penghapusan data sebelum digunakan untuk memastikan atau meningkatkan kinerja, dan merupakan langkah penting dalam proses penambangan data. sehingga penelitian ini menggunakan kolom Nama Puskesmas, gizi kurang, gizi buruk, dan gizi stunting. selanjutnya dilakukan normalisasi terhadap data tersebut sehingga rentang data tidak terlalu jauh.

Langkah selanjutnya adalah menerapkan algoritma tersebut pada data Puskesmas dengan menggunakan dua algoritma yaitu K-Means dan Fuzzy C-Means. Aplikasi ini menggunakan perangkat lunak Knime, perangkat lunak data mining dan analisis prediktif. Knime sangat mudah digunakan karena fitur drag and drop operator yang disediakan. Sebelum mengimplementasikan algoritma, terlebih dahulu dilakukan pengolahan data yaitu raw mentah yang dikumpulkan biasa disebut data mentah diubah menjadi data yang lebih bersih yang dapat digunakan untuk diproses lebih lanjut.

4. HASIL DAN PEMBAHASAN

Proses analisis komparatif algoritma k-means dan fuzzy c-means dilakukan dengan menggunakan perangkat lunak KNIME (Konstanz Information Miner). KNIME adalah perangkat lunak untuk mengintegrasikan, memproses, dan menganalisis data dari berbagai sumber. KNIME menyediakan berbagai fitur visualisasi yang memungkinkan pengguna membuat aliran analisis data dengan mudah tanpa baris kode.



Gambar 3. Workflow KNIME

Pada gambar 3 merupakan tampilan dari workflow pada KNIME yang dimulai dengan membaca data menggunakan node CSV Reader. Sebelum mulai diproses, perlu dilakukan tahap pre-processing data terlebih dahulu node Normalizer. Selanjutnya, dilakukan analisis menggunakan masing-masing node berbeda sesuai dengan algoritma terkait. Dari masing-masing algoritma tersebut, diperlukan validasi menggunakan node Silhouette Coefficient.

Tabel 2. Silhouette Coefficient K-Means

Preprocessing	Cluster							
	2	3	4	5	6	7	8	
Standard Scaler	0.387	0.3119	0.3313	0.3300	0.3093	0.3381	0.3190	
Normalization	0.5183	0.4902	0.5170	0.4560	0.4798	0.4297	0.4487	
Min Max Scaler	0.40785	0.3335	0.3349	0.3376	0.2733	0.2529	0.2614	

Tabel 2 menunjukkan silhouette score untuk algoritma K-Means dari i dari cluster 2 hingga cluster 8. Hasil silhouette score tertinggi dari masing-masing perlakuan preprocessing data pada algoritma K-Means berturut-turut adalah 0.3381 untuk Standard scaler, 0.5183 untuk Normalization, dan 0.40785 untuk Min Max Scaler. Sehingga akan dikelompokkan menjadi 2 cluster dengan preprocessing Normalization.

Tabel 3. Silhouette Coefficient Fuzzy C-Means

Preprocessing	Cluster							
	2	3	4	5	6	7	8	
Standard Scaler	0.3660	0.3035	0.2422	0.2191	0.2303	0.2295	0.2171	
Normalization	0.4966	0.4846	0.4202	0.4505	0.3857	0.4170	0.4487	
Min Max Scaler	0.3916	0.3225	0.2679	0.2317	0.2507	0.2243	0.2211	

Tabel 3 menunjukkan silhouette score untuk algoritma Fuzzy C-Means dari cluster 2 hingga cluster 8. Hasil silhouette score tertinggi dari masing-masing perlakuan preprocessing data pada algoritma K-Means berturut-turut adalah 0.3660 untuk Standard scaler, 0.4966 untuk Normalization, dan 0.3916 untuk Min Max Scaler. Semua perolehan skor tertinggi terjadi pada cluster 2.

Tabel 4. Perbandingan Algoritma

Metode	Preprocessing	Cluster	Silhouette Score
K-Means	Normalization	2	0.51833215835383
Fuzzy C-Means	Normalization	2	0.49666983222478

Pada Tabel 4 menunjukkan bahwa algoritma K-Means n dengan menggunakan 2 cluster memiliki nilai Silhouette Score yang lebih tinggi dibanding dengan algoritma fuzzy C-Means meskipun dan memiliki selisih yang tidak terlalu jauh.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan dengan algoritma K-Means dan Fuzzy c-Means yang keduanya dibagi menjadi dua kluster, k-Means menghasilkan 18 puskesmas menjadi cluster 0 dan 45 Puskesmas menjadi cluster 1. Sedangkan Fuzzy c-Means menghasilkan 20 puskesmas menjadi cluster 0 dan 43 Puskesmas menjadi cluster 1

Hasil terbaik silhouette coefficient dari K-Means dengan proses Normalization adalah 0.51833215835383. Sedangkan hasil terbaik pada algoritma Fuzzy C-Means dengan proses Normalization adalah 0.49666983222478. Sehingga dapat disimpulkan bahwa pada clustering Puskesmas berdasarkan gizi balita Kota Surabaya algoritma K-Means lebih baik.

5.2. Saran

Saran untuk penelitian selanjutnya adalah mencari silhouette coefficient yang lebih tinggi dengan mencoba menggabungkan atribut yang dipilih, atau dengan mengurangi atau menambah atribut yang digunakan. Selain itu juga menggunakan algoritma untuk clustering yang lainnya.

DAFTAR PUSTAKA

- [1]. Adha, R., Nurhaliza, N., Sholeha, U., & Mustakim, M. (2021). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 18(2), 206-211
- [2]. Agustina, N., & Prihandoko, P. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(3), 621-626.
- [3]. Buaton, R., Sundari, Y., & Maulita, Y. (2016). Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean. *MEANS (Media Informasi Analisa dan Sistem)*, 47-53.
- [4]. Dwididanti, S., & Anggoro, D. A. (2022). Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit. *Emitor: Jurnal Teknik Elektro*, 22(2).
- [5]. Fakhri, D. A., & Defit, S. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering. *Jurnal Informasi dan Teknologi*, 160-166.
- [6]. Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan algoritma k-means dan k-medoids dalam klusterisasi produk asuransi perusahaan nasional. *Jurnal Ilmiah Fifo*, 11(2), 208-214.
- [7]. Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangasambung. *Journal of Applied Informatics and Computing*, 5(2), 109-116.
- [8]. Fatonah, N. S., & Pancarani, T. K. (2022). Analisa Perbandingan Algoritma Clustering Untuk Pemetaan Status Gizi Balita Di Puskesmas Pasir Jaya. *Konvergensi Teknologi Informasi & Komunikasi*, 18(1), 1-9.
- [9]. Firdaus, H. S., Nugraha, A. L., Sasmito, B., & Awaluddin, M. (2021). Perbandingan metode fuzzy c-means dan k-means untuk pemetaan daerah rawan kriminalitas di kota semarang. *Elipsoida: Jurnal Geodesi dan Geomatika*, 4(01), 58-64.
- [10]. Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275-282
- [11]. Malik, R. A., Defit, S., & Yuhandri, Y. Perbandingan Algoritma K-Means Clustering Dengan Fuzzy C-Means Dalam Mengukur Tingkat Kepuasan Terhadap Televisi Dakwah. *vol*, 3, 10-21.
- [12]. Nastiti, P. R., & Putra, A. B. W. (2017). Perbandingan Algoritma K-Means dan Fuzzy C-Means Clustering Untuk Kualifikasi Data Kinerja Dosen di Jurusan Teknologi Informasi POLNES.
- [13]. Pramitasari, A. E., & Nataliani, Y. (2021). Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(3), 1119-1132.
- [14]. Putri, A. L. R., & Dwidayati, N. (2021). Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran COVID-19 Indonesia. *Unnes Journal of Mathematics*, 50-55.
- [15]. Ramadhan, A., Efendi, Z., & Mustakim, M. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 219-226).

- [16].Sangga, V. A. P. (2018). Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015.
- [17].Saputra, P. R. N., & Chusyairi, A. (2020). Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(6), 1077-1084.
- [18].Selviana, N. H. (2016). Analisis Perbandingan k-means dan fuzzy c-means untuk pemetaan motivasi belajar mahasiswa. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*.
- [19].Suryanto, J. (2017). Analisa Perbandingan Pengelompokkan Curah Hujan 15 Harian Provinsi Diy Menggunakan Fuzzy Clustering Dan K-Means Clustering. *Jurnal AGRIFOR*, 16.
- [20].Tamtelahitu, T. M. (2020). Komparasi Algoritma Clustering Dengan Dataset Penyebaran Covid-19 Di Indonesia Periode Maret-Mei 2020. *JURNAL TEKNOLOGI TECHNOSCIENTIA*, 27-34.