

**IMPLEMENTASI ALGORITMA *K-MEANS* DENGAN MENGGUNAKAN
METODE *ELBOW* DAN *SILHOUETTE* PADA KLASTERISASI TINGKAT
GIZI BALITA DI PUSKESMAS MARGOYOSO TAHUN 2020**

(Skripsi)

Oleh

ATIF SETIAWAN



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

IMPLEMENTATION OF THE *K-MEANS* ALGORITHM USING *ELBOW* AND *SILHOUETTE* METHOD ON LEVEL CLASSTERIZATION TODDLER NUTRITION AT MARGOYOSO HEALTH CENTER IN 2020

By

ATIF SETIAWAN

Nutrition is a measure to determine the health status of a toddler or child. Good child development depends on adequate nutrition obtained. Research by clustering children based on nutritional levels is an effort to find out and measure how well parents provide their children with nutritional needs. Clustering using the K-Means algorithm can be used to perform clustering by looking at the size of the closeness of the data. The measure of closeness of the data in question is the age of the child himself. In clustering using this algorithm, the elbow and silhouette methods are also used, where these two methods are considered the best for clustering the nutritional status of toddlers. By using this method, the optimal number of clusters is 2 clusters. This means that from this 288 data sets, 2 clusters of toddlers were obtained, with the first cluster of toddlers with adequate nutritional status, and the second cluster of toddlers with good nutritional status. A silhouette value of 0.5 was also obtained, which means that data clustering has produced a good structure. So it can be concluded that clustering using the K-Means algorithm produces 2 optimal clusters which are determined using the elbow method, and produces a silhouette value of 0.5, where the data structure of grouping these 2 clusters is good.

Keywords : K-Means Algoritihm, Nutrition, Cluster, Elbow Method, Silhouette Score

ABSTRAK

IMPLEMENTASI ALGORITMA *K-MEANS* DENGAN MENGGUNAKAN METODE *ELBOW* DAN *SILHOUETTE* PADA KLASTERISASI TINGKAT GIZI BALITA DI PUSKESMAS MARGOYOSO TAHUN 2020

Oleh

Atif Setiawan

Gizi merupakan suatu ukuran untuk menentukan status kesehatan bagi seorang balita atau anak. Perkembangan anak yang baik bergantung dari kecukupan gizi yang diperoleh. Penelitian dengan pengklasteran anak berdasarkan tingkat gizi adalah salah satu upaya untuk mengetahui dan mengukur seberapa baik pemenuhan kebutuhan gizi yang diberikan orang tua kepada anaknya. Pengklasteran dengan menggunakan algoritma K-Means dapat digunakan untuk melakukan pengklasteran dengan melihat ukuran kedekatan data. Ukuran kedekatan data yang dimaksud adalah usia dari anak itu sendiri. Dalam pengklasteran menggunakan algoritma ini digunakan pula metode elbow dan silhouette dimana kedua metode ini dirasa paling baik untuk melakukan pengklasteran status gizi balita. Dengan menggunakan metode ini diperoleh jumlah kluster optimal sebanyak 2 kluster. artinya dari data sebanyak 288 data ini diperoleh 2 kluster balita dengan kluster pertama balita dengan status gizi cukup, dan kluster 2 balita dengan status gizi baik. Diperoleh juga nilai silhouette sebesar 0,5 yang artinya pengklasteran data sudah menghasilkan struktur yang baik. Sehingga dapat disimpulkan bahwa pengklasteran menggunakan algoritma K-Means menghasilkan kluster optimal sebanyak 2 yang ditentukan dengan menggunakan metode elbow, dan menghasilkan nilai silhouette sebesar 0,5 dimana struktur data dari pengelompokan 2 kluster ini sudah baik.

Kata Kunci : Algoritma K-Means, Gizi, Kluster, Metode Elbow, Nilai Silhouette

**IMPLEMENTASI ALGORITMA *K-MEANS* DENGAN MENGGUNAKAN
METODE *ELBOW* DAN *SILHOUETTE* PADA KLASTERISASI TINGKAT
GIZI BALITA DI PUSKESMAS MARGOYOSO TAHUN 2020**

Oleh

ATIF SETIAWAN

SKRIPSI

**Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika Dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **IMPLEMENTASI ALGORITMA *K-MEANS* DENGAN MENGGUNAKAN METODE *ELBOW* DAN *SILHOUETTE* PADA KLASTERISASI TINGKAT GIZI BALITA DI PUSKESMAS MARGOYOSO TAHUN 2020**

Nama Mahasiswa : **Atif Setiawan**

Nomor Pokok Mahasiswa : **1917031017**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing



Drs. Eri Setiawan, M. Si.
NIP. 195811011988031002



Drs. Nusyirwan, M. Sc.
NIP. 196610101992031028

2. Ketua Jurusan Matematika



Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua : Drs. Eri Setiawan, M. Si.



Sekretaris : Drs. Nusyirwan, M. Sc.



**Penguji
Bukan Pembimbing : Prof. Dr. Ir. Netti Herawati, M. Sc.**



**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002



Tanggal Lulus Ujian Skripsi : 29 September 2023

PERNYATAAN

Yang bertandatangan di bawah ini :

Nama : **Atif Setiawan**

Nomor Pokok Mahasiswa : **1917031017**

Jurusan : **Matematika**

Judul Skripsi : **IMPLEMENTASI ALGORITMA *K-MEANS*
DENGAN MENGGUNAKAN METODE
ELBOW DAN *SILHOUETTE* PADA
KLASTERISASI TINGKAT GIZI BALITA DI
PUSKESMAS MARGOYOSO TAHUN 2020**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri. Apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 29 September 2023

Atif Setiawan Menyatakan,



Atif Setiawan

NPM. 1917031017

RIWAYAT HIDUP

Penulis bernama lengkap Atif setiawan lahir di Campang, 23 Agustus 2000. Penulis merupakan putra dari bapak Mulyadi dan ibu Aryanti. Penulis merupakan anak pertama dari dua bersaudara.

Penulis menempuh pendidikan sekolah dasar di SD Negeri 2 Simpangkanan tahun 2007 sampai 2013. Pendidikan menengah pertama di SMP Negeri 1 Sumberejo tahun 2013 sampai 2016. Pendidikan menengah atas di SMA Negeri 1 Sumberejo tahun 2016 sampai 2019.

Pada tahun 2019 penulis terdaftar sebagai mahasiswa Program Studi S1 Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) dan mendapatkan bantuan beasiswa BIDIKMISI. Selama menjadi mahasiswa penulis aktif dalam organisasi yaitu sebagai Anggota Bidang Minat Dan Bakat Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) tahun 2020. Selama menjadi mahasiswa penulis juga aktif dalam keanggotaan kepanitiaan Dies Natalis Jurusan Matematika (DINAMIKA) yaitu sebagai anggota divisi perlengkapan DINAMIKA XXI, Koordinator divisi Humas dan Publikasi DINAMIKA XXII, serta *Steering Comitte* (SC) divisi Humas DINAMIKA XXIII.

Pada awal tahun 2022 penulis melakukan Kerja Praktik (KP) di Lembaga Penjaminan Mutu Pendidikan Provinsi Lampung. Pada pertengahan tahun 2022 penulis melakukan Kuliah Kerja Nyata sebagai wujud pengabdian kepada masyarakat yang dilaksanakan penulis di Desa Argomulyo, Kecamatan Sumberejo, Kabupaten Tanggamus, Lampung selama 40 hari.

KATA INSPIRASI

“Orang yang hebat adalah orang yang memiliki kemampuan menyembunyikan kesusahan, sehingga orang lain mengira bahwa kamu selalu senang”

(Imam Syafi'i)

“Angin tidak berhembus untuk menggoyangkan pepohonan, melainkan menguji kekuatan akarnya”

(Ali Bin Abi Thalib)

“Dunia ini cukup untuk menghidupi seluruh manusia, tetapi tak akan cukup untuk satu orang serakah”

(Mahatma Gandhi)

PERSEMBAHAN

Dengan mengucap puji dan syukur kehadirat Allah SWT yang telah memberikan taufik dan hidayah-Nya untuk menyelesaikan skripsi ini, kupersembahkan karya kecil dan sederhana ini kepada :

Bapak dan Mama Tercinta

Yang selalu bekerja dengan keras agar penulis dapat menempuh pendidikan dan mendapat gelar sarjana, yang tidak pernah lelah untuk selalu mendoakan, memberikan dukungan, nasehat dan kasih sayang yang tidak mungkin terbalas oleh apapun.

Adik Tersayang

Yang telah memberikan semangat, motivasi, doa dan dukungan.

Dosen Pembimbing dan Penguji

Yang senantiasa meluangkan waktu untuk mengarahkan dan memotivasi penulis

Sahabat-sahabatku

Yang selalu memberikan doa, dukungan, motivasi, canda dan tawa yang telah menemani penulis dalam setiap langkahnya

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji dan syukur penulis panjatkan atas kehadiran Allah SWT karena berkat taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma K-Means Dengan Menggunakan Metode Elbow Dan Silhouette Pada Klasterisasi Tingkat Gizi Balita Di Puskesmas Margoyoso Tahun 2020”.

Skripsi ini dibuat sebagai salah satu syarat memperoleh gelar Sarjana Matematika di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan bimbingan, dukungan, motivasi, serta saran sehingga skripsi ini dapat terselesaikan. Oleh karena itu penulis ingin mengucapkan terima kasih kepada :

1. Bapak Drs. Eri Setiawan, M.Si., selaku Pembimbing I yang selalu bersedia memberikan waktu, arahan, bimbingan, saran serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Drs. Nusyirwan, M.Sc., selaku Pembimbing II yang telah bersedia memberikan waktu, arahan, bimbingan, dan dukungan kepada penulis.
3. Ibu Prof. Dr. Ir. Netti Herawati, M.Sc., selaku Penguji yang telah bersedia memberikan kritik dan saran serta evaluasi kepada penulis.
4. Ibu Dian Kurniasari S.Si., M.Sc., selaku Pembimbing Akademik yang selalu bersedia memberikan bimbingan, saran serta dukungan kepada penulis pada hal yang berkaitan dengan akademik.

5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Bapak, Mama, dan adik tercinta terima kasih atas cinta, kasih sayang, doa serta dukungan yang tiada hentinya kepada penulis.
8. Teman-teman seperjuangan yaitu Dea, Surya, Irma, Anisa, dan Ardelia yang selalu berbagi suka duka, saling membantu dan saling memotivasi.
9. Teman-teman Jurusan Matematika Angkatan 2019
10. Lanang Galih, Laduna, Sopian, Adrian, dan Fahnan selaku teman-teman yang selalu memberikan support.
11. Fiki Oktavian, Diar Azari selaku sahabat perjuangan di Bandar Lampung.
12. Dian Lestari dan Shafira Nurhaliza selaku teman seperjuangan dari SMA.
13. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Bandar Lampung, 29 September 2023
Penulis

Atif Setiawan

DAFTAR ISI

DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
I. PENDAHULUAN.....	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	5
2.1 Analisis Klaster.....	5
2.2 Ukuran Kemiripan	6
2.3 <i>K-Means</i>	8
2.4 Metode <i>Elbow</i>	11
2.5 <i>Silhouette Coefficient</i>	11
2.6 Simpangan Baku	12
2.7 Multikolinearitas.....	13
2.8 <i>Xie-Beni Index (XBI)</i>	14
2.9 <i>Davies Bouldin Index (DBI)</i>	15
2.10 Gizi	15
III. METODOLOGI PENELITIAN	17
3.1 Waktu dan Tempat Penelitian.....	17
3.2 Data Penelitian.....	17
3.3 Metode Penelitian	17
IV. HASIL DAN PEMBAHASAN	19
4.1 Deskripsi Data	19
4.2 Praproses Data	20
4.2.1 Normalisasi Data	20
4.2.2 Penentuan Jumlah Klaster Optimal dengan Metode <i>Elbow</i>	21
4.2.3 Perhitungan Nilai <i>Silhouette Score</i>	23
4.2.4 Pengujian Performa	25

4.3 Penerapan Algoritma K-Means dengan Menggunakan Metode <i>Elbow</i> dan <i>Silhouette Score</i>	25
4.3.1 Penentuan Titik Pusat Kluster atau <i>Centroid</i>	25
4.3.2 Pengklasteran dengan Jumlah Kluster = 2	26
4.3.3 Menghitung Jarak Masing-Masing Data dengan Titik Pusat	28
4.3.4 Mendeskripsikan Karakteristik Kluster	29
4.3.5 Evaluasi Hasil	29

V. KESIMPULAN	30
----------------------------	-----------

DAFTAR PUSTAKA	31
-----------------------------	-----------

LAMPIRAN

DAFTAR TABEL

Tabel	Halaman
1. Deskripsi Statistik	19
2. Hasil Nilai <i>Elbow</i>	22
3. Nilai <i>Silhouette Score</i> masing-masing klaster.....	23
4. Perbandingan Nilai DBI.....	25
5. Jarak Data dengan Titik Pusat.....	28

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Grafik Persebaran Data	20
Gambar 2. Grafik Hasil Normalisasi.....	21
Gambar 3. Grafik Metode Elbow.....	22
Gambar 4. Cluster dengan Silhoutte Score	24
Gambar 5. Grafik hasil klasterisasi.	26
Gambar 6. Anggota masing-masing klaster.....	27

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis kluster merupakan salah satu teknik pengelompokan data dalam jumlah besar menjadi beberapa kluster atau kelompok yang ditentukan sesuai dengan jarak tiap data. Analisis kluster membagi data ke dalam beberapa himpunan berdasarkan kesamaan datanya. Analisis ini dapat berpotensi mengetahui struktur dari data yang dipakai untuk dilakukan analisis lebih lanjut dalam berbagai aplikasi luas. Tujuan dari analisis kluster adalah untuk mengelompokkan data dengan memaksimalkan kemiripan data dalam suatu kelompok yang dibentuk dan memaksimalkan perbedaan yang terdapat antar kelompok.

Analisis kluster dilakukan karena dengan membagi data atau observasi ke dalam beberapa kelompok yang berbeda memungkinkan untuk menunjukkan pola-pola dan informasi yang dapat dianalisis sehingga mungkin akan berguna ke depannya. Agar lebih memahami data dan observasi yang dilakukan, maka dapat dilakukan proses *clustering* atau klusterisasi berdasarkan variabel yang berbeda sehingga mendapatkan kluster-kluster yang berbeda. Dalam analisis kluster dituntut untuk mampu menginterpretasikan hasil analisis yang telah dilakukan sehingga hasil analisis kluster yang telah dilakukan dapat benar-benar berguna dan bermanfaat.

Lloyd (1957) mengungkapkan bahwa algoritma *K-Means* merupakan salah satu metode untuk melakukan *clustering* atau klusterisasi yang di publikasikan pada tahun 1982. Algoritma *K-Means* ini akan membagi n pengamatan menjadi k

kelompok sedemikian sehingga tiap-tiap kelompok akan masuk ke dalam nilai rata-rata terdekat.

Beberapa penelitian tentang *clustering* dengan menggunakan algoritma *K-Means* telah banyak dilakukan sebelumnya, seperti yang dilakukan Arora (2015). Dalam penelitian ini dilakukan analisis kluster menggunakan algoritma *K-Means* dengan bantuan *software* R-Studio untuk memperoleh kluster dari data dalam jumlah besar. Data yang digunakan pada penelitian ini adalah data penjualan yang berjumlah 10000 data. Bezdán (2021) juga melakukan penelitian analisis kluster dengan menerapkan *K-Means* untuk mengelompokkan dokumen teks. Adapun penelitian lain dilakukan oleh Khairul Islam (2021) yang melakukan penelitian dengan mengaplikasikan *machine learning* dengan metode analisis kluster untuk melakukan analisis penyakit tumor.

Analisis kluster dapat diterapkan dalam beberapa aplikasi luas dalam kehidupan nyata, salah satunya yaitu untuk menentukan status gizi pada balita. Gizi merupakan suatu hasil ekstraksi yang terjadi pada suatu organisme hidup yang diproses kembali untuk menjadi zat-zat yang dibutuhkan oleh tubuh. Gizi dapat berupa zat padat ataupun cair yang dihasilkan dari makanan yang dikonsumsi sehingga diproses kembali untuk memelihara kehidupan, pertumbuhan dan fungsinya organ tubuh suatu organisme sehingga tercipta kehidupan yang layak.

Pemenuhan gizi yang baik sangat bergantung kepada kesadaran orang tua akan pentingnya gizi yang cukup bagi balita dan faktor penunjang lainnya seperti misalnya faktor ekonomi, kesediaan bahan pangan, dan faktor-faktor lainnya. Dalam pertumbuhan balita, gizi merupakan salah satu hal terpenting yang harus dipenuhi sehingga nantinya balita dapat tumbuh dan berkembang dengan baik dan menjadi anak yang sehat. Karena dengan gizi yang cukup dan pertumbuhan yang baik, maka sistem imunitas tubuh dan sistem kerja otak juga akan terbentuk

dengan baik yang akan menjadikan anak menjadi yang tumbuh dengan normal dan maksimal.

Dalam jurnal yang ditulis (Yunita & Widodo, 2015) membandingkan algoritma *K-Means* dan *K-Medoids* dengan menggunakan data sekunder puskesmas yang berjumlah 121 di DIY pada tahun 2015. Dalam penelitian ini diperoleh metode terbaik adalah metode *K-Means* dikarenakan memiliki nilai rasio simpangan baku terkecil untuk pengklasteran.

Beberapa contoh penggunaan analisis kluster beberapa waktu terakhir ini adalah untuk melakukan analisis tingkat persebaran kasus COVID-19 pada provinsi di Indonesia, analisis segmentasi pasar, dan analisis brain tumor, serta indikator tingkat kemiskinan masyarakat. Metode ini digunakan karena cukup efektif dalam menyelesaikan beberapa kasus diatas sehingga nantinya terdapat gambaran kluster yang dapat memberikan gambaran informasi bermanfaat.

Di era perkembangan teknologi yang semakin maju dan canggih ini, dapat diperoleh berbagai macam cara dan informasi untuk melakukan analisis yang dapat mempermudah untuk mendapatkan gambaran informasi tentang status gizi balita. Salah satu yang dapat dilakukan yaitu dengan melakukan analisis kluster mengenai status gizi balita. Dalam hal ini nantinya akan dikumpulkan data tinggi dan berat badan balita lalu akan diambil data balita dengan kelahiran yang memenuhi standar normal sehingga nantinya data ini dapat dikelompokkan menjadi beberapa kluster yang optimal.

Dari pemaparan diatas, penulis tertarik melakukan analisis kluster menggunakan algoritma *K-Means* dengan menggunakan metode *Elbow* dan *Silhouette* pada status gizi balita yang akan dilakukan dengan bantuan bahasa pemrograman *Python*. Dalam penelitian ini nantinya akan digunakan metode *Elbow* dan *Silhouette* untuk menentukan jumlah kluster optimal sehingga diperoleh hasil pengklasteran dan jumlah kluster terbaik menggunakan algoritma *K-Means*.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menentukan jumlah kluster optimal dengan menggunakan metode *Elbow* dan *Silhouette* pada *K-Means* untuk data tinggi dan berat badan balita di Puskesmas Margoyoso tahun 2020.
2. Menentukan hasil pengklasteran data tinggi dan berat badan balita di Puskesmas Margoyoso tahun 2020.
3. Mengevaluasi hasil pengklasteran.
4. Memperoleh tingkat status gizi balita di Puskesmas Margoyoso tahun 2020.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memperoleh jumlah kluster optimal dengan menggunakan metode *Elbow* dan *Silhouette* pada *K-Means* untuk data tinggi dan berat badan balita di Puskesmas Margoyoso tahun 2020.
2. Mendapatkan hasil kluster optimal data tinggi dan berat badan balita di Puskesmas Margoyoso tahun 2020.
3. Menambah wawasan dan informasi untuk penerapan ilmu statistika dalam kehidupan nyata khususnya analisis kluster untuk tingkat status gizi balita.
4. Mendapatkan informasi tentang gizi balita yang ada di Puskesmas Margoyoso tahun 2020.

II. TINJAUAN PUSTAKA

2.1 Analisis Klaster

Hair, *et al* (2014) menyatakan analisis klaster adalah salah satu dari beberapa kumpulan teknik-teknik pengolahan data yang bertujuan untuk mengelompokkan beberapa objek menjadi kumpulan objek yang berdekatan berdasarkan dari karakteristik apa yang dimiliki.

Tujuan dari analisis klaster yaitu membuat kelompok (klaster) dengan memaksimalkan kemiripan dari kelompok dan memaksimalkan perbedaan antar kelompok. Kriteria klaster yang baik yaitu:

- Homogenitas (persamaan) dalam klaster yang tinggi atau besar.
- Heterogenitas (perbedaan) antar klaster yang tinggi atau besar.

Metode analisis klaster dibagi menjadi dua, yaitu analisis klaster *non- hirarki* dan analisis klaster *hirarki*.

1. Analisis klaster *non-hirarki*

Merupakan analisis klaster yang melakukan proses pengklasteran secara langsung pada semua observasi yang terdapat pada *dataset*, sehingga pengklasteran hanya terjadi dalam satu tingkatan atau level. Beberapa algoritma yang termasuk ke dalam analisis klaster non-hirarki adalah *K-Means Clustering*, *K-Medoids*, dan DBSCAN.

2. Analisis kluster hirarki

Merupakan analisis kluster dengan menggunakan proses klusterisasi bertingkat.

2.2 Ukuran Kemiripan

Ukuran kemiripan merupakan tingkat kemiripan yang dimiliki suatu data. Ukuran kemiripan yang biasa digunakan dalam analisis kluster yaitu jarak *euclidean* dan jarak mahalanobis. Jarak *euclidean* digunakan jika variabel saling bebas atau tidak berkorelasi satu sama lain (tidak terjadi multikolinearitas). Namun jika terjadi multikolinearitas, maka dapat digunakan jarak Mahalanobis.

1. Jarak *Euclidean*

Jarak *euclidean* atau metrik *euclidean* adalah perhitungan jarak antara dua titik pada ruang *Euclidean* (Derisma, Firdaus, & Yusya, 2016). Dengan jarak ini, jarak *euclidean* menjadi sebuah jarak metrik yang diasosiasikan dengan istilah norma *euclidean*. Persamaan jarak *euclidean* digunakan untuk mengukur dua titik dalam satu dimensi seperti halnya perhitungan *pythagoras* (Mustofa & Suasana, 2018).

Berikut adalah persamaan *euclidean distance*:

$$d_x = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

untuk mengkalkulasi jarak x dan y didefinisikan dalam rumus :

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

jika semakin banyak nilai x dan y , dengan asumsi $\mathbf{x} = (x_1, x_2, \dots, x_n)$ dan $\mathbf{y} = (y_1, y_2, \dots, y_n)$, sehingga dapat dirumuskan bahwa:

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_n - y_n)^2}$$

sehingga diperoleh persamaan rumus *Euclidean* adalah:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

dengan:

$d(x, y)$	= jarak data x ke pusat kluster y
x_i	= data x pada observasi ke- i
y_i	= titik pusat ke- y pada observasi ke- i
n	= banyaknya observasi

2. Jarak Mahalanobis

Jarak mahalanobis merupakan perhitungan jarak dengan cara mengukur setiap objek dengan ukuran pusat dari himpunan objek tersebut (Kantacarini, Herwindiaty, & Hendryli, 2021). Ukuran pusat yang digunakan adalah nilai pusat mean dan median. Dan nilai median dinyatakan dengan persamaan sebagai berikut:

Jika N ganjil maka, $\text{Med} = x_{(\frac{n}{2}+1)}$

Jika N genap maka, $\text{Med} = \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}$

dengan:

x = data ke- I pada himpunan data

n = banyaknya sampel dalam suatu himpunan data

sehingga perhitungan jarak Mahalanobis dalam persamaan umum adalah sebagai berikut:

$$d^2(x, y) = \sqrt{(x_i - y_i)^T S^{-1} (x - y)} \quad (2.2)$$

dimana x_i dan y_i sebagai vektor dari objek data dan S^{-1} merupakan *invers* matriks kovarians.

2.3 *K-Means*

Arora, *et al* (2015) menyatakan bahwa algoritma *K-Means* merupakan salah satu metode partisi *non-hierarki* yang terkenal untuk pengelompokan atau klaster suatu data. Metode ini mengelompokkan atau mengklaster data berdasarkan jarak kedekatannya satu sama lain menurut jarak *euclidean*. Data yang sama akan dikelompokkan ke dalam satu klaster dan data yang berbeda akan dikelompokkan ke dalam klaster yang lain.

Pada *K-Means*, misal didefinisikan $X = \{x_1, x_2, \dots, x_n\}$ adalah sebuah himpunan data dalam ruang berdimensi D , yang dinotasikan R^D sedangkan k adalah sebuah bilangan integer positif lebih dari satu. $X_n \in R^D$, maka algoritma *K-Means clustering* akan membagi (mempartisi) X ke dalam k kelompok, dapat dinyatakan dengan himpunan X_1, X_2, \dots, X_k yang saling lepas, sehingga $X_1 \cup X_2 \cup \dots \cup X_k = X$, dimana setiap kelompok memiliki nilai tengah (*centroid*) dari data-data dalam kelompok tersebut. Algoritma *K-Means* secara acak menentukan k buah data sebagai titik tengah (*centroid*), kemudian dengan pendekatan jarak *euclidean* dihitung jarak antara data dengan *centroid*, untuk selanjutnya data akan ditempatkan ke dalam kelompok yang terdekat dihitung dari titik tengah kelompok. Proses penentuan *centroid* dan penempatan data kelompok dilakukan sampai nilai *centroid* konvergen (*centroid* dari semua kelompok tidak berubah lagi).

Dalam bentuk yang paling sederhana langkah-langkah algoritma *K-Means* terdiri dari tiga tahap :

1. Bagi objek-objek ke dalam k kelompok dan tentukan pula *centroid* pada tiap kelompok.
2. Masukkan objek ke suatu kelompok berdasarkan nilai rata-rata terdekat. Jarak yang digunakan biasanya menggunakan jarak *euclidean*, dan hitung kembali rata-rata untuk kelompok baru yang terbentuk.
3. Ulangi langkah 2 sampai tidak ada lagi pemindahan objek antar kelompok.

Cara yang sederhana untuk menentukan kualitas kelompok adalah dengan mengukur seberapa dekat masing-masing titik-titik kelompok dengan *centroid*, dengan demikian ukuran kepadatan kelompok dinyatakan sebagai total jarak dari rata-rata kelompok (Indraputra & Fitriana, 2020).

$$\sum_{x_n \in C_k} \|x_n - m_k\|^2 = \sum_{n=1}^N z_{kn} \|x_n - m_k\|^2 \quad (2.3)$$

dimana pusat kluster didefinisikan sebagai :

$$m_k = \frac{1}{N} \sum_{x_n \in C_k} x_n \quad (2.4)$$

dengan z_{kn} menunjukkan keanggotaan data X_n terhadap kelompok k , dimana $z_{kn} = 1$ jika $X_n \in C_k$, dan 0 untuk yang lainnya dan $N_k = \sum_{n=1}^N z_{kn}$ adalah total banyaknya titik yang dialokasikan ke kelompok k . Maka kebaikan kelompok berdasarkan jumlah ukuran kepadatan kelompok untuk masing-masing kelompok k . dengan menggunakan indikator variabel bisa ditentukan ukuran kebaikan kelompok secara keseluruhan sebagai berikut :

$$\varepsilon_k = \sum_{n=1}^N \sum_{k=1}^k z_{kn} \|x_n - m_k\|^2 \quad (2.5)$$

Setelah memperoleh ukuran secara keseluruhan dari kualitas kelompok, langkah selanjutnya adalah menyusun algoritma yang akan memungkinkan untuk mengoptimalkan ukuran ini. Kriteria yang harus dioptimalkan adalah $\varepsilon_k = \sum_{n=1}^N \sum_{k=1}^k z_{kn} \|x_n - m_k\|^2$, dimana terdapat dua set parameter, yaitu rata-rata kelompok m_k dan variable indikator alokasi kluster z_{kn} . Optimalisasi pada setiap set variabel dengan memegang set yang bersifat tetap (*fixed*).

Kriteria pengklusteran ketika menggunakan algoritma *K-Means* adalah meminimalkan nilai *clustering error* (ε_k), atau secara sederhana sebagai berikut:

$$\varepsilon_k = \sum_{n=1}^N \sum_{k=1}^k z_{kn} \|X_n - m_k\|^2$$

dapat dijabarkan sebagai berikut

$$\begin{aligned}
\varepsilon_k &= \sum_{n=1}^N \sum_{k=1}^K z_{kn} \|X_n - m_k\|^2 \\
&= \sum_{n=1}^N \sum_{k=1}^K z_{kn} (x_n - m_k)^T (X_n - m_k) \\
&= \sum_{n=1}^N \sum_{k=1}^K z_{kn} (X_n^T X_n - 2m_k^T X_n + m_k^T m_k)
\end{aligned} \tag{2.6}$$

dengan catatan :

$$m_x^T X_n = \frac{1}{N} \sum_{m=1}^N z_{km} X_m^T X_n \tag{2.7}$$

dan

$$m_x^T X_n = \left(\frac{1}{N} \sum_{p=1}^N z_{kp} X_p \right)^2 \tag{2.8}$$

$$= \frac{1}{N_k^2} \sum_{p=1}^N \sum_{l=1}^N z_{kp} z_{kl} X_p^T X_l \tag{2.9}$$

Fitriyiah (2017) menjelaskan algoritma yang lebih sistematis untuk melakukan analisis kluster menggunakan metode *K-Means* adalah sebagai berikut:

1. Menentukan banyaknya k , dimana k merupakan jumlah kluster yang akan dibentuk.
2. Menentukan *centroid* atau titik pusat kluster secara acak dari setiap kluster yang dibentuk.
3. Menghitung jarak setiap objek atau data ke *centroid* atau pusat kluster dengan menggunakan *euclidean* pada persamaan (2.1). Lalu alokasikan semua data ke *centroid* terdekat.
4. Mengelompokkan setiap data yang berdasarkan pada jarak kedekatannya (jarak terkecil) dengan *centroid* dengan persamaan (2.10) sebagai berikut.

$$\text{Min} \sum_{i=1}^k d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2.10}$$

5. Menghitung nilai *centroid* baru. Nilai ini diperoleh dari rata-rata setiap kluster dengan persamaan (2.11) berikut.

$$C_{ij} = \frac{\sum_{i=1}^p X_{ij}}{p} \quad (2.11)$$

dengan:

C_{ij} = *centroid* terbaru pada iterasi ke- k

X_{ij} = anggota kluster ke- k

p = banyaknya anggota kluster ke- k

6. Ulangi langkah 2-5 hingga mendapatkan kondisi konvergen, dimana anggota setiap kluster tidak mengalami perubahan tata letak kluster.

2.4 Metode *Elbow*

Merliana, Ernawati & Santoso (2012) menyebutkan bahwa metode *elbow* merupakan salah satu metode yang digunakan untuk mencari jumlah kluster optimal dengan melihat nilai persentase perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik dan pada titik inilah merupakan jumlah kluster yang optimal untuk digunakan. Selain dari nilai persentasenya, untuk menentukan jumlah kluster yang optimal dengan metode ini juga dapat dilihat dari hasil grafik yang dihasilkan dimana grafik akan membentuk suatu sudut siku ataupun mendekati sudut siku. Dari grafik dapat dilihat penurunan nilai *Sum Square Error* (SSE) yang paling signifikan yang akan membentuk lekukan sudut mendekati sudut siku yang akan menjadi tolak ukur dalam menentukan jumlah kluster optimal.

2.5 *Silhouette Coefficient*

Merupakan metode yang digunakan untuk evaluasi kluster yang menggabungkan metode *cohesion* dan *separation*. *Cohesion* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah kluster dan *separation* diukur dengan

menghitung jarak rata-rata setiap objek dalam sebuah kluster dengan kluster terdekatnya (Pramessti, dkk 2017). Nilai *silhouette* untuk keseluruhan data dengan jumlah kluster k , dapat didefinisikan sebagai $sil(k)$ yang dihitung dengan persamaan rata-rata *silhouette value* untuk semua kluster.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \quad (2.12)$$

dengan:

$sil(k)$ = nilai *silhouette* semua kluster

$|k|$ = banyaknya kluster k

$sil(c_i)$ = rata-rata nilai *silhouette*

2.6 Simpangan Baku

Fithriyyah (2017) dalam jurnalnya menuliskan bahwa simpangan baku merupakan salah satu cara untuk mengetahui metode pengklusteran yang mempunyai kinerja terbaik. Simpangan baku ini dapat digunakan untuk menghitung rata-rata simpangan dalam kluster dan antar kluster. Adapun persamaan simpangan baku dalam kluster adalah:

$$\sigma_w = K^{-1} \sum_{k=1}^K \sigma_k \quad (2.13)$$

dengan σ_w simpangan baku dalam kluster, K adalah banyaknya kluster yang terbentuk, dan σ_k simpangan baku kluster ke k . Selanjutnya persamaan untuk simpangan baku kluster ke- k adalah:

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{X}_i - \mu_k)^2} \quad (2.14)$$

dengan N adalah jumlah anggota kluster ke- k , \bar{X}_i adalah rata-rata variabel data ke I , dan μ_k adalah rata-rata kluster ke- k . Sedangkan persamaan untuk simpangan baku antar kluster adalah

$$\sigma_b = \left[(K)^{-1} \sum_{k=1}^N (\mu_k - \mu)^2 \right]^{\frac{1}{2}} \quad (2.15)$$

dengan σ_b adalah simpangan baku antar klaster dan μ adalah rata-rata keseluruhan klaster. Sehingga nilai rasio simpangan baku dapat dihitung dengan persamaan sebagai berikut:

$$\sigma = \frac{\sigma_w}{\sigma_b} \times 100\% \quad (2.16)$$

metode dengan nilai rasio simpangan baku terkecil merupakan metode pengklasteran terbaik dimana klaster yang baik adalah klaster yang mempunyai homogenitas yang tinggi antar anggota dalam satu klaster dan heterogenitas yang tinggi antar klaster.

2.7 Multikolinearitas

Sriningsih, Hatidja & Prang (2018) menjelaskan bahwa multikolinearitas merupakan kondisi dimana adanya hubungan antara variabel bebas dan tak bebas dalam variabel penelitian. Uji multikolinearitas merupakan salah satu uji asumsi klasik. Uji ini merupakan salah satu asumsi yang harus dilakukan dalam pengujian analisis klaster menggunakan metode *K-means*. Tujuan dari uji multikolinearitas adalah untuk mengetahui apakah terdapat korelasi atau hubungan antar variabel bebas dalam penelitian. Persamaan uji multikolinearitas adalah sebagai berikut :

$$VIF_i = \frac{i}{1 - R_i^2} \quad (2.17)$$

dengan:

VIF = *Variance Inflation Factor* (rasio varians dalam memperkirakan beberapa variabel)

R_i^2 = koefisien determinasi antara x_i dengan variabel bebas lainnya pada persamaan atau model

$i = 1, 2, \dots, p$

jika nilai VIF < 10 atau *tolerance value* < 0,1 maka terjadi multikolinearitas, namun jika nilai VIF > 10 atau *tolerance value* > 0,1 maka tidak terjadi multikolinearitas.

2.8 Xie-Beni Index (XBI)

Xie-Beni Index (XBI) merupakan langkah yang digunakan untuk menentukan jumlah kluster optimum (Purnamasari, 2014). Indeks ini bertujuan mengukur rasio total variasi dalam kluster (S) dan pemisah antar kluster (d_{min}). Banyaknya kluster optimum ditentukan dengan meminimalkan nilai indeks. *Xie-Beni Index* dinyatakan sebagai berikut:

$$XB(c) = \frac{S}{d_{min}} \quad (2.18)$$

dimana:

$$S = \frac{\sum_{i=1}^n \sum_{k=1}^n (u_{ik})^m d_{ik}^2(x_i, v_k)}{n} \quad (2.19)$$

$$d_{min} = \min_{ik} d_{ik}^2(v_i, v_k) \quad (2.20)$$

dengan:

$d_{ik}^2(v_i, v_k)$ = jarak data terhadap pusat kluster

u_{ik} = nilai kenaggotakan objek ke- k dengan pusat kelompok

ke- i

n = banyaknya kluster

m = bobot pangkat

2.9 *Davies Bouldin Index (DBI)*

DBI merupakan metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin yang digunakan untuk mengevaluasi kluster. Evaluasi menggunakan DBI memiliki skema evaluasi internal kluster, dimana baik atau tidaknya hasil kluster dilihat dari kuantitas dan kedekatan antar data hasil kluster. Semakin kecil nilai DBI diperoleh (*non-negative* ≥ 0), maka akan semakin baik kluster yang diperoleh dari pengelompokkan menggunakan algoritma *clustering* (Harmain, dkk. 2021).

2.10 **Gizi**

WHO (*World Health Organization*), dalam Soekirman, (2004) menjelaskan bahwa gizi merupakan suatu hasil ekstraksi yang terjadi pada suatu organisme hidup yang diproses kembali untuk menjadi zat-zat yang dibutuhkan oleh tubuh. Gizi dapat berupa zat padat ataupun cair yang dihasilkan dari makanan yang dikonsumsi sehingga diproses kembali untuk memelihara kehidupan, pertumbuhan dan berfungsinya organ tubuh suatu organisme sehingga tercipta kehidupan yang layak. Beberapa gizi yang diperlukan oleh tubuh diantaranya yaitu:

- a. Karbohidrat
- b. Protein
- c. Vitamin dan mineral
- d. Kalsium
- e. Zat besi

Semua kebutuhan gizi tersebut dapat dipenuhi melalui asupan makanan yang diberikan sehari-hari. Asupan makanan yang baik untuk menjaga pertumbuhan dan memperoleh gizi yang baik di antaranya yaitu :

- 1) Karbohidrat

Dengan takaran saji 28 gram setara dengan satu lembar roti, segelas sereal atau setengah gelas beras.

2) Sayuran

Disajikan dengan ukuran terkecil namun masih dapat dimakan dengan takaran satu gelas.

3) Buah

Dapat dihaluskan dengan takaran 480 ml.

4) Susu

Satu gelas setara dengan yogurt dan keju.

5) Daging

Dengan takaran berat 28 gram setara dengan satu butir telur.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung tahun 2023.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yaitu data tinggi dan berat badan 288 balita yang diperoleh dari PUSKESMAS Margoyoso tahun 2020 untuk dilakukan analisis kluster dengan algoritma *K-Means* menggunakan metode *elbow* dan *silhouette*.

3.3 Metode Penelitian

Metode yang digunakan pada penelitian ini yaitu analisis kluster menggunakan algoritma *K-Means* dengan metode *elbow* dan *silhouette* dengan bantuan bahasa pemrograman *python*. Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Menentukan banyaknya kluster yang akan dibentuk dengan menggunakan titik siku pada grafik metode *elbow* dan *silhouette score*.
2. Melakukan analisis kluster dengan menggunakan metode *K-means*.

Tahapan analisis kluster menggunakan metode *K-Means* adalah sebagai berikut:

- a. Menentukan titik pusat (*centroid*) awal
 - b. Menghitung jarak setiap data ke titik pusat (*centroid*)
 - c. Mengalokasikan semua data ke titik pusat terdekat
 - d. Menghitung nilai titik pusat (*centroid*) terbaru
 - e. Memeriksa kriteria pemberhentian iterasi
3. Mendeskripsikan karakteristik kluster

Dalam mendeskripsikan karakteristik kluster digunakan persamaan sebagai berikut.

$$X = \frac{\sum x_{ij}}{n_j} \quad i = 1,2,3, \dots, n \quad j = 1,2,3, \dots, q \quad (3.1)$$

Dengan:

X = rata-rata sampel (rata-rata pada kluster tertentu)

n_j = banyak anggota pada kluster ke- j

x_{ij} = nilai data ke- i pada variabel ke- j

4. Evaluasi hasil

V. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa:

1. Pada klasterisasi tingkat status gizi balita dengan menggunakan data berat dan tinggi badan balita di Puskesmas Margoyoso Kabupaten Tanggamus diperoleh jumlah klaster optimal sebanyak 2 yang ditentukan dengan menggunakan metode *elbow* dan *silhouette score* dengan pusat klaster pertama [10.96361702 83.19202128] ditandai dengan angka 0 dan klaster kedua dengan pusat [12.3721 89.543] ditandai dengan angka 1. Pada pengklasteran ini diperoleh nilai *silhouette score* sebesar 0.5 yang berarti pengelompokan data sudah menghasilkan struktur yang baik.
2. Klaster pertama berisikan balita dengan berat dan tinggi badan cukup baik sebanyak 188 balita dan klaster kedua berisikan balita dengan berat badan dan tinggi badan baik sebanyak 100 balita.
3. Tingkat status gizi pada Puskesmas Margoyoso rata-rata cukup baik, salah satu faktor yang mempengaruhi pengklasteran ini adalah usia saat ukur dan jenis kelamin dari balita itu sendiri.

DAFTAR PUSTAKA

- Arora, P., Deepali, Varshney. 2016. Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia Computer Science*. **7**: 507-5012.
- Badruttaman, A., Sudarno, maruddani, D. A. 2020. Penerapan Analisis Kluster K-Modes Dengan Validasi Davies Bouldin Dalam Menentukan Karakteristik kanal Youtube di Indonesia. *Jurnal Gaussian*. **9**(3): 263-272.
- Bezdan, T., *et all*. 2021. Hybrid Friut-Fly Optimization Algorithm with K-Means for Text Document Klastering. *Jurnal Mathematics*. **9**(16): 1-19.
- Derisma, Firdaus, & Yusya, R. P. 2016. Perancangan Ikat Pinggang Elektronik Untuk Tunanetra Menggunakan Mikrokontroller Dan Global Positioning System (Gps) Pada Smartphoe Android. *Jurnal Teknik elektro ITP*. **5**(2): 130-136.
- Fithriyyah. 2017. *Analisis Kluster Spasial Tingkat Kerawanan Demam Berdarah (DBD) di Provinsi Daerah Istimewa Yogyakarta*. Skripsi. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Gustientiedina, Adiya, M. H., Desnelita, Y. 2019. Penerapan Algoritma K-Means Untuk Klastering Data Obat-obatan di Pekanbaru. *Jurnal Nasional Teknologi dan Sistem Informasi*. **5**(1): 17-24.
- Hair, J. F., Black, W.C., Babin, B. J. & Anderson, R.E. 2014. *Multivariate Data Analysis*. 7th Edition. Pearson Education, Upper Saddle River.
- Han, J., Kamber. 2006. *Data Mining: Concept and Techniques*. Waltham Morgan Kaufmann Publisher.

- Harmain, A., Paiman., Kurniawan, H., Kusri., & Maulina, D. 2021. Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA*. **2**(2): 83-89.
- Indraputra, R. A. & Fitriana, R. 2020. K-Means Clustering Data Covid-19. *Jurnal Teknik Industri*. **10**(3): 275-282.
- Islam, M. K., Ali, M. S., Miah, M. S., Rahman, M. M., Alam, M. S., & Hossain, M. A. 2021. Brain Tumor Detection in MR Image Using Superpixels, Pricipal Component Analysis and Template based K-Means Klastering Algorithm. *Elsevier*. **5**: 1-8.
- Matdoan, M. Y., Matdoan, U. A., & Far-far, M. S. 2022. Algoritma K-Means Untuk Klasifikasi provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting. *PANRITA Journal of Science, Technology , and Arts*. **1**(2): 41-46.
- Miftahuddin, Y., Umaroh, S., & Karim, F. R. 2020. Perbandingan Metode Perhitungan jarak Euclidean, Haversine, dan Manhattan Dalam Penentuan Posisi Karyawan. *Jurnal Tekno Intensif*. **14**(2): 69-77.
- Murpartiwi, S. I., Indrawan, I. G. A., & Aranta, A. 2021. Analisis Pemilihan Kluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan Teknologi dan Kejuruan*. **18**(2): 152-163.
- Mustofa, Z., & Suasana, I. S. 2018. Algoritma Clustering K-Medoids pada E-Government Bidang Information And Communication Technology Dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi dan Komunikasi*. **9**(1): 1-10.
- Purnamasari, S.B. 2014. Pemilihan Kluster Optimum Pada Fuzzy C-Means. *Jurnal Gaussian*. **3**(3): 491-498.
- Sinaga, P. & Yang, M. S. 2020. Unsupervised K-Means Klastering Algorithm. *Digital Object Identifier*. **8**: 80716-80727.
- Soekirman. 2004. *Ilmu Gizi dan Aplikasinya*. Jakarta : Ditjen DIKTI, Depdiknas.

- Soemartini, Supartini, E. 2017. Analisis K-Means Klaster Untuk Pengelompokkan Kabupaten / Kota di Jawa Barat Berdasarkan Indikator Masyarakat. *Prosiding*, 144-154.
- Sriningsih, M., Hatidja, D., Prang, J. D. 2018. Penanganan Multikolineritas Dengan Menggunakan Analisis regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulawesi. *Jurnal Ilmiah Sains*. **18**(1): 19-20.
- Talakua, M. W., Leleury, Z.A., Talluta, A. W. 2017. Anaisis Klaster dengan Menggunakan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Jurnal Ilmu Matematika dan Terapan*. **11**(2): 119-128.
- Turban, dkk. 2005. *Decision Support Systems and Intelligent Systems* 7th Edition Jilid 1. Yogyakarta: Andi.
- Yunita, Widodo, E. 2015. Perbandingan K-Means dan K-Medoids Klastering Terhadap Kelayakan Puskesmas di DIY Tahun 2015. *Jurnal Statistika FMIPA Universitas Islam Indonesia*. **1**(1): 116-122.