



Implementasi Algoritma *K-Means* Clustering Status Gizi Balita

Nurul Rizki Octaviani¹, Rini Mayasari², Susilawati³

Universitas Singaperbangsa Karawang

Received: 16 Juli 2022

Revised: 18 Juli 2022

Accepted: 22 Juli 2022

Abstract :

Nutritional problems are a common problem in Indonesia, especially the problem of malnutrition. One of the factors that affect the nutritional status of toddlers is the lack of parental knowledge, because just looking at the physical development of toddlers is certainly not enough to know the category of nutritional status of toddlers. This study uses data mining techniques by applying the K-Means method, the grouping process using the K-Means algorithm is expected to make it easier for the community, especially mothers, to know the nutritional status of their children, we can know that the relevant K-Means Algorithm is used for the process of grouping toddler nutritional data. The output of this study is to classify the nutritional status of toddlers based on weight and height. The validation test in this study used the Davies Bouldin Index, according to the results of testing and data validation, the final result menu Submit the DBI Kmeans result 0.522483983 (Non-Negative). DBI is declared to be essentially optimal if the final value obtained is as low as possible (non-negative ≥ 0) in order to measure whether the final cluster obtained in the calculation is good or not. This value describes the high level of similarity/similarity of membership in a cluster that has a large degree of similarity/similarity and the distance between one cluster and other clusters is also large. The nutritional clustering process for toddlers is grouped according to their nutritional status

Keywords: Data Mining, Clustering, Toddler Nutrition

(*) Corresponding Author: nurul.rizki18034@student.unsika.ac.id, rini.mayasari@staff.unsika.ac.id, susilawati.sobur@staff.unsika.ac.id

How to Cite: Octaviani, N., Mayasari, R., & Susilawati, S. (2022). Implementasi Algoritma K-Means Clustering Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 370-381. <https://doi.org/10.5281/zenodo.6962588>.

PENDAHULUAN

Menurut laporan dari Global Nutrition Report (2014), Indonesia masuk kedalam 17 negara yang memiliki 3 permasalahan gizi sekaligus, yaitu stunting (pendek), wasting (kurus), dan juga overweight (obesitas). Berdasarkan data riset kesehatan dasar (Riskesdas, 2013) disebutkan bahwa jumlah balita yang mengalami gangguan obesitas menurut BB/TB pada anak usia 0-59 bulan sebesar 11,8% sedangkan berdasarkan data survey pemantauan status gizi (PSG, 2015) disebutkan bahwa jumlah balita yang mengalami gangguan obesitas menurut BB/TB usia 0-59 bulan sebesar 5,3%. Obesitas dapat menjadi awal dari berbagai penyakit. Maka dari itu pemberian gizi yang berlebihan bagi balitapun tidak baik, penerapan gizi seimbanglah yang seharusnya diberikan dan dapat dikontrol oleh para orang tua (Kementerian Kesehatan Republik Indonesia, 2018)

Status gizi pada balita dapat diketahui berdasarkan indeks Berat Badan menurut umur (BB/U), Tinggi Badan menurut umur (TB/U), dan Berat Badan menurut Tinggi Badan (BB/TB) dengan memasukan data nilai dari berat badan, tinggi badan serta jenis kelamin balita. Salah satu cara untuk mengukur tubuh



manusia dalam menentukan status gizi anak yaitu antropometri. Pengukuran antropometri secara umum berdasarkan Berat Badan menurut umur (BB/U) dan Tinggi Badan menurut umur (TB/U). penggunaan antropometri (pengukuran tubuh) yang sering digunakan untuk mengukur dari status gizi balita, hal ini disebabkan sederhana dan aman pada tahap pengukuran serta tidak membutuhkan tenaga ahli. Hasil dari antropometri berupa data yang akurat serta mampu mendeteksi riwayat gizi di masa lampau. Model pengujian dilakukan menggunakan clustering Data mining dengan metode K-Means berdasarkan berat badan dan tinggi badan (Sari et al., 2017).



Gambar 1.1 Prevalensi kekurangan gizi pada balita
(sumber : BPS, Maret 2020)

Gizi balita di Indonesia dari tahun ke tahun mengalami penurunan seperti pada gambar 1.1 terdapat data Prevalensi kekurangan gizi pada anak balita di Indonesia yang diposting oleh BPS yaitu pada bulan maret 2020, yang memrepresentasi status gizi balita Indonesia. Data tersebut diperoleh merupakan tahun 2010, 2013, 2017, 2018. Anak di usia dibawah lima tahun (Balita) rentan terhadap masalah kesehatan dan gizi. Masalah gizi pada anak balita sangat berbeda sifatnya dengan orang dewasa karena masalah gizi pada anak balita tidak mudah dikenali oleh masyarakat, bahkan keluarga. Akibatnya, bila suatu desa terdapat sejumlah anak yang menderita masalah gizi, mereka tidak segera mendapat perhatian karena anak-anak tersebut kadang tidak tampak sakit.

Untuk itu penulis tertarik untuk mengadakan penelitian mengenai pengelompokan data balita, karena pengelompokan item data ke dalam sejumlah kecil grup sehingga masing-masing grup mempunyai persamaan yang esensial. Pengelompokan data mining di bagi menjadi enam bagian antara lain *description*, *estimation*, *prediction*, *klasification*, *clustering*, dan *association* yang diambil didalam jurnal ilmiah yang dibuat Lutfi.S dan Kusri, Larose. K-Means adalah metode pemecahan pada data mining yang sanggup dipakai buat melakukan pengelompokan/clustering suatu data. Metode K-Means merupakan metode yg termasuk pada prosedur pemecahan clustering berbasis jeda yang membagi-bagi data kedalam sejumlah cluster & prosedur pemecahan tersebut hanya dapat optimal jika atribut tersebut berbentuk angka/numerik

METODOLOGI PENELITIAN

Metodologi penelitian yang dipakai pada penelitian ini menggunakan metodologi Knowledge Discovery in Database (KDD), Adapun tahapan pada KDD sebagai berikut :

- Seleksi Data
Pemilihan Data berdasarkan data yang ingin digunakan pada penelitian.
- Pembersihan Data

Data yang hilang/data yang ganda akan dibuang karena keberadaannya dapat mengurangi akurasi dari hasil penelitian

- **Transformasi Data**

Melakukan perubahan jenis data agar dimodelkan sesuai dengan kebutuhan algoritma yang dipakai pada penelitian

- **Data Mining**

Proses analisis data dalam jumlah besar dengan algoritma K-means untuk menentukan hasil akurat yang bisa di manfaatkan

- **Pengujian Hasil untuk Knowledge**

Melakukan pengujian data, hasil perhitungan algoritma untuk mengetahui tingkat keakurasian hasil penelitian

HASIL DAN PEMBAHASAN

Hasil penelitian data mining yang telah dilakukan adalah bagaimana menerapkan algoritma *K-means* dalam melakukan clustering status gizi balita di posyandu perum pemda teluk jambe timur karwang sehingga hasil clustering dapat digunakan sebagai sumber informasi pihak posyandu dalam melakukan clustering status gizi balita. Penelitian ini menggunakan metode KDD (*Knowledge Discovery and Data Mining*) sebagai metode standart dalam data mining untuk mengolah data mentah yang didapatkan dari hasil observasi, dengan tahapan-tahapan yakni: Seleksi Data, *Cleaning Data*, *Transformation Data* / Permodelan Data, Proses Data Mining (*Clustering*), & Pengujian Data

Pengumpulan Data

Pengumpulan data perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas

Tabel 1 pengumpulan data

| Balita | Berat Badan | Tinggi Badan | Lingkar Kepala |
|----------------|-------------|--------------|----------------|
| Balita1 | 12 | 92 | 50 |
| Balita2 | 6 | 63 | 48 |
| Balita3 | 12 | 79 | 50 |
| Balita4 | 9 | 72 | 51 |
| Balita5 | 11 | 75 | 49 |
| | | | |

Seleksi Data

Proses Seleksi Data ini dilakukan, dengan cara membuang atribut yang tidak diperlukan dalam penelitian dan mengambil data nilai yang menunjang kebutuhan pada penelitian ini

Tabel 2 Data hasil seleksi

| Balita | Berat Badan | Tinggi Badan |
|-----------------|-------------|--------------|
| Balita 1 | 12 | 92 |
| Balita 2 | 6 | 63 |
| Balita 3 | 12 | 79 |
| Balita 4 | 9 | 72 |
| Balita 5 | 11 | 75 |

Pembersihan Data

Tahapan selanjutnya adalah pembersihkan data, yaitu membuang data yang duplikat dan menghilangkan kesalahan data jika terdapat data hilang, terhubung dalam penelitian ini tidak ada data yang duplikat atau double dan juga tidak ada data yang hilang, maka dilanjutkan dengan proses selanjutnya yaitu transformasi data / permodelan data

Transformasi Data (Normalisasi)

Transformasi Data Pada tahap ini dilakukan proses perubahan data, tujuannya adalah agar data dapat diolah dengan menggunakan algoritma K-Means Clustering. karena perbedaan ukuran angka tinggi badan dan berat badan sangat berbeda. Solusi untuk mengurangi ukuran angka yaitu untuk menormalisasikan ukuran angka variabel tinggi badan dan berat badan hasil tranformasi data sebagai berikut

Tabel 4. 3 Data Hasil Normalisasi

| Balita | Berat Badan | Tinggi Badan | BB_Normalisasi | TB_Normalisasi |
|------------------------|-------------|--------------|----------------|----------------|
| Balita 1 | 12 | 92 | -0,41342297 | -0,34842571 |
| Balita 2 | 6 | 63 | -1,01455608 | -1,92950134 |
| Balita 3 | 12 | 79 | -0,41342297 | -1,05718375 |
| Balita 4 | 9 | 72 | -0,71398952 | -1,43882270 |
| Balita 5 | 11 | 75 | -0,51361182 | -1,27526315 |
| Balita 6 | 10 | 88 | -0,61380067 | -0,56650510 |
| | | | | |
| rata-rata | 16,12643678 | 98,3908046 | | |
| standar deviasi | 9,981150371 | 18,3419435 | | |

Proses Clustering

Setelah proses normalisasi data maka dapat dilakukan proses clustering status gizi balita. Proses clustering dengan memakai teknik algoritma K-Means akan dilakukan dalam pengujian 87 data sampel balita. Adapun langkah-langkah proses clustering untuk status gizi Balita yakni sebagai berikut :

1. Menentukan jumlah Cluster.

Tahap pertama adalah menentukan jumlah cluster, di penelitian ini dibagi menjadi 5 kelompok, cluster 0, cluster 1, cluster 2, cluster 3 dan cluster 4.

2. Tentukan nilai centroid awal.

Tahap selanjutnya menentukan nilai centroid awal dengan mengambil data yang mewakili cluster, seperti yang disajikan pada tabel 4.4

Tabel 4. Nilai Centroid Awal

| Kelas | Centroid | | Balita |
|-------|----------|----|--------|
| | BB | TB | |
| | | | |

| | | | |
|-----------|-------------|-------------|-----------|
| Cluster 0 | -0.11285641 | -0.02130661 | Balita 20 |
| Cluster 1 | 0.88903211 | 2.37756677 | Balita 57 |
| Cluster 2 | 7.80206292 | -1.38430285 | Balita 21 |
| Cluster 3 | 1.49016523 | 2.15948737 | Balita 56 |
| Cluster 4 | -1.01455608 | -1.92950134 | Balita 2 |

3. Hitung jarak setiap data yang ada terhadap setiap Cluster.

Setelah menentukan nilai centroid awal, langkah selanjutnya adalah hitung jarak setiap data yang ada terhadap setiap Cluster.

$$D_{(x,y)} = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i= 1,2,3,\dots,n$$

Dimana :

(x) merupakan objek data

(y) merupakan centroid

4. Tentukan anggota dari suatu cluster.

Langkah selanjutnya adalah menentukan suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat cluster nya. Misalkan untuk data pertama atau Balita 1, jarak terkecil diperoleh pada C0, sehingga data pertama akan menjadi anggota dari C0.

Berdasarkan hasil perhitungan jarak pada masing-masing klastering, masing-masing data akan menjadi anggota salah satu klaster yang mempunyai jarak/range paling dekat(nilai terakhir paling kecil) dari titik tengah klasternya Perhitungan hasil klaster untuk masing-masing klaster dapat dilihat pada Tabel di bawah ini.

Tabel 5. Data Hasil Perhitungan Anggota suatu Cluster

| Balita | Cluster |
|----------|---------|
| Balita 1 | c0 |
| Balita 2 | c4 |
| Balita 3 | c4 |
| Balita 4 | c4 |
| Balita 5 | c4 |
| Balita 6 | c4 |
| | |

Menghitung centroid klaster baru. Perhitungan centroid klaster baru dikerjakan dengan mencari nilai Average(rata-rata) dari setiap kriteria dari keseluruhan anggota yang menjadi anggota dari setiap klasternya,

Tabel 6. Nilai Centroid Baru

| Cluster | Berat | Tinggi | Keterangan |
|---------|-----------|-------------|-------------|
| c0 | 8 s/d 19 | 90 s/d 115 | Gizi Kurang |
| c1 | 20 s/d 31 | 138 s/d 142 | Gizi Buruk |
| c2 | 94 | 73 | Obesitas |
| c3 | 16 s/d 27 | 109 s/d 121 | Gizi Lebih |
| c4 | 6 s/d 13 | 61 s/d 89 | Gizi Baik |

| Centroid Baru | BB_Normalisasi | TB_Normalisasi |
|---------------|----------------|----------------|
| c0 | -0,01969835 | 0,24746669 |
| c1 | 0,62854110 | 1,88688813 |
| c2 | 7,80206292 | -1,38430285 |
| c3 | 1,13950425 | 1,27353982 |
| c4 | -0,71899896 | -1,36249491 |

Setelah nilai centroid baru dihitung, langkah selanjutnya adalah bandingkan dengan nilai centroid sebelumnya, jika nilainya sama maka proses iterasi dihentikan. Namun jika nilainya tidak sama langkah-langkah proses pengelompokan data diulangi kembali. Pada pengujian pertama ini datanya belum sama. tahapan berhenti pada perhitungan ke-9, pada klustering dengan hasil titik pusat terakhir

Tabel 8 Hasil perbandingan centroid awal & baru

| Centroid Baru | BB_Normalisasi | TB_Normalisasi |
|---------------|----------------|----------------|
| c0 | -0,08928256 | 0,23365387 |
| c1 | 0,88903211 | 2,28215704 |
| c2 | 7,80206292 | -1,38430285 |
| c3 | 0,48326726 | 0,84828499 |
| c4 | -0,64242606 | -1,16622345 |

Tabel 9 Keterangan Kategori

Klasifikasi Hasil Clustering

Perhitungan pada uji coba clustering k-means status gizi balita berhenti pada perhitungan ke-9. Hasil pengelompokan clustering k-means status gizi balita

Tabel 4. 10 Klasifikasi Hasil Clustering

| Balit a | Berat Badan | Tinggi Badan | BB_Normalisasi | TB_Normalisasi | Cluster |
|----------|-------------|--------------|----------------|--|---------|
| Balita 1 | 12 | 92 | -0,41342297 | -0,34842571 | c0 |
| Balita 2 | 6 | 63 | -1,01455608 | -1,92950134 | c4 |
| Balita 3 | 12 | 79 | -0,41342297 | -1,05718375 | c4 |
| Balita 4 | 9 | 72 | -0,71398 | $SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$ | c4 |
| Balita 5 | 11 | 75 | -0,51361 | | c4 |
| Balita 6 | 10 | 88 | -0,61380067 | -0,56650510 | c4 |
| | | | | | |

Menurut tabel di atas balita yang berada di cluster 0 terdapat 34 balita, cluster 1 terdapat 4 balita, cluster 2 terdapat 1 balita, cluster 3 terdapat 20 balita, cluster 4 terdapat 28 balita

Pengujian dan Tahap Validasi

Tahapan selanjutnya adalah perhitungan nilai akurasi clustering, pengukuran Index Davies Bouldin (DBI) adalah memaksimalkan jarak antar cluster diantara Ci dan Cj dan pada waktu yang bersamaan mencoba untuk meminimalkan jarak antar titik dalam sebuah cluster

Tabel 11. pengukuran jarak DBI

| |
|-------------------|
| Jarak |
| 0,66624592 |
| 0,84916070 |
| 0,25363768 |
| 0,28183626 |
| 0,16876837 |
| 0,60040112 |
| |

Untuk evaluasi dari Davies Bouldin Index dimulai dengan mencari nilai SSW. SSW atau yang disebut juga “Sum of Square Within” yaitu teknik menghitung jarak/range ke dalam klaster ke-i dengan mencari nilai average jarak enclidean setiap data ke- dalam cluster ke-i. bertujuan mencari nilai SSW, data harus dikelompokkan menurut klaster yang diperoleh pada perhitungan K-Means.

Keterangan :

Mi = jumlah data dalam cluster

X = data dalam cluster

D(x,c) = jarak data ke centroid

Xj = data pada cluster tersebut

Ci = centroid cluster ke-i

Tabel 12. Data perhitungan SSW

| Cluster | Jarak | SSW |
|---------|------------|------------|
| 0 | 0,66624592 | |
| 0 | 0,52337464 | |
| 0 | 0,15385491 | 0,34264711 |
| 0 | 0,07591175 | |
| 0 | 0,25604798 | |
| | | |
| 1 | 0,50261032 | |
| 1 | 0,61352169 | 0,32816337 |
| 1 | 0,09540974 | |
| 1 | 0,10111173 | |
| 2 | 0 | 0 |
| 3 | 0,16091155 | |
| 3 | 0,33913928 | 0,34574219 |
| 3 | 0,45937153 | |
| | | |

| | | |
|-------|------------|------------|
| 4 | 0,84916070 | |
| 4 | 0,25363768 | 0,37377357 |
| 4 | 0,28183626 | |
| | | |

Sedangkan untuk mencari nilai separasi antar kluster contohnya kluster i dan j maka dipakai formula perhitungan SSB(Sum of Square Between) dengan mencari nilai jarak/range antar titik pusat Ci dan Cj seperti pada formula :

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

SSB didapatkan dengan mencari jarak antar centroid, maka nilai perolehan SSB:

Tabel 13. Data perhitungan SSB

| SSB | C0 | C1 | C2 | C3 | C4 |
|-----|------------|------------|------------|------------|-------------|
| 0 | 0 | 2,27012441 | 8,05550231 | 0,83999091 | 1,50519906 |
| 1 | 2,27012441 | 0 | 7,82514684 | 1,49017923 | 3,77315413 |
| 2 | 8,05550231 | 7,82514684 | 0 | 7,65174610 | 8,44730446 |
| 3 | 0,83999091 | 1,49017923 | 7,65174610 | 0 | 2,307689257 |
| 4 | 1,50519906 | 3,77315413 | 8,44730446 | 2,30768926 | 0 |

Langkah ketiga yaitu menghitung Rasio, untuk mengetahui seberapa bagus nilai cluster satu dengan yang lain, Menghitung nilai rasio dengan menjumlahkan nilai ssw lalu dibagi nilai SSB Rumus :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Keterangan :

Rij = Rasio antar cluster

SSWi = cluster 1

SSWj = cluster 2

SSBij = separasi dari cluster 1 dan 2

Tabel 4. 2 Data Perhitungan Rasio antar cluster

| RASI | C0 | C1 | C2 | C3 | C4 |
|------|------------|------------|------------|------------|------------|
| 0 | 0 | 0,29549503 | 0,04253578 | 0,81951994 | 0,47596408 |
| 1 | 0,29549503 | 0 | 0,04193702 | 0,45223121 | 0,18603453 |
| 2 | 0,04253578 | 0,04193702 | 0 | 0,04518474 | 0,04424767 |
| 3 | 0,81951994 | 0,45223121 | 0,04518474 | 0 | 0,31179057 |
| 4 | 0,47596408 | 0,18603452 | 0,04424767 | 0,31179057 | 0 |

Langkah keempat yaitu DBI, fakta DBI yaitu semakin bagus Ketika mendekati nol, tetapi tidak negative. Rumus :

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

Keterangan :

K = cluster yang ada
 Max = dicari rasio antar cluster yang terbesar
 Rij = Rasio antar cluster I dan j

Tabel 4. 3 Hasil perhitungan DBI

Dari hasil perhitungan, diperoleh nilai DBI untuk K-Means Clustering status gizi balita **0,522483983** (Non-Negatif).

Berdasarkan Davies-Bouldin Indeks, nilai clustering yang baik adalah yang mempunyai nilai Davies-Bouldin Indeks paling kecil atau mendekati angka 0. Semakin kecil nilai *Sum of Square Within* maka hasil clustering yang diperoleh semakin optimal. Sedangkan pada *Sum of Square Between* semakin besar nilainya maka semakin optimal. Maka dari itu, artinya semakin jauhnya perbedaan/ketidaksamaan jarak/range data satu klaster dengan klaster lainnya. sama halnya dengan nilai DBI, DBI dinyatakan optimal secara esensial jika hasil akhir nilai yang didapat itu sekecil mungkin (non-negatif ≥ 0) untuk mengukur baik tidaknya klaster akhir yang diperoleh dalam perhitungan

Implementasi Rapid Miner

Pada penelitian ini penulis menggunakan tool rapidminer sebagai alat visualisasi hasil dataset. Adapun tahapan-tahapan dalam mem-visualisasikan hasil dataset yang dilakukan yaitu sebagai berikut :

Langkah awal dapat dilihat pada Gambar 1, peneliti memasukan hasil data perhitungan K-Means balita dalam tools Rapidminer



Gambar 1 Proses import data k-means

Kemudian, langkah yang ke-2 dapat dilihat pada gambar 4.2 yaitu tampilan data yang telah di lakukan proses import ke dalam tools rapidminer



Gambar 2 list dataset hasil import

Langkah selanjutnya yaitu proses design, dapat dilihat pada gambar 4.4. setelah data di import maka kita melakukan proses design dan menentukan jumlah K-cluster agar lebih memudahkan user untuk dapat melakukan proses clustering dengan algoritma K-Means.

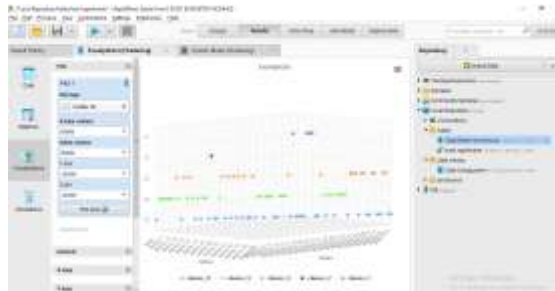


Gambar 3 Tampilan Design Retrieve Data dengan Algoritma K-Means



Gambar 4. Tampilan Hasil Clustering K-Means Balita

Lalu selanjutnya adalah tahapan memvisualisasikan hasil cluster model kedalam bentuk diagram scatter 3D, sehingga memudahkan dalam melakukan uji coba clusterin, gambar 4.7 adalah tampilan scatter 3D tools rapidminer dari hasil clustering dataset K-Means Gizi Balita



Gambar 4 hasil clustering dalam diagram scatter 3D

KESIMPULAN

Dilihat menurut hasil akhir dari penelitian, bisa kita ketahui Algoritma K-Means klastering dapat dipakai guna melakukan pemrosesan pengelompokan data status gizi balita layaknya teknik ataupun metode dalam menentukan status gizi balita.

1. Berdasarkan hasil pengujian, hasil akhir menunjukkan hasil DBI Kmeans sebesar 0.522483983 menggambarkan clustering terhadap pengelompokan data status gizi balita dinilai cukup baik. DBI dinyatakan optimal secara esensial jika hasil akhir nilai yang didapat itu se-rendah mungkin (non-negatif ≥ 0) untuk mengukur baik tidaknya cluster akhir yang diperoleh dalam perhitungan. Nilai tersebut mendefinisikan tingkat kemiripan keanggotaan didalam suatu klaster yang mempunyai tingkat kemiripan/kesamaan yang besar dan jarak perbedaan/ketidakmiripan antar klaster yang juga besar. Proses clustering data siswa dikelompokkan menurut setiap status gizinya.

2. Agar hasil tingkat akurasi dataset mudah dibaca dan dipahami oleh pembaca, maka diimplementasikan lah hasil dataset tersebut kedalam tools Rapidminer, sehingga visualisasi data dalam bentuk scatter 3D lah yang menunjukkan tingkat keakurasian hasil dataset tersebut

DAFTAR PUSTAKA

- A Aditya, I. J. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019
- Arhami, M., & Nair, M. (2020). *Data Mining - Algoritma dan Implementasi*. Yogyakarta: Penerbit ANDI.
- Ayyad, U. (2016). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: DEEPUBLISH
- E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi Yogyakarta, 2012.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *JOIN*, 36-41.
- Ediyanto, M. N. Mara, and N. Satyahadewi, —Pengklasifikasian Karakteristik Dengan Metod KMeans Cluster Analysis, *l Bul. Ilm.*, vol. 2, no. 2, pp. 133–136, 2013
- Firdaus. (2021). *Metodologi Penelitian Kuantitatif*. Riau: DOTPLUS Publisher
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concept and Techniques Third Edition*. United States of America: Morgan Kaufmann.
- J. O. Ong, “Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university,” *J. Ilm. Tek. Ind.*, vol. 12, no. 01, pp. 10–20, 2013
- Khasanah dan Sulistyawati. (2018). Karakteristik Ibu dengan Jurnal, S., Kesehatan, I., Khasanah, N. A., Sulistyawati, W., Tinggi, S., & Majapahit, I. K. (2018). Karakteristik Ibu dengan Kejadian Gizi Kurang pada Balita 6-24 Bulan di Kecamatan Selat , Kapuas Tahun 2016. *7(1)*, 1–8. *Kejadian G. Jurnal Strada Kesehatan Lmiah*, *7(1)*, 1–8
- Larose, D. T. (2017). *Discovering Knowledge in Data : An Introduction to Data Mining*. . John Willey & Sons, Inc
- L Ardiantoro, S. Z. (2019). Pemanfaatan Knowledge Data Discovery (Kdd) Pada Pola Permainan Atlet Bulutangkis. *Explore It*
- Prasetyowati, E. (2017). *Data Mining : Pengelompokan Data untuk Informasi dan Evaluasi*. Pamekasan: Duta Media
- Nasari, F., & Darma, S. (2013). Penerapan KMeans Clustering pada Data Penerimaan Mahasiswa Baru (Studi Kasus: UNIVERSITAS POTENSI UTAMA). *SEMNAS TEKNOMEDIA ONLINE*, *3(1)*, 2-1
- P. Sari, B. Pramono, and L. ode H. S. Sagala, —Improve K-Means Terhadap Status Nilai Gizi Pada Balita, *l semanTIK*, vol. 3, no. 1, pp. 143–148, 2017.
- Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data Menjadi Informasi dengan Rapid Miner*. CV Kekata Group
- Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data*

Menjadi Informasi dengan Rapid Miner. CV Kekata Group

Wahyudi Prayoga, H. S. (2019). Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means.

Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques 2nd edition. *Morgan Kaufmann, San Francisco*.