

Data Mining

Penerapan Algoritma *K-Means* dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi

Putri Apriyani, Arif Rinaldi Dikananda, Irfan Ali

Program Studi Manajemen Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia



INFORMASI ARTIKEL

Diterima Redaksi: 18 Februari 2023
Revisi Akhir: 01 Maret 2023
Diterbitkan Online: 03 Maret 2023

KATA KUNCI

Stunting; Data Mining; Klasterisasi; Algoritma K-Means

KORESPONDENSI

Phone: 085882450053
E-mail: putriapriyani014@gmail.com

A B S T R A K

Di Indonesia, dengan prevalensi stunting sebesar 37,2%, naik dari 35,6% pada tahun 2019 dan 36,8% pada tahun 2020, masalah stunting masih menjadi masalah kesehatan masyarakat yang signifikan. Ketika seorang anak pendek (kerdil), panjang atau tinggi badannya di bawah normal untuk usianya. Stunting berdampak buruk pada masalah kesehatan mental, tingkat IQ rendah, infeksi menular, dan tumbuh kembang anak. Permasalahan yang ada di Desa Tegalwangi saat ini adalah orang tua dan kader posyandu masih belum memiliki pemahaman dasar tentang stunting balita, balita dengan kasus stunting teridentifikasi saat pertumbuhan dan perkembangannya tidak sesuai dengan usianya dan tidak ada pengelompokan data berdasarkan faktor tersebut. K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan suatu data. Penggunaan algoritma K-Means pada penelitian dimana algoritma tersebut banyak dipakai oleh para peneliti sebagai proses klasterisasi untuk menentukan status stunting balita. Teknik pengumpulan data dalam penelitian ini yaitu teknik observasi pasif, dimana peneliti datang langsung ke tempat pelaksanaan Posyandu. Adapun data yang digunakan dalam penelitian ini sebanyak 395 dataset. Penelitian ini didukung dengan metode Knowledge Discovery In Databases (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data. Tujuan dari penelitian ini adalah untuk mengetahui kelompok atau cluster status stunting balita yang didapat berdasarkan parameter usia, berat badan dan tinggi badan. Hasil dari penelitian ini adalah didapat jumlah cluster optimal dengan $K=2$. Untuk cluster 0 terdapat 392 balita yaitu Shanum, Rizka, Nurjanah dan lainnya, cluster 1 dengan terdapat 3 balita yaitu Ezra, M Abidza dan Abd Mahmud. Dengan jumlah total balita stunting 287 balita dan jumlah total balita status normal 108 balita dan nilai DBI yang paling optimal sebesar 0.007 dimana nilai tersebut mendekati 0 yang berarti klaster yang di evaluasi menghasilkan klaster yang baik.

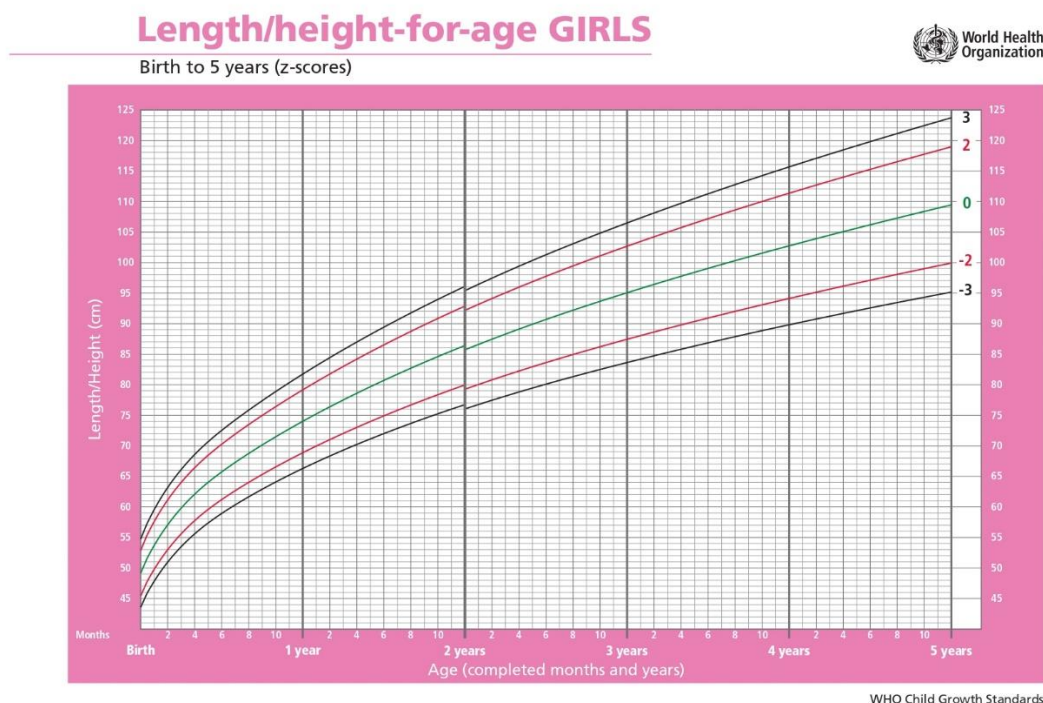
PENDAHULUAN

Stunting masih mempengaruhi sebagian besar penduduk di Indonesia, dimana frekuensinya adalah 37,2%, naik dari 35,6% pada tahun 2019 dan 36,8% pada tahun 2020. Menurut Kementerian Kesehatan Republik Indonesia, angka prevalensi stunting adalah 38,9 % pada tahun 2020 [1] Balita yang kerdil (kerdil) memiliki panjang atau tinggi yang lebih pendek dari seseorang seusianya. Penyakit ini ditentukan oleh panjang atau tinggi badan yang lebih dari minus dua standar deviasi dari standar rata-rata pertumbuhan anak Organisasi Kesehatan Dunia (WHO) [1] Ketidakseimbangan nutrisi menyebabkan stunting, masalah pertumbuhan fisik yang ditandai dengan penurunan laju pertumbuhan [2]. Faktor stunting pada anak dapat berkembang selama kehamilan, persalinan, menyusui, atau waktu setelah melahirkan. seperti MPASI yang tidak memberikan nutrisi yang cukup untuk balita [3]. Penilaian status stunting balita dapat ditentukan melalui pengukuran tubuh manusia yang dikenal dengan istilah “*Anthropometri*”. Jenis *anthropometri* antara lain Umur (U), Berat Badan (BB), Tinggi Badan (TB) [4].

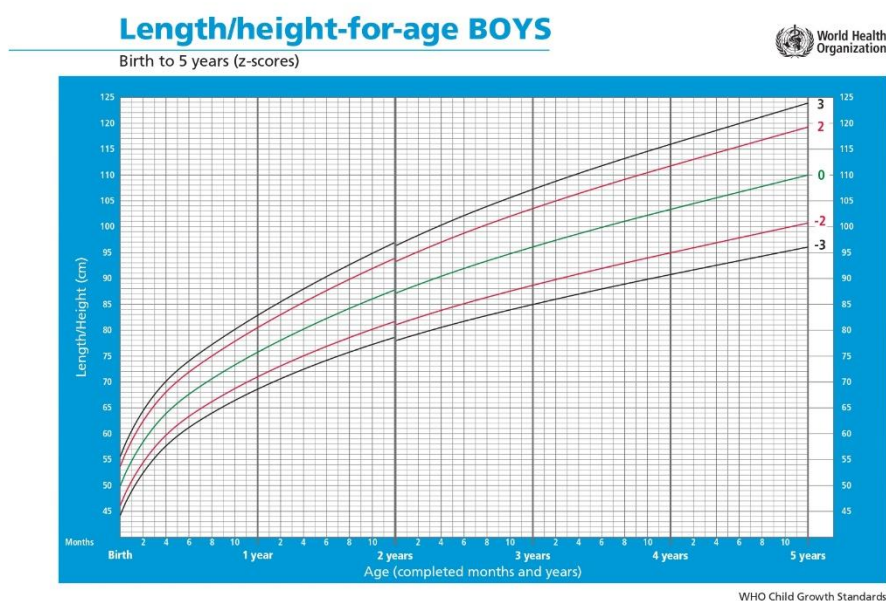
Menurut penelitian [5] tentang penggunaan algoritma *K-Medoids* untuk mengkategorikan balita stunting di Indonesia, algoritma tersebut dapat digunakan untuk membagi balita stunting menjadi dua cluster, tinggi dan rendah, dengan 28

provinsi memiliki cluster tertinggi dan 6 provinsi. memiliki cluster terendah. Nilai gizi balita dikelompokkan menjadi beberapa kategori berdasarkan parameter tinggi dan berat badan balita antara lain obesitas, gizi lebih, gizi baik, gizi buruk, dan gizi buruk, dengan menggunakan algoritma K-means *Clustering*, yang didasarkan pada penelitian tambahan dari artikel [6] algoritma K-means *Clustering* digunakan agar nilai gizi balita dapat dikelompokkan berdasarkan parameter tinggi badan dan berat badan balita menjadi beberapa kategori yaitu obesitas, gizi lebih, gizi baik, gizi kurang dan gizi buruk. Dari beberapa uraian di atas dapat diketahui bahwa relevansi kedua penelitian tersebut dengan proyek tugas akhir ini adalah sama-sama membahas tentang kasus stunting pada balita, sedangkan perbedaannya yaitu pada penelitian ini metode yang digunakan menggunakan algoritma K-means dan penentuan *cluster* terbaik yang di peroleh berdasarkan nilai DBI terkecil yang paling mendekati 0.

Pada posyandu Desa Tegalwangi, parameter yang umum digunakan dalam penentuan status stunting balita hanya berdasarkan Berat Badan menurut Umur (BB/U), yang terdapat pada Kartu Menuju Sehat (KMS). Setelah itu di catat pada formulir pemantauan status gizi balita dan di cocokkan status gizi balita tersebut berdasarkan tabel baku rujukan WHO. Tetapi Berat Badan menurut Umur (BB/U) tidak spesifik menunjukkan apakah balita tersebut tergolong pada status Normal atau Stunting. Sementara itu anak yang sehat semakin bertambah umurnya semakin bertambah berat badan dan tinggi badannya [4].



Gambar 1. Grafik tabel baku perempuan rujukan WHO



Gambar 2. Grafik Tabel baku Laki – Laki rujukan WHO

Tujuan dilakukannya penelitian ini adalah untuk mengetahui klusterisasi status stunting balita yang didapat berdasarkan parameter usia, berat badan dan tinggi badan yang akan di jadikan kategori pengelompokan yaitu terdiri dari Normal dan Stunting. Oleh karena itu, pada proyek tugas akhir ini digunakan metode K-Means *Clustering*. Klusterisasi (*clustering*) merupakan metode yang cukup populer dan paling banyak digunakan dalam pengolahan data karena kesederhanaannya serta efektif dalam mengelompokan data dengan ukuran besar berdasarkan kecepatan pemrosesan dengan menempatkan objek ke dalam kelas yang memiliki kesamaan/kemiripan. Algoritma K-Means untuk proses klusterisasi penentuan status stunting pada balita. Algoritma K-Means *clustering* yaitu salah satu algoritma *cluster* yang paling banyak digunakan karena kesederhanaan dan kinerjanya [7]. Adapun data yang digunakan dalam penelitian ini sebanyak 395 dataset. Proyek tugas akhir ini didukung dengan metode *Knowledge Discovery in Databases* (KDD) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat serta dapat dimengerti [8]. Sehingga dengan adanya proyek Tugas Akhir ini dapat membantu para petugas posyandu untuk mengolah data dan menentukan status stunting balita apakah termasuk balita dengan status Normal atau Stunting serta dapat dijadikan acuan pemerintah sekitar dalam mengatasi serta memperhatikan setiap rt/rw atau blok yang terdapat stunting di wilayah Desa tersebut.

Oleh karena itu diusulkan penelitian dengan judul “Klusterisasi Kasus Stunting Balita di Desa Tegalwangi Menggunakan Algoritma K-Means. Adapun yang menjadi alasan dilakukannya tugas akhir ini dengan judul tersebut adalah untuk mengetahui status stunting balita di wilayah desa karena belum ada penelitian sebelumnya dari desa tersebut kemudian alasan lainnya jumlah balita di Desa Tegalwangi cukup banyak sehingga perlu di ketahui juga status stunting nya.

TINJAUAN PUSTAKA

Stunting

Balita yang kerdil (kerdil) memiliki tubuh yang lebih pendek dari rata-rata anak seusianya dalam hal panjang atau tinggi badan.[9]. Menurut Keputusan Menteri Kesehatan Nomor 1995/MENKES/SK/XII/2010 tentang Standar Antropometri Penilaian Status Gizi Anak, pengertian pendek dan sangat pendek adalah status gizi yang didasarkan pada Indeks Panjang Badan menurut Umur (PB/U) atau Tinggi Badan menurut Umur (TB/U) yang merupakan istilah stunted (pendek) dan severely stunted (sangat pendek). Balita pendek adalah balita dengan status gizi berdasarkan panjang atau tinggi badan menurut umur bila dibandingkan dengan standar baku WHO, nilai Zscorenya kurang dari -2SD dan dikategorikan sangat pendek jika nilai Zscorenya kurang dari -3SD[10].Jadidapat disimpulkan bahwa *stunting* merupakan gangguan pertumbuhan yang dialami oleh balita yang mengakibatkan keterlambatan pertumbuhan anak yang tidak sesuai dengan standarnya sehingga mengakibatkan dampak baik jangka pendek maupun jangka panjang.

Data Mining

Proses penggalian pola yang berguna dari volume data yang sangat besar dikenal sebagai data mining. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi dan data dapat disimpan dalam database, data warehouse, atau penyimpanan data lainnya. Data mining berkaitan dengan bidang ilmu –ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing[11]

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan sebuah proses dalam menemukan informasi yang terdapat dalam database[12]. Dengan pertumbuhan teknologi yang sangat pesat ini kebutuhan-kebutuhan juga semakin banyak juga untuk mengimbangi pesatnya pertumbuhan teknologi seperti halnya kebutuhan akan tempat penyimpanan data yang sangat besar pula. Database atau tempat penyimpanan data haruslah memiliki ruang yang sangat besar yang dianalisa sehingga bermanfaat dan banar-benar menjadi tempat penyimpanan data yang berguna[12].

Clustering/Klasterisasi

Clustering/Klasterisasi merupakan teknik mengelompokkan data pada *data base* yang mengolah banyak data dalam database tersebut yang berdasarkan kriteria yang telah ditentukan sebelumnya. Pengelompokan data pada metode *clustering* menentukan *cluster* tanpa berdasarkan kelas-kelas tertentu. *Clustering* juga dapat dipakai untuk mengelompokkan data yang kelasnya belum diketahui sama sekali. Bahkan *clustering* dapat digunakan untuk memberikan label pada kelas data yang belum diketahui, sehingga *clustering* dapat digolongkan kedalam metode unsupervised learning. Konsep dari *clustering* sangatlah sederhana yaitu mengelompokkan objek kedalam *cluster* yang memiliki kesamaan, semakin tinggi kesamaan dari objek tersebut maka semakin akurat juga hasil dari *cluster* tersebut [12].

Algoritma K-means Clustering

K-means merupakan salah satu algoritma dari *clustering*. Algoritma ini termasuk yang cukup sederhana dalam melakukan pengklasteran, yang mana proses pengklasteran algoritma ini dengan cara mempartisi atau mengelompokkan dataset menjadi beberapa klaster [12]. Pada metode K-Means data dikelompokkan menjadi beberapa kelompok dimana setiap kelompok mempunyai karakteristik yang mirip atau sama dengan lainnya namun dengan kelompok lainnya memiliki karakteristik yang berbeda. Metode ini meminimalisasi perbedaan antar data di dalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain[6]

Dalam melakukan klasterisasi, metode K-means memiliki beberapa langkah, yaitu sebagai berikut[12]:

1. Tentukan jumlah *cluster* k
2. Inisialisasikan ke *cluster* k ke pusat *cluster*.

Penginisialisasian ini dapat dilakukan dengan beberapa cara. Cara yang paling sering digunakan dalam melakukan penginisialisasian ini dengan cara acak (*random*) sehingga pusat-pusat *cluster* diberi nilai awal dengan angka-angka acak. Pada tahap iterasi digunakan rumus seperti dibawah ini[13]

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \dots \dots \dots (1)$$

Keterangan :

V_{ij} = Centroid rata-rata *cluster* ke-I untuk variabel ke-j
 N_i = Jumlah anggota *cluster* ke-i
 i, k = Indeks dari *cluster*
 j = Indeks dari variable
 X_{kj} = nilai data ke-k variabel ke-j untuk *cluster* tersebut

3. Tempatkan seluruh data data kedalam *cluster* terdekat. Dua objek dikatakan dekat ditentukan dengan jarak dari kedua objek tersebut. Begitu juga dengan kedekatan data ke *cluster*.

4. Hitung jarak semua data ke setiap titik pusat *cluster* yang ada dengan menggunakan rumus *euclidean distance* sebagai berikut:

Keterangan:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots \dots \dots (2)$$

$D(i,j)$ = jarak data i ke pusat j

X_{ki} = data ke i pada atribut data ke k

X_{kj} = titik pusat ke j pada atribut ke k

Selanjutnya lakukan penghitungan kembali jarak pusat *cluster* dengan menggunakan kenggotaan *cluster* yang sekarang. Pusat *cluster* merupakan nilai rata-rata dari keseluruhan data yang terdapat pada *cluster* tertentu. Apabila memungkinkan juga dapat menggunakan median dari *cluster* yang sama. Oleh karena itu *mean* atau rata-rata bukanlah satu-satunya ukuran yang bisa digunakan dalam mengklasterkan suatu data. Setiap objek yang ada ditugaskan lagi untuk menggunakan pusat *cluster* yang baru. Apabila pada saat proses *cluster* tidak terjadi perubahan lagi pada pusat *cluster*, maka proses *clustering* dinyatakan selesai.

Rapidminer

RapidMiner adalah *software* yang dapat diakses oleh siapa saja dan bersifat terbuka (*open source*). RapidMiner ini dijadikan sebuah solusi untuk menganalisa terhadap data *processing*. Pada RapidMiner ini digunakan berbagai teknik seperti teknik deskriptif dan prediksi [14]. RapidMiner merupakan mesin pengolahan atau penambangan data yang dapat diintegrasikan ke dalam produknya sendiri dan tersedia sebagai perangkat lunak mandiri untuk analisis data [15].

Evaluasi Cluster

Metode yang digunakan dalam menentukan *evaluasi cluster* ini menggunakan *Davies-Bouldin Index* (DBI). *Davies-bouldin index* adalah salah satu metode *evaluasi internal* yang mengukur *evaluasi cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi [13].

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies bouldin index* (DBI) dari persamaan berikut [13]:

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

Dari persamaan tersebut, k merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh dari pengelompokan K-means yang digunakan.

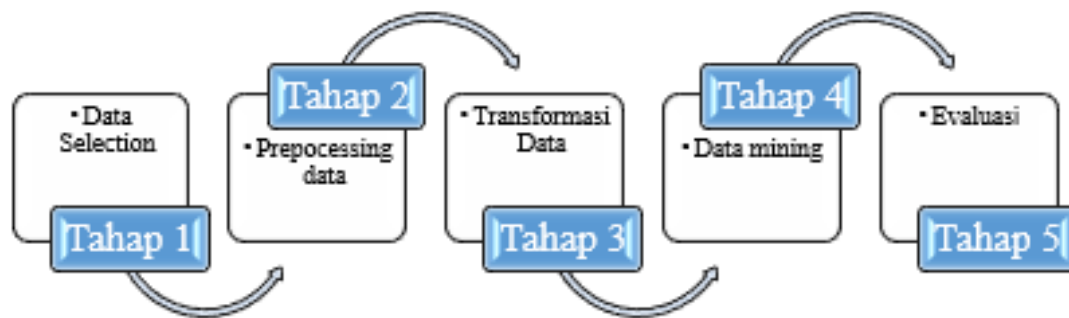
METODOLOGI

Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan pada penelitian ini, yaitu Teknik observasi yang digunakan pada proyek tugas akhir ini adalah teknik observasi pasif, dimana peneliti datang langsung ke tempat pelaksanaan Posyandu untuk mengamati kegiatan yang dilakukan tetapi tidak terlibat langsung dalam kegiatan tersebut. Kemudian hasil pencatatan Penimbangan dari tiap posyandu di ambil sebagai bahan proyek tugas akhir.

Tahap perancangan

Tahapan penelitian mencakup langkah – langkah penelitian dari awal sampai akhir, pada penelitian ini penulis menggunakan tahapan data mining Knowledge Discovery in Databases dalam pengolahan data. Fase-fase dimulai dari data mentah dan berakhir dengan pengetahuan atau informasi yang telah diolah, yang didapatkan sebagai hasil dari tahapan-tahapan berikut pada Gambar 3:



Gambar 3. Tahapan Data Mining Menggunakan KDD

Data Selection

Tahap pertama adalah proses pemilihan data yang digunakan dalam proses pengelompokan data yang bertujuan untuk memilih atau menyeleksi data yang di peroleh. Dataset yang digunakan adalah data penimbangan Posyandu yang ada di Desa Tegalwangi pada bulan januari 2023 yang diperoleh dari hasil observasi menggunakan *tools software Microsoft excel*, Dataset terdiri dari 395 record dengan 7 atribut. Atribut tersebut di antaranya No, Nama Balita, Alamat, Jenis Kelamin, Umur, Berat Badan dan Tinggi Badan terlihat pada Tabel 1, dibawah ini.

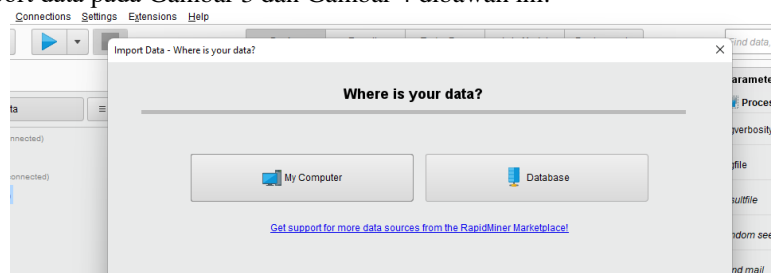
Tabel 1. Dataset Posyandu Desa Tegalwangi

No	Nama Balita	Alamat	Jenis Kelamin	Umur	Berat Badan	Tinggi Badan
1	Shanum	Posyandu Tunaswangi	Perempuan	23	9,9	85
2	Riska	Posyandu Tunaswangi	Perempuan	29	10,4	104
3	Nurjanah	Posyandu Tunaswangi	Perempuan	7	5,8	64
.....
394	Alisha.K	PosyanduTunasMekar1	Perempuan	58	15,1	104
395	Aulia.P	PosyanduTunasMekar1	Perempuan	58	16,9	102

Proses seleksi data pada penelitian ini dilakukan dengan menggunakan operator *Select Atribut* yang ada pada *tools RapidMiner*. Sebelum dilakukannya proses pemilihan data dataset yang telah disiapkan terlebih dahulu di import ke dalam RapidMiner. Selanjutnya pembuatan lembar proses dimana lembar proses ini berisi penerapan algoritma maupun pengujian menggunakan metode-metode pengujian yang ada pada operator RapidMiner. Langkah-langkah tersebut dapat dilihat pada penjelasan-penjelasan di bawah ini.

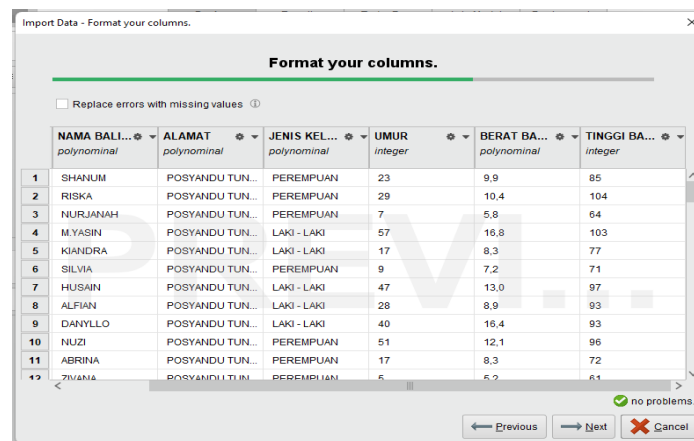
Import Data

Sebelum menerapkan algoritma K-Means maka data dengan format *.xls di import terlebih dahulu ke dalam RapidMiner dengan klik “Add Data” selanjutnya pilih lokasi penyimpanan dataset yang akan digunakan. Karena dataset tersebut berbentuk file yang memiliki format *.xls dan bukan berasal dari database seperti sql maka pilih lokasi dataset dari “My Computer”. Proses import data pada Gambar 3 dan Gambar 4 dibawah ini.



Gambar 3. Import Data

Selanjutnya pilih data import ke dalam RapidMiner. Apabila data yang dipilih telah terindikasi tidak ada masalah maka akan muncul “no problem” pada proses import data. Lalu lanjutkan hingga finish, sehingga data berhasil di import ke dalam RapidMiner.

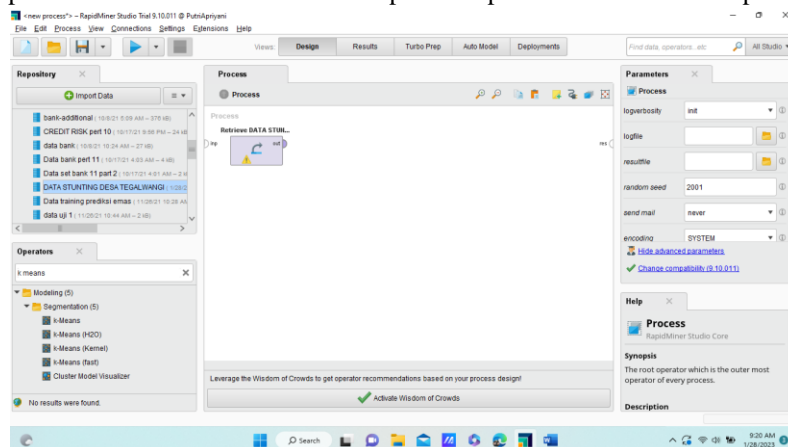


Gambar 4. Proses Import Data

Pada proses import data ini juga dilakukan penambahan id atau identitas pada dataset yang akan digunakan yaitu pada atribut Nama Balita di ubah dengan meng’Klik’ logo *setting* kemudian mengubah “change role type” di”Klik” menjadi “id”.

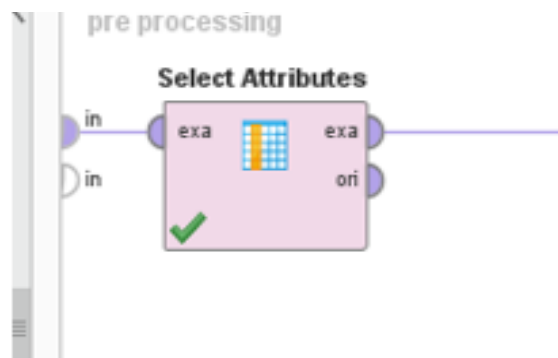
Membuat Lembar Proses

Setelah data di import, data dimasukkan kedalam halaman Proses. Dengan menyeret data yang telah di import. Contoh file data yang digunakan dalam proyek tugas akhir kali ini adalah Data Stunting Desa Tegalwangi. Kemudian Drag dan Drop down data ke halaman proses. Gambar 5 dibawah ini merupakan implementasi ke dalam RapidMiner.



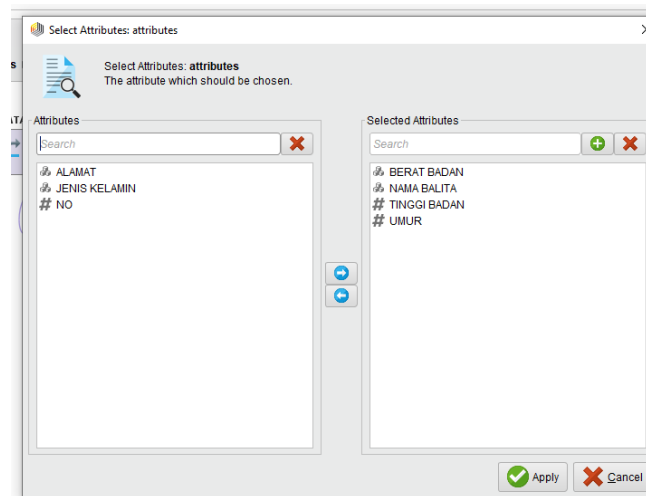
Gambar 5. Retrive Data yang telah di import

Select Attribut



Gambar 6. Operator Select Attribut

Pada dataset awal posyandu Desa Tegalwangi terdapat 7 Atribut. Atribut tersebut di antaranya No, Nama Balita, Alamat, Jenis Kelamin, Umur, Berat Badan dan Tinggi Badan. Operator *Select Atribut* pada *tools RapidMiner* akan menyeleksi atribut yang akan digunakan dengan pemilihan atribut berupa *subset* pada parameter operator *Select Atribut*. Proses seleksi data terlihat pada Gambar 6 dan Gambar 7.



Gambar 7. Parameter Operator Select Atribut

Row No.	NAMA BALITA	UMUR	BERAT BAD...	TINGGI BADAN
1	SHANUM	23	9,9	85
2	RISKA	29	10,4	104
3	NURJANAH	7	5,8	64
4	M.YASIN	57	16,8	103
5	KIANDRA	17	8,3	77
6	SILVIA	9	7,2	71
7	HUSAIN	47	13,0	97
8	ALFIAN	28	8,9	93
9	DANYLLO	40	16,4	93
10	NUZI	51	12,1	96
11	ABRINA	17	8,3	72
12	ZIVANA	5	5,2	61
13	ALFIYAH	30	10,1	83
14	ANDINI	32	12,1	88
15	GIBRAN	45	15,3	93

ExampleSet (395 examples, 1 special attribute, 3 regular attributes)

Gambar 8. Data yang telah diseleksi

Pre-Processing Data (Pembersihan data)

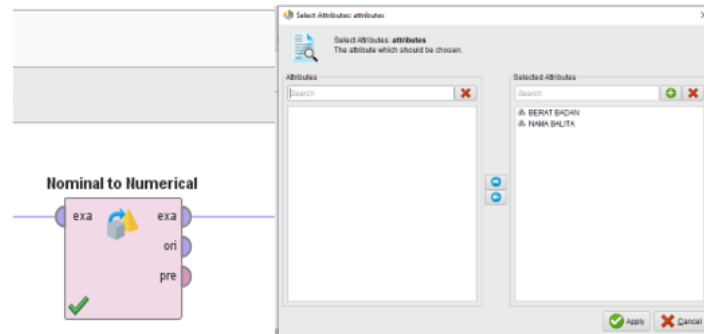
Tahap *Pre-Processing Data* yaitu proses pembersihan data yang variable nya tidak memenuhi dalam perhitungan (*Missing value*). Dalam tahap ini tidak ditemukan data yang *missing* pada dataset atau *No Missing value* dapat terlihat dari hasil *Statistics* dalam *tools RapidMiner* pada Gambar 9 berikut.

Name	Type	Missing	Statistics
NAMA BALITA	Polynomial	0	Least: ZULFAN ASHAR RAIHA... Most: AZKIYA (3) Values: AZKIYA (3), ADELIA (2), ... [381 more]
UMUR	Integer	0	Min: 0 Max: 60 Average: 29.678
BERAT BADAN	Polynomial	0	Least: 9.7 (1) Most: 9 (17) Values: 9 (17), 12 (10), ... [136 more]
TINGGI BADAN	Integer	0	Min: 8 Max: 113 Average: 83.916

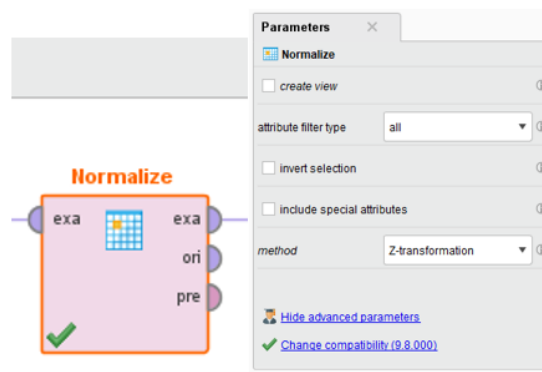
Gambar 9. Pengecekan *Missing Value*

Transformation Data (Data transformasi)

Pada tahap transformasi data digunakan operator *Nominal to Numerical* dan juga operator *Normalize* agar dataset yang telah melalui tahapan sebelumnya dapat di analisis dan diolah dengan menggunakan algoritma *K-Means Clustering*. Operator *Normalize* pada dataset proyek tugas akhir ini di gunakan untuk menormalkan nilai DBI sebelum di clusterkan dengan menggunakan algoritma *K-Means*. Proses *Transformation data* terlihat pada Gambar 10 dan Gambar 11.



Gambar 10. parameter operator *Nominal to Numerical*



Gambar 11. Parameter operator *Normalize*

Data Mining

Pada Tahap data mining dalam Penelitian ini proses klusterisasi menggunakan algoritma *K-Means* dalam mencari cluster terbaik penentuan kasus stunting balita di Desa Tegalwangi dan juga dengan bantuan *tools RapidMiner*. Operator yang digunakan dalam tahapan tersebut yaitu operator *clustering K-means* dan juga operator *cluster distance performance* untuk dengan metode evaluasi nilai DBI.

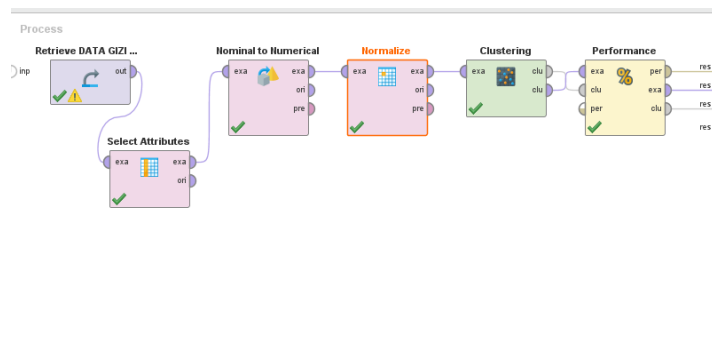
Evaluasi

Pola – pola yang telah dihasilkan dalam proses data mining ini ditampilkan dalam bentuk visualisasi scatter dan bar column agar lebih mudah dimengerti dengan tujuan agar informasi yang telah ditemukan tidak bertentangan dengan hipotesis yang telah dilakukan sebelumnya juga bisa dipahami dengan mudah.

HASIL DAN PEMBAHASAN

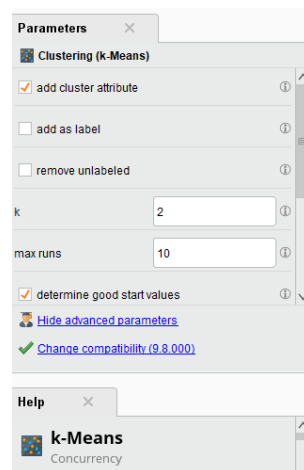
Hasil Penerapan Algoritma K-Means Clustering dengan Tools RapidMiner

Menerapkan algoritma *K-Means* ke dalam pada proses yang telah terbentuk dibawah ini untuk mengetahui pengelompokan/*clustering* dari proses yang telah dibuat. Operator *clustering* yang di gunakan adalah *K-Means*. Setelah selesai melakukan pembuatan model kemudian lakukan running pada proses untuk mendapatkan hasil *clustering*. Gambar 12 dibawah ini menunjukan model proses terbaik dengan menerapkan algoritma *K-Means*.



Gambar 12. Menerapkan K-Means

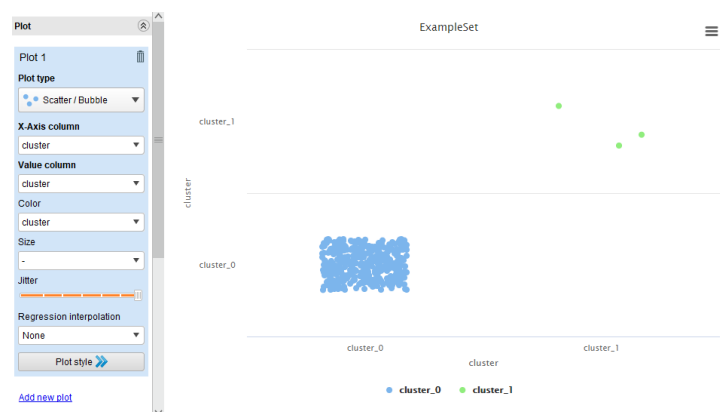
Selanjutnya melakukan pengaturan algoritma K-Means, yang diatur pada menu *Parameter Clustering K-Means*, seperti yang terlihat pada Gambar 13 berikut.



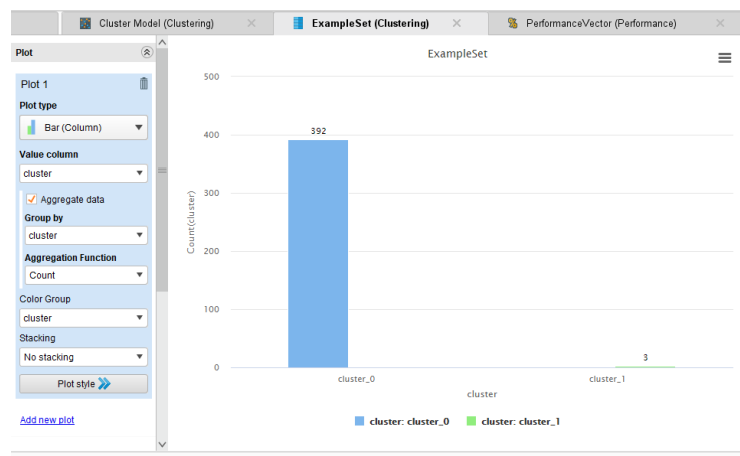
Gambar 13. Parameter penentuan cluster

Berdasarkan Gambar 13 merupakan proses memasukkan nilai K yang diinginkan, pada proyek tugas akhir ini dilakukan iterasi sebanyak 10 kali dengan nilai k yang di masukan yaitu k=2,3,4,5,6,7,8,9,10. Untuk menentukan *cluster* mana yang nilai DBI nya mendekati 0 (*cluster* terbaik). Setelah itu lakukan running pada proses untuk mendapatkan hasil clustering dari algoritma K-Means yang terapkan pada tools RapidMiner. Pada iterasi yang telah dilakukan didapat nilai K terbaik yaitu k=2, kemudian didapat hasil *cluster* dengan menggunakan algoritma K-Means.

Pada hasil pengujian data terdapatlah beberapa *output* yang dihasilkan oleh *software* RapidMiner 9.10. Pada gambar visualisasi, kita dapat melihat garis dimana ditentukan jumlah clustering grup kasus stunting di Desa Tegalwangi. Untuk visualisasi disisi kiri untuk sumbu x, sumbu y, dan warna custom pilih (*cluster*) semuanya akan terlihat pada gambar di bawah ini. Dalam visualisasi tersebut, kita dapat mengetahui yang mendapatkan posisi *clustering* pada level 0 dan 1.



Gambar 14. Visualisasi Cluster dengan Scatter

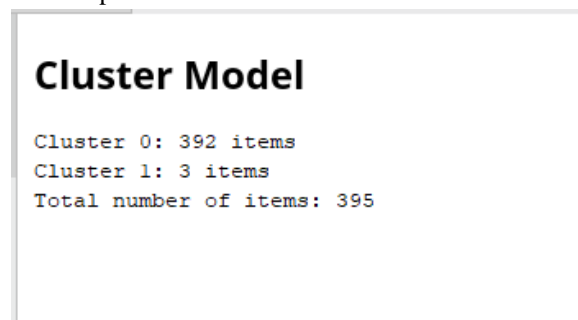


Gambar 15. Visualisasi dengan Bar Column

Pada Gambar 14 dan Gambar 15 diatas merupakan tampilan akhir dari aplikasi K-Means untuk mengelompokkan data *clustering*. Dari pengujian diatas dapat disimpulkan bahwa cluster 0 berjumlah 392 balita dilambangkan dengan warna biru, cluster 1 berjumlah 3 balita dilambangkan dengan warna hijau.

Cluster Model (clustering)

Pada *Cluster Model (clustering)* dapat dilihat beberapa tampilan hasil *cluster*, yaitu *text view* yang merupakan tampilan hasil pengelompokan berdasarkan *cluster* dan jumlah anggotanya *cluster* terdiri dari 392 items dan *cluster* 1 terdiri dari 3 items. Tampilan *text view* akan terlihat pada Gambar 16.

Gambar 16. Tampilan *text view*Gambar 17. Tampilan *Folder View*

Berdasarkan tampilan pada Gambar 17 *Folder view* merupakan tampilan data bagian-bagian *cluster* secara keseluruhan, dimana masing-masing anggota dari 2 *cluster* menampilkan nama balita. Setelah mengimplementasikan pengelompokkan kasus stunting dengan *tools RapidMiner*. Berikut Tabel 3 Merupakan indikator tinggi badan ideal anak berdasarkan standar Kementerian Kesehatan RI dan WHO (World Health Organization).

Tabel 3. Standar *Anthropometri* tinggi badan ideal berdasarkan usia

No	Standar <i>Anthropometri</i> (Tinggi Badan Ideal Berdasarkan Usia)
1	Bayi Usia 0 - 3 Bulan: Tinggi Badan 40,4 – 60 cm
2	Bayi Usia 4 - 6 Bulan: Tinggi Badan 60,5 – 66,0 cm
3	Bayi Usia 7 – 9 Bulan: Tinggi Badan: 67,5 – 70,5 cm
4	Bayi Usia 10 – 12 Bulan: Tinggi Badan: 72 – 74,5 cm
5	Balita Usia 13 – 24 Bulan: Tinggi Badan: 82 – 92 cm
6	Balita Usia 25 – 32 Bulan: Tinggi Badan: 83 – 95 cm
7	Balita Usia 33 – 44 Bulan: Tinggi Badan: 84 – 97 cm
8	Balita Usia 44 – 58 Bulan: Tinggi Badan: 85 – 98 cm

Berdasarkan Tabel 4 dibawah dapat disimpulkan bahwa informasi status Normal dan Stunting pada balita dari 395 data diperoleh anggota *Cluster* sebagai berikut .

Tabel 4. informasi jumlah balita Normal dan Stunting

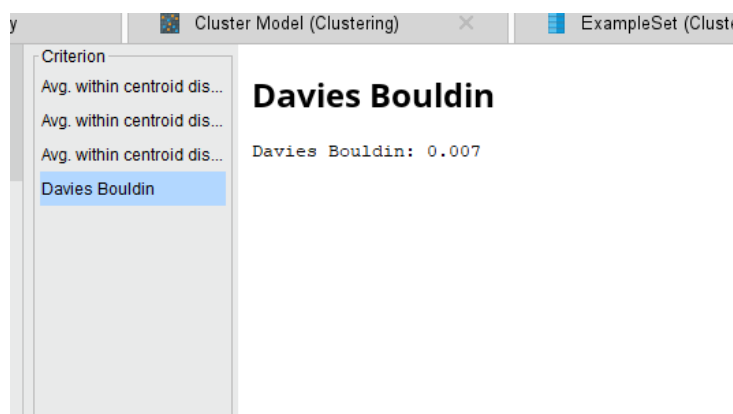
Cluster	Anggota Cluster	Informasi
C0	392 Balita	Jumlah Balita Normal: 285 Balita Jumlah Balita Stunting: 107 Balita
C1	3 Balita	Jumlah Balita Normal: 2 Balita Jumlah Balita Stunting: 1 Balita
Jumlah		Jumlah total balita Normal: 287 balita Jumlah total balita Stunting: 108 balita

Hasil Evaluasi Cluster teroptimal menggunakan algoritma K-Means dan perhitungan DBI

Berdasarkan Tabel 5 dibawah ini Melalui penggunaan metode k-means clustering, dan dibantu menggunakan perhitungan *Davies-bouldin index* (DBI) untuk penentuan cluster yang paling optimal. Didapat 10 kali percobaan iterasi untuk menentukan nilai DBI terbaik.

Tabel 5. Nilai *Davies-bouldin index* (DBI)

K-Means		
K	Avg. within centroid distance	DBI
2	0.990	0.007
3	0.982	0.028
4	0.968	0.029
5	0.961	0.024
6	0.954	0.022
7	0.945	0.027
8	0.939	0.017
9	0.932	0.018
10	0.924	0.020



Gambar 18. nilai DBI iterasi ke 2

Pada Gambar 18 dapat diketahui informasi mengenai nilai DBI dan performace dari *clustering*. Informasi pengujian mencari nilai *Davies-bouldin index* (DBI) menggunakan perhitungan dengan menggunakan bantuan *tools RapidMiner*. Hasil DBI yang terbaik dengan nilai $K=2$ di dapat sebesar 0.007 mendekati 0 yang artinya semakin kecil nilai DBI yang didapat (non-negatif ≥ 0), Semakin baik clusternya. Diperoleh dari pengelompokkan K-Means yang digunakan. Untuk hasil perhitungan dengan menggunakan *tools RapidMiner*.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan sebagai berikut, Berdasarkan hasil pengujian data menggunakan *tools RapidMiner* pada pengelompokkan kasus Stunting balita di Desa Tegalwangi per akhir bulan Januari 2023 menggunakan algoritma *K-Means Clustering* melalui 3 parameter yaitu umur, berat badan dan tinggi badan dimana dibagi menjadi 2 *cluster* yaitu *cluster 0* terdapat 392 balita yaitu Shanum, Rizka, Nurjanah, dan lainnya, *cluster 1* terdapat 3 balita yaitu Ezra, M Abidza dan Abd Mahmud. Jumlah balita status Normal sebanyak 287 balita dan jumlah balita status Stunting sebanyak 108 balita berdasarkan Standar *Antropometri* standar ideal balita. Sehingga perlu adanya pendampingan dari posyandu serta puskesmas terkait kepada Orang Tua balita sehingga balita yang termasuk stunting (pendek) bisa berkurang bahkan tidak ada di bulan berikutnya. Berdasarkan hasil pengujian nilai Davies Bouldin Index (DBI) perhitungan menggunakan bantuan *tools RapidMiner* menghasilkan nilai evaluasi performa yang optimal berada di $K=2$ memiliki nilai 0.007 dimana nilai tersebut mendekati 0 yang berarti klaster yang di evaluasi menghasilkan klaster yang baik.

Saran

Adapun saran untuk pengembangan lebih lanjut dari proyek tugas akhir ini yaitu, Diharapkan penelitian ini dapat dikembangkan dengan cara memodifikasi atau menggabungkan algoritma K-Means dengan algoritma lainnya. Selain itu jumlah data yang digunakan dapat ditambah dengan data yang lebih banyak lagi, dikarenakan pada metode K-Means semakin banyak data semakin optimal hasilnya. Diharapkan pada penelitian selanjutnya dapat dilakukan secara lebih detail dengan menggunakan data per Kecamatan yang ada di Kabupaten/Kota Cirebon dengan data kasus Stunting di Desa Tegalwangi pada bulan-bulan berikutnya, dengan menggunakan Algoritma yang berbeda misalnya *K-Medoids* atau *Fuzzy C-Means*. Uji validasi dapat dilakukan dengan metode lainnya seperti *Silhouette* dsb. Penelitian selanjutnya diharapkan untuk mencari informasi mengenai perbedaan hasil evaluasi DBI yang dilakukan secara perhitungan manual dan menggunakan *tools RapidMiner*.

DAFTAR PUSTAKA

- [1] Titimeidara & Hadikurniawati, "PENERAPAN ALGORITMA KMEANS CLUSTERING DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI STATUS STUNTING PADA BALITA," 2021, Diakses: 19 Januari 2023. [Daring]. Tersedia pada: repository.upnjatim.ac.id

- [2] G. Apriluana dan S. Fikawati, "Analisis Faktor-Faktor Risiko terhadap Kejadian Stunting pada Balita (0-59 Bulan) di Negara Berkembang dan Asia Tenggara," *Media Penelitian dan Pengembangan Kesehatan*, vol. 28, no. 4, hlm. 247–256, Des 2018, doi: 10.22435/mpk.v28i4.472.
- [3] T. Prasetya, I. Ali, C. L. Rohmat, dan O. Nurdian, "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS FOR EDUCATORS AND PROFESSIONALS*, vol. 4, no. 2, hlm. 93–104, 2020.
- [4] H. Saleh, M. Faisal, dan R. I. Musa, "KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR," <http://ejournal.catursakti.ac.id/>, vol. 4, no. 2, 2019, Diakses: 19 Januari 2023. [Daring]. Tersedia pada: <http://ejournal.catursakti.ac.id/index.php/simtek/article/view/60/58>
- [5] H. Pohan dkk., "Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia," *JUKI : Jurnal Komputer dan Informatika*, 2021.
- [6] E. Irfiani, S. Sulistia Rani, J. Kamal Raya No, R. Road Barat Cengkareng Jakarta Barat, S. Nusa Mandiri Jl Kramat Raya No, dan J. Pusat, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," 2018.
- [7] dkk Gustientiedina, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," 2018. Diakses: 20 Januari 2023. [Daring]. Tersedia pada: <http://repository.potensi-utama.ac.id/jspui/bitstream/123456789/4745/3/BAB%20II.pdf>
- [8] U. Fayyad, G. Piatetsky-Shapiro, dan P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework," 1996. [Daring]. Tersedia pada: www.aaai.org
- [9] A. Dwi, N. Yadika, K. N. Berawi, dan S. H. Nasution, "Pengaruh Stunting terhadap Perkembangan Kognitif dan Prestasi Belajar," 2019.
- [10] Kemenkes RI 2020, "PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA STANDAR ANTROPOMETRI ANAK."
- [11] R. A. Indraputra dan R. Fitriana, "K-Means Clustering Data COVID-19," *Jurnal Teknik Industri*, 2020.
- [12] ary bagus jiwandono, "Analisa Pengelompokan Kelas BPJS Kesehatan Dengan Menggunakan Metode K-Means," 2021.
- [13] Z. Nabila, A. Rahman Isnain, dan Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, hlm. 100, 2021, [Daring]. Tersedia pada: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [14] rahmawati, "Menentukan Tingkat Kesejahteraan Provinsi Kalimantan Tengah Dengan Penerapan Algoritma K-Means Clustering Menggunakan Rapidminer," 2023.
- [15] D. Pascalina, R. Widhiastono, dan C. Julian, "Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner," *Technomedia Journal*, vol. 7, no. 3, hlm. 293–302, Des 2022, doi: 10.33050/tmj.v7i3.1914.