

BAB II LANDASAN TEORI

2.1 Clustering

Pengelompokan objek (objek clustering) adalah salah satu proses dari objek mining yang bertujuan untuk mempartisi objek yang ada kedalam satu atau lebih cluster objek berdasarkan karakteristiknya. Objek dengan karakteristik yang sama dikelompokkan dalam satu cluster dan objek dengan karakteristik berbeda dikelompokkan kedalam cluster yang lain [2].

Perbedaan antara klasifikasi dan clustering terletak pada proses pengelompokan obyek. Jika pada klasifikasi proses pengelompokan obyek dilakukan dengan membagi obyek berdasarkan kelompok / kategori yang telah didefinisikan sebelumnya, maka proses pengelompokan obyek pada *clustering* dilakukan dengan melihat kemiripan antar obyek, sehingga kategori belum terdefinisi sebelumnya.

Clustering merupakan jenis khusus dari klasifikasi tapi ada perbedaan antara clustering dan klasifikasi seperti pada Tabel 2.1. [4]

Tabel 2. 1 Perbedaan antara clustering dan klasifikasi

	Clustering	Klasifikasi
Jenis data	Tidak berlabel	Berlabel
Metode	Unsupervised Learning	Supervised Learning
Tujuan	Membentuk grup berdasarkan pola kemiripan antar sampel	Membentuk class berdasarkan Feature yang disediakan oleh data set
Fungsi	Untuk memprediksi class dimasa depan	Untuk memahami data
Algoritama	K-Mean, KNN, Naïve Bayes Classifier, dsb	Decision Tree, SVM, Neural Network, dsb

Clustering sebagai proses pengorganisasian objek data ke dalam set kelas yang saling berhubungan, yang disebut cluster. Clustering merupakan contoh dari klasifikasi tanpa arahan (unsupervised). Klasifikasi merujuk kepada prosedur yang menetapkan objek data set kelas. Unsupervised berarti bahwa pengelompokan tidak tergantung pada standar kelas dan pelatihan atau training.

Clustering merupakan salah satu teknik data mining yang digunakan untuk mendapatkan kelompok-kelompok dari objek-objek yang mempunyai karakteristik yang umum di data yang cukup besar. Tujuan utama dari metode clustering adalah pengelompokan sejumlah data atau objek ke dalam cluster atau grup sehingga dalam setiap cluster akan berisi data yang semirip mungkin. Clustering melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode unsupervised learning.

Clustering dapat dibagi menjadi dua, yaitu hierarchical clustering dan non-hierarchical clustering. Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut.

Berbeda dengan metode hierarchical clustering, metode non-hierarchical clustering justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang

diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan K-Means Clustering

2.1.1 Algoritma Clustering

a. K-Means

K-Means Cluster Analysis merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain. Tujuan pengelompokan adalah untuk meminimalkan objective function yang di set dalam proses clustering, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu cluster dan memaksimalkan variasi antar cluster. [2]

Algoritma K-Means dimulai dengan pemilihan secara acak K , K disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid secara random, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidean hingga ditemukan jarak yang paling dekat dari setiap data dengan

centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah. [5]

Jarak Euclidean yang dirumuskan sebagai berikut:

$$d = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2} \dots\dots\dots (1)$$

b. Fuzzy Clustering C-Means (FCM)

Pada logika fuzzy, metode yang dapat digunakan untuk melakukan pengelompokan sejumlah data dikenal dengan nama fuzzy clustering. Fuzzy Clustering lebih alami jika dibandingkan dengan pengklasteran secara klasik. Suatu algoritma clustering dikatakan sebagai fuzzy clustering jika algoritma tersebut menggunakan parameter strategi adaptasi secara soft competitive. Sebagian besar algoritma fuzzyclustering didasarkan atas optimasi fungsi obyektif atau modifikasi dari fungsi obyektif tersebut. [6]

c. K-Medoids

Algoritma K-medoids atau dikenal pula dengan PAM (Partitioning Around Medoids) menggunakan metode partisi clustering untuk mengelompokkan sekumpulan n objek menjadi sejumlah k cluster. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster. Objek yang terpilih untuk mewakili sebuah cluster disebut medoids. Cluster dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek non-medoids.

2.2 Data Mining

Data mining disisi lain adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [7].

2.3 Malnutrisi

Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia. Pada salah satu riset skala nasional yaitu Riset Kesehatan Dasar (RISKESDAS) Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan Republik Indonesia menunjukkan bahwa jumlah balita di Indonesia usia 12 bulan sampai dengan 59 bulan penderita gizi buruk pada tahun 2013 mencapai 28,1 persen. Jumlah balita yang meninggal dunia akibat kekurangan gizi adalah 29 dari 1000 kelahiran. [1]

Gizi kurang atau malnutrisi pada balita membawa dampak negatif terhadap pertumbuhan fisik maupun mental, yang selanjutnya akan menghambat beberapa proses belajar yang dilakukan oleh balita seperti belajar berbicara, berjalan, makan dan lain-lain. Dampak yang lebih parah dari malnutrisi pada balita adalah timbulnya kecacatan, tingginya angka kesakitan dan percepatan kematian. [8]

Kepedulian orang tua dan aparat desa (dalam hal ini petugas Pusat Pelayanan Kesehatan Masyarakat – PUSKESMAS) untuk memantau gizi balita sangat diperlukan. Malnutrisi pada balita tidak terjadi secara tiba-tiba seperti

penyakit pada umumnya. Tanda-tanda seperti berat badan kurang dari standar, stunting bisa menjadi indikator awal terjadinya malnutrisi pada balita. [1]

2.4 Penelitian Terkait

Tabel 2.2

Penelitian Terkait

Tabel 2. 2 Penelitian terkait

NO	1
Nama & Tahun	Austin Almayda, Sudin Saepudin 2021
Judul	Penerapan Data Mining K-Means Clustering untuk Mengelompokkan Berbagai Jenis Merk Smartphone
Metode	K-Means
Hasil	Terbentuk beberapa cluster produk smartphone. Hal ini diharapkan bisa membantu kelak siapa saja yang akan membeli smartphone
Perbedaan	Penelitian ini mengelompokkan berbagai jenis merk smartphone
NO	2

Nama & Tahun	Darmawan, Izzati Muhimmah, Kariyam 2017
Judul	Identifikasi Masalah Kesehatan di Sulawesi Tenggara
Metode	K-Means
Hasil	Pengelompokan kecamatan di Sulawesi Tenggara berbasis indikator kesehatan
Perbedaan	Penelitian ini mengelompokan kecamatan di Sulawesi Tenggara berbasis indikator kesehatan
NO	3
Nama & Tahun	Windha Mega Pradnya Duhita 2015
Judul	Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita
Metode	K-Means
Hasil	Pengelompokan balita di desa Karang Songo kedalam 5 cluster status gizi menggunakan parameter yang berbeda yaitu hanya tinggi badan balita dan berat badan balita yang tingkat akurasi hasilnya 34%

Perbedaan	Parameter yang digunakan untuk melakukan pengelompokan status gizi balita berjumlah 2 yaitu tinggi badan balita (TB) dan berat badan balita (BB)
NO	4
Nama & Tahun	Ediyanto, Muhlasah Novitasari Mara, Neva Satyahadewi 2013
Judul	Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis
Metode	K-Means K-Means
Hasil	Metode K-Means Cluster Analysis cukup efektif diterapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian
Perbedaan	Penelitian ini menerapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian
NO	5
Nama & Tahun	Muhardi, Nisar 2015

Judul	Penentuan Penerima Beasiswa Dengan Algoritma Fuzzy C-Means Di Universitas Megow Pak Tulang Bawang
Metode	Fuzzy C-Means
Hasil	Penentuan beasiswa di kelompokan menjadi tiga cluster yaitu menerima, dipertimbangkan dan tidak berhak menerima beasiswa
Perbedaan	Penelitian ini menentukan beasiswa menggunakan metode Fuzzy C-Means
NO	6
Nama & Tahun	Fitri Kurnia Rahim 2014
Judul	Faktor Risiko Underweight Balita Umur 7-59 Bulan
Metode	Cluster random sampling
Hasil	Menunjukkan balita yang tergolong status gizi underweight sebanyak 31,40 %, yang mengalami diare kronik 14,90 %, dan pneumonia 8,80 %

Perbedaan	Penelitian ini melakukan survei pada balita umur 7-59 bulan di wilayah Puskesmas Leuwimunding sebanyak 200 sampel, menggunakan cluster random sampling
NO	7
Nama & Tahun	Heni Sulastri, Acep Irham Gufroni 2017
Judul	Penerapan Data Mining dalam Pengelompokan Penderita Thalassaemia
Metode	K-Means
Hasil	Pengelompokan data penderita Thalassaemia menggunakan algoritma K-means clustering menjadi 3 kelompok
Perbedaan	Penelitian ini mengelompokkan data penderita Thalassaemia
NO	8
Nama & Tahun	Totok Suprawoto 2016

Judul	Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran
Metode	K-Means
Hasil	Clustering data mahasiswa menggunakan metode K-Means, terdiri dari nilai rerata ujian nasional (UN) dan indeks prestasi kumulatif (IPK) mahasiswa
Perbedaan	Penelitian ini mengelompokkan data mahasiswa
NO	9
Nama & Tahun	Liliana Swastina 2013
Judul	Penerapan Algoritma C4.5 untuk Penentuan Jurusan Mahasiswa
Metode	Algoritma C4.5
Hasil	Algoritma Decision Tree C4.5 akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat akurasi 93,31 % dan akurasi rekomendasi jurusan sebesar 82,64%

Perbedaan	Penelitian ini menggunakan Algoritma Decision Tree C4.5 untuk penentuan kesesuaian jurusan mahasiswa
NO	10
Nama & Tahun	Santi Thomas, Noviyanti P 2021
Judul	Analisis Data untuk Pengelompokan Mahasiswa dengan Metode K-Mean (Studi Kasus : Institut Shanti Bhuana)
Metode	K-Mean
Hasil	Perhitungan secara teoritis dengan menggunakan excel dan pengujian dengan menggunakan software Rapid Miner
Perbedaan	Penelitian ini mengelompokkan mahasiswa dengan menggunakan excel