

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Prinsip nilai gizi seimbang adalah dengan membiasakan pola konsumsi beragam secara seimbang sesuai dengan masing-masing kebutuhan anggota keluarga. Kebiasaan pola hidup sehat secara rutin dengan melakukan aktifitas fisik dan memantau berat badan secara teratur. Pola makanan pada balita berpengaruh terhadap keadaan gizi. Gizi yang optimal sangat penting untuk pertumbuhan normal serta perkembangan fisik dan kecerdasan pada anak. Dengan gizi yang baik tubuh tidak mudah terkena serangan penyakit dan infeksi serta terlindung dari penyakit kronis [1].

Data Mining merupakan penyelesaian masalah dengan menganalisa data yang ada pada *database*, dimana data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer. *Data Mining* merupakan ekstraksi informasi yang tersembunyi dari *database* yang besar. *Data Mining* adalah teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan agar fokus pada informasi yang paling penting dalam gudang data mereka [2].

Data mining telah dipakai diberbagai bidang seperti ilmu sains, bisnis dan industri, teknik, kesehatan, serta pertahanan dan keamanan. Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan, menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam limpahan data tersebut, yang

kemudian melahirkan data mining. data mining merupakan proses untuk menemukan pengetahuan (*knowledge discovery*) yang ditambang dari sekumpulan data yang volumenya sangat besar.

*Cluster* juga dikenal dengan istilah *clustering* adalah metode yang digunakan untuk membagi rangking data menjadi beberapa grup berdasarkan kesamaan yang telah ditentukan sebelumnya. *Clustering* adalah proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelompok-kelompoknya berdasarkan tingkat kesamaan. Sehingga dalam suatu *cluster* merupakan data dengan karakteristik yang hampir sama [2].

*K-Means* adalah suatu teknik pengklusteran data yang mana keberadaan tiap data dalam suatu *cluster* ditentukan oleh nilai keanggotaannya. Pada kondisi awal, pusat cluster masih belum akurat. Tiap data memiliki derajat keanggotaan untuk tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap data secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi obyektif [3].

Permasalahan yang terjadi di Puskesmas Rambah Hilir 1 yaitu sulitnya menentukan status gizi balita, dimana dokter umum di Puskesmas hanya menentukan berdasarkan Berat Badan (BB), Tinggi Badan dan usia yang terdapat pada Kartu Menuju Sehat (KMS) dan mencocokkan berdasarkan status gizi balita pada tabel buku rujukan nilai Z-score (WHO 2005). Sehingga dokter umum di Puskesmas maupun bidan di puskesmas sulit mengetahui apakah balita tersebut tergolong dalam anak yang gizi normal, gizi tidak normal atau obesitas [1].

Dengan menerapkan data mining dengan metode *k-means clustering* dokter umum di puskesmas dapat dengan mudah menggunakan aplikasi data mining tersebut, serta tingkat efisiensi dan interaktifitas, dimana metode ini bisa mengelompokkan status gizi balita dengan mudah. Algoritma *K-Means* merupakan salah satu metode *clustering non-hirarki*. Algoritma ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain. Algoritma ini sederhana untuk diterapkan dan dijalankan, relatif cepat, mudah diadaptasi dan umum digunakan dalam praktek [4].

Untuk itu peneliti mencoba membuat sebuah Data Mining dengan menggunakan Algoritma *K-means Clustering* dengan judul “**Data Mining Penentuan Potensi Status Gizi Balita Menggunakan Algoritma *K-Means Clustering*”**”.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah dijelaskan di atas, maka penulis merumuskan masalah sebagai berikut :

1. Bagaimana cara menerapkan metode *k-means clustering* dalam penentuan potensi status gizi balit di Puskesmas Rambah Hilir 1 ?
2. Bagaimana cara mengelompokkan status gizi balita dengan menggunakan algoritma *k-means clustering* ?

### **1.3 Ruang Lingkup Permasalahan**

Agar permasalahan tidak menyimpang dari apa yang telah ditetapkan, maka permasalahan yang dibahas sebagai berikut :

1. Variabel yang digunakan hanya tinggi badan, berat badan dan usia.
2. Batas usia balita hanya dari 24 bulan sampai 60 bulan, atau dari 2 tahun sampai 5 tahun.
3. *Cluster* yang digunakan hanya gizi normal, gizi tidak normal dan obesitas.
4. Bahasa pemrograman yang digunakan dalam sistem ini adalah PHP dan *database* yang digunakan adalah MYSQL.

### **1.4 Tujuan dan Manfaat penelitian**

#### **1.4.1 Tujuan Penelitian**

Adapun tujuan penelitian dalam tugas akhir ini adalah :

1. Merancang dan membangun data mining yang dapat diakses oleh dokter serta pegawai Puskesmas dengan mengimplementasikan metode *k-means clustering*
2. Untuk mendapatkan informasi yang berguna tentang penentuan status gizi balita dan mempermudah dokter serta pegawai Puskesmas dalam penentuan gizi balita.

#### **1.4.2 Manfaat Penelitian**

1. Manfaat bagi instansi
  - a. Dapat digunakan sebagai data mining untuk mengelompokkan status gizi balita.

- b. Mempermudah dokter serta pegawai Puskesmas untuk menentukan status gizi balita.
2. Manfaat bagi mahasiswa (peneliti)
  - a. Mengetahui bagaimana menganalisis data mining penentuan status gizi balita
  - b. Merancang dan membangun data mining penentuan status gizi balita menggunakan algoritma *k-means clustering*.
  - c. Dapat menambah wawasan dan pengalaman peneliti.

### **1.5 Metode Pengumpulan Data**

Dalam melakukan penelitian diperlukan data yang berkaitan dengan masalah penelitian untuk membantu memecahkan masalah. Pada penelitian ini penulis menggunakan 3 teknik pengumpulan data [5], yaitu :

#### **1. Studi Kepustakaan**

Pada penelitian ini langkah pertama yang penulis lakukan adalah dengan mempelajari bahan-bahan yang berhubungan dengan bidang masalah tersebut diatas. Studi kepustakaan ini dilakukan untuk mendapatkan informasi yang berhubungan dengan masalah yang penulis teliti, dengan cara membaca literatur, buku teks dan dokumen-dokumen dari sistem yang sedang berjalan.

#### **2. Penelitian Lapangan**

Setelah studi kepustakaan dilakukan, langkah selanjutnya adalah dengan melakukan penelitian dibagian Datamining penentuan status gizi balita menggunakan algoritma *k-means clustering*.

### 3. Wawancara

Untuk memperkuat penelitian yang dilakukan penulis, maka penulis mengadakan wawancara dengan dokter serta pegawai Puskesmas untuk mendapatkan informasi data dan informasi untuk mendukung penelitian.

#### **1.6 Sistematika Penulisan**

Sistematika penulisan Tugas Akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan Tugas Akhir ini adalah sebagai berikut :

##### **BAB 1. PENDAHULUAN**

Dalam bab ini peneliti menerapkan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

##### **BAB 2. LANDASAN TEORI**

Bab ini membahas mengenai dasar-dasar teori yang digunakan sebagai landasan dalam pembuatan sistem aplikasi atau sistem informasi.

##### **BAB 3. METODOLOGI PENELITIAN**

Bab ini penelitian menguraikan tentang metodologi yang digunakan dalam data mining penentuan potensi status gizi balita menggunakan algoritma *k-means clustering*.

##### **BAB 4. ANALISA DAN PERANCANGAN**

Analisa dan perancangan dan usulan rancangan sistem yang baru digambarkan pada bab ini. Dimulai dari *Use case*, *Class Diagram*,

*Sequence Diagram*, serta *Activity diagram* dalam data mining penentuan potensi status gizi balita menggunakan algoritma *k-means clustering* yang digunakan, hingga program keluarannya semua tergambar dalam bab ini.

#### BAB 5. TESTING DAN IMPLEMENTASI

Testing dan implementasi program merupakan tahapan yang harus dilakukan pada bab ini dijelaskan langkah-langkah testing yang digunakan yaitu metode *black box* Testing dan implementasi program tersebut.

#### BAB 6. KESIMPULAN DAN SARAN

Bab ini berisi tentang hasil kesimpulan yang telah diperoleh dari hasil tugas akhir beserta pengembangan aplikasinya.

#### DAFTAR PUSTAKA

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 Konsep Dasar Data Mining**

Istilah data mining mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis hingga medis [2]. Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Dalam berbagai literatur, teori-teori pada data mining sudah ada sejak lama seperti antara lain *Naïve-Bayes* dan *Nearest Neighbour*, Pohon Keputusan, aturan asosiasi, *K-Means Clustering* dan *text mining* [2].

Data mining adalah suatu konsep yang digunakan untuk mencari nilai tambah yang disembunyi dalam *database*. Data mining merupakan proses semi otomatis yang digunakan teknik statistik, matematik, kecerdasan buatan dan *machine learning* untuk mengestraksi dan mengidentifikasi informasi pengetahuan potensial dan bermanfaat yang tersimpan dalam *database* besar [4].

Data Mining merupakan penyelesaian masalah dengan menganalisa data yang ada pada *database*, dimana data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer. Data Mining merupakan ekstraksi informasi yang tersembunyi dari *database* yang besar. Data Mining adalah teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan agar fokus pada informasi yang paling penting dalam gudang data mereka [2].

Data Mining merupakan proses menemukan kolerasi baru yang bermanfaat, pola dan trend dengan manambang sejumlah repositori data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika. Data Mining semakin menyebar dan berkembang dengan pesat belakangan ini karena kemampuannya dalam menambang pola bermanfaat dan trend dari basis data yang sudah ada. Perusahaan-perusahaan telah menghabiskan dana milyaran untuk mengumpulkan data dalam jumlah *megabytes* atau *terabytes* tapi tidak mendapatkan keuntungan yang bernilai didalamnya, padahal didalamnya terbatat informasi yang berharga namun tersembunyi pada repositori data [6]. Data mining terbagi menjadi beberapa kelompok antara lain sebagai berikut :

a. Deskripsi (*Description*)

Terkadang penelitian analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

b. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

c. Prediksi (*Prediction*)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwadalam prediksi nilai dari hasil akan ada di masa datang.

d. Klasifikasi (*Classification*)

Klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

e. Pengclusteran (*Clustering*)

Pengclusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan record-record dalam *cluster* lain. Pengclusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam Pengclusteran. Pengclusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma Pengclusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogen*), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

f. Asosiasi (*Assosiation*)

Asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Data mining telah dipakai diberbagai bidang seperti ilmu sains, bisnis dan industri, teknik, kesehatan, serta pertahanan dan keamanan. Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan, menimbulkan kebutuhan untuk dapat memanfaatkan

informasi dan pengetahuan yang terkandung didalam limpahan data tersebut, yang kemudian melahirkan data mining. data mining merupakan proses untuk menemukan pengetahuan (*knowledge discovery*) yang ditambang dari sekumpulan data yang volumenya sangat besar [2]. Adapun kelebihan dan kekurangan data mining antara lain :

1. Kelebihan data mining
  - a. Mampu mengolah data dalam jumlah yang besar
  - b. Mudah diimplementasi dan dijalankan
  - c. Pencarian data secara otomatis
2. Kekurangan data mining
  - a. Hasil pengelompokan selalu berubah-ubah
  - b. Jumlah cluster harus ditentukan
  - c. Memungkinkan suatu gerobol tidak mempunyai anggota.

Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut [7]. Data mining adalah metoda yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada *database*, ini adalah teknologi yang sangat potensial bagi perusahaan yang sangat potensial bagi perusahaan dalam memberdayakan data *warehouse* [8]. Secara garis besar, data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu:

1. *Deskriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah *clustering*, *asosiation*, dan *sequential mining*.
2. *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi.

Data mining adalah bagian dari proses KDD (*Knowledge Discoveriyin Database*) yang terdiri dari beberapa tahap seperti pemilihan data, prapengolahan, transpormasi, data mining dan evaluasi hasil [4]. Data mining mempunyai fungsi yang sangat penting untuk membantu endapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Data mining mempunyai fungsi dasar [2] antara lain :

1. prediksi (*prediction*)
2. deskripsi (*descriptio*)
3. klasifikasi (*clasification*)
4. asosiasi (*associatioan*)

## **2.2 Konsep Dasar *Clustering***

*Cluster* juga dikenal dengan istilah *clustering* adalah metode yang digunakan untuk membagi rangking data menjadi beberapa grup berdasarkan kesamaan yang telah ditentukan sebelumnya. *Clustering* adalah proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelompok-kelompoknya berdasarkan tingkat kesamaan. Sehingga dalam suatu *cluster* merupakan data dengan karakterisitik yang hampir sama [2].

*Clustering* adalah proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelompok-kelompoknya berdasarkan tingkat kesamaan. Sehingga dalam suatu *cluster* merupakan data yang karakteristik yang hampir sama [4]. Analisis *clustering* adalah sebuah teknik dari analisis *multivariable* yang digunakan untuk mengelompokkan obyek-obyek (*variable* atau data) sehingga dapat menghasilkan suatu informasi untuk membantu pelaksanaan pengujian terhadap obyek dan pada akhirnya dapat menyajikan suatu hipotesis berdasarkan relasi yang terjadi [2]. Jadi secara umum dapat dikatakan bahwa :

1. Data dalam suatu kluster memiliki tingkat kesamaan yang tinggi.
2. Data dalam kluster yang berbeda memiliki tingkat kesamaan yang rendah.

Teknik yang digunakan untuk mengukur tingkat kesamaan antara lain :

1. *Minowski Distance*
2. *Tanimoto Measure*
3. *Pearson's r Measure*
4. *Mahalanobis Measure*

Tujuan dari teknik ini adalah untuk melakukan pengelompokan berdasarkan kriteria tertentu sehingga obyek-obyek tersebut mempunyai variasi di dalam *cluster* relatif kecil dibandingkan variasi antar *cluster* [4].

### **2.3 Konsep Algoritma *K-Means***

*K-means clustering* pertama kali dikemukakan oleh *Dunn* (1973) dan dikembangkan oleh *Bezdek* (1981) yang banyak digunakan dalam *pattern recognition*. Untuk menghitungnya dibuat terlebih dahulu jumlah kelas yang akan dijadikan basis klasifikasi. Kemudian dilakukan iterasi sampai mendapat

keanggotaan kelompok tersebut. Metode ini juga memberi hasil yang *smooth* (halus) karena pembobotan yang digunakan berdasarkan himpunan *fuzzy* [2].

*K-Means* adalah suatu teknik pengklusteran data yang mana keberadaan tiap data dalam suatu *cluster* ditentukan oleh nilai keanggotaannya. Pada kondisi awal, pusat *cluster* masih belum akurat. Tiap data memiliki derajat keanggotaan untuk tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap data secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi obyektif [3].

*K-Means* adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numeric. Algoritma *K-Means* termasuk *partitioning clustering* yang memisahkan data ke *k* daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma *K-Means*, setiap data harus termasuk ke *cluster* tertentu dan bisa dimungkinkan bagi setiap data yang termasuk *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* lainnya [8].

Kehalusan disini berarti objek pengamatan tidak mutlak untuk menjadi anggota satu kelompok saja, tapi juga mungkin menjadi anggota kelompok yang lain dengan ukuran tingkat keanggotaan yang berbeda. Objek akan cenderung menjadi anggota kelompok tertentu dimana tingkat keanggotaan objek dalam kelompok itu paling besar dibanding dengan kelompok yang lain [2].

*K-Means* merupakan salah satu metode *clustering non-hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Algoritma ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain. Algoritma ini sederhana untuk diterapkan dan dijalankan, relatif cepat, mudah diadaptasi dan umum digunakan dalam praktek [4]. Metode *K-Means* memiliki karakteristik sebagai berikut :

1. *K-Means* merupakan metode pengelompokan yang sederhana dan dapat digunakan dengan mudah
2. Pada jenis data set tertentu, *K-Means* tidak dapat melakukan segmentasi data dengan baik dan dimana hasil segmentasi tidak dapat menentukan pola kelompok yang mewakili karakteristik bentuk alami data
3. *K-Means* bisa mengalami masalah ketika mengelompokkan data yang mengandung *outlier*.

Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidak miripan data pada kelompok lain. Secara umum metode *K-means* menggunakan algoritma dasar adalah sebagai berikut [1]:

1. Tentukan nilai  $k$  sebagai jumlah *cluster* yang ingin dibentuk. Penentuan banyak jumlah cluster  $k$  dilakukan dengan beberapa faktor seperti pertimbangan teoritis dan konsep tual yang diusulkan untuk menentukan beberapa banyak *cluster*.

2. Bangkitkan k centroid (titik pusat *cluster*) awal secara random. Untuk menentukan centroid awal dilakukan secara acak dari beberapa objek yang tersedia sebanyak k *cluster*.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance*.

$$De = \sqrt{(M_{ix} - C_{ix})^2 + (M_{iy} - C_{iy})^2}$$

Keterangan:

De= jarak antara data dengan pusat *cluster*

$M_{ix}$ =nilai variabel x

$C_{ix}$ =nilai *cluster* variabel x

$M_{iy}$ =nilai variabel y

$C_{iy}$ =nilai *cluster* untuk variabel y

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya*.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru ( $C_k$ ) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Dimana :

$n_k$  adalah jumlah dokumen dalam *cluster* k

$d_i$  adalah dokumen dalam *cluster* k.

6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Proses penggabungan titik dilakukan dengan membandingkan matrik kumpulan tugas-tugas pada iterasi sebelumnya dengan matrik kumpulan tugas pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma *k-means* clustering analysis sudah konvergen, tetapi jika berbeda maka belum konvergen sehingga perlu dilakukan iterasi berikutnya [1].

Salah satu solusi yang digunakan untuk memperkecil besaran angka antar variabel adalah dengan melakukan normalisasi angka-angka yang ada di variabel tinggi badan, berat badan dan usia menggunakan rumus:

$$\text{Nilai Normalisasi} = \frac{(\text{Nilai Awal} - \text{Nilai Minimal})}{(\text{Nilai Maksimal} - \text{Nilai Minimal})}$$

Nilai variabel tinggi badan, berat badan dan usia akan dinormalisasi ke dalam rentang 0 – 1. Normalisasi angka pada tiap variabel ini sangat dibutuhkan sebelum proses perhitungan nilai centroid oleh algoritma *K-Means* agar tidak ada parameter yang mendominasi dalam perhitungan jarak antar data.

## **2.4 Alat Bantu Perancangan Program**

### **2.4.1 Konsep Dasar MYSQL dan Database**

Menurut *Gordon C. Everest*, *database* adalah koleksi atau kumpulan data yang mekanis, terbagi (*shared*), terdefinisi secara formal dan dikontrol terpusat pada organisasi. *Database* adalah koleksi “data operasional” yang tersimpan dan dipakai oleh sistem aplikasi dari suatu organisasi [9].

*Database* memiliki sifat internal dan terbagi. Internal yang dimaksud adalah kesatuan (*integritas*) dari file-file yang terlibat. Sedangkan yang dimaksud terbagi

(*shared*) adalah elemen-elemen *database* dapat dibagikan pada para user baik secara sendiri-sendiri maupun secara serentak dan pada waktu yang sama (*Concurrent Sharing*) [9].

MySQL adalah sebuah *database* relasional gratis namun memiliki fitur lengkap. MySQL dikembangkan pada tahun 1990 untuk memenuhi kebutuhan basis data yang terus tumbuh untuk komputer mengelola informasi secara cerdas [9].

#### **2.4.2 Konsep Dasar HTML**

HTML (*Hypertext Mark up Language*) adalah semacam bahasa pengkodean bukan sebagai bahasa pemrograman dan juga sekumpulan perintah yang terformat yang digunakan untuk membuat halaman dokumen web *Hypertext* berarti halaman yang dibuat link (rangkai) [9].

HTML digunakan untuk membangun suatu halaman web. Sekalipun banyak orang menyebutnya sebagai suatu bahasa pemrograman. HTML sebenarnya sama sekali bukan bahasa pemrograman, karena seperti tercermin dari namanya, HTML adalah suatu bahasa *mark up*. HTML digunakan untuk mark up (penandaan) terhadap sebuah dokumen teks. Tanda tersebut digunakan untuk menentukan format atau style dari teks yang ditandai [9].

#### **2.4.3 Konsep Dasar *Hypertext Preprocessor* (PHP)**

*Hypertext Preprocessor* adalah bahasa pemrograman yang berjalan disisi server, biasanya digunakan untuk membuat aplikasi web dengan menggunakan web server, seperti *Apache*. PHP juga dapat digunakan untuk membuat skrip baris perintah mirip dengan *Perl* atau *shell script*, namun penggunaannya kurang

umum, karena itu PHP lebih banyak digunakan sebagai bahasa pemrograman web [9]

#### **2.4.4 Konsep Dasar XAMPP**

XAMPP adalah perangkat lunak open source yang mendukung banyak sistem operasi. Fungsi XAMPP adalah sebagai server yang berdiri sendiri (localhost) yang terdiri atas program *Apache* HTTP Server, MySQL, dan penerjemah atau *interpreter* bahasa yang ditulis dengan bahasa pemrograman PHP dan Perl [9].

XAMPP adalah perangkat lunak bebas, yang mendukung banyak sistem operasi, merupakan kompilasi dari beberapa program. Fungsinya adalah sebagai server yang berdiri sendiri (localhost), yang terdiri atas program *Apache* HTTP Server, MySQL database, dan penerjemah bahasa yang ditulis dengan bahasa pemrograman PHP dan Perl. Nama XAMPP merupakan singkatan dari X (empat sistem operasi papun), *apache*, MySQL, PHP dan Perl. Program ini tersedia dalam *GNU General Public License* dan bebas, merupakan web server yang mudah digunakan yang dapat melayani tampilan halaman web yang dinamis. Untuk mendapatkannya dapat mendownload langsung dari web resminya [10].

#### **2.4.5 Konsep Dasar Web**

*Web* adalah halaman informasi yang disediakan melalui jalur internal sehingga bisa diakses di seluruh dunia selama terkoneksi dengan jaringan internet. *Webseit* merupakan komponen atau kumpulan komponen yang terdiri dari teks, gambar, suara animasi sehingga lebih merupakan media informasi yang menarik untuk dikunjungi. Layanan *Web* atau *World Wide Web (WWW)* sangat banyak dimanfaatkan dalam internet [10].

#### 2.4.6 Konsep Dasar Notepad ++

Notepad++ adalah sebuah aplikasi text editor yang bersifat gratis. Notepad menitik beratkan kegunaan aplikasi untuk editing text dalam waktu yang cepat dan praktis. Notepad++ mendukung banyak format bahasa pemrograman seperti PHP, HTML, Java Script dan CSS [1].

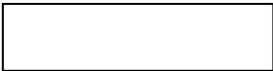
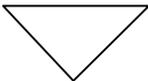
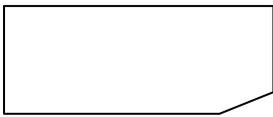
#### 2.5 Alat Bantu Perancangan Aplikasi

Pengenalan perangkat keras dan perangkat lunak sangat diperlukan, sehingga perancangan (pedesaianan) sistem akan lebih bagus . Dalam hal ini dapat diketahui alat-alat analisis dan perancangan system antara lain sebagai berikut :

##### 2.5.1 Aliran Sistem Informasi (ASI)

Bagian aliran sistem informasi merupakan bagan yang menunjukkan arus pekerjaan secara keseluruhan dari suatu sistem. Aliran sistem informasi sangat berguna untuk mengetahui permasalahan yang ada pada suatu sistem [12].

**Tabel 2.1 Simbol-simbol aliran sistem informasi**

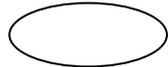
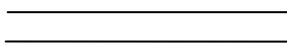
Simbol	Keterangan
	Menunjukkan kegiatan proses dari operasi program komputer
	Untuk menggambarkan penyimpanan data baik dalam bentuk arsip atau file komputer. Bisa ditulis F atau A
	Dokumen merupakan formulir yang digunakan untuk merekam data yang menunjukkan input dan output, baik untuk proses manual maupun komputer
	Menunjukkan kegiatan manual

	Menunjukkan dokumen input dan output baik untuk proses manual, maupun komputer
---	--

### 2.5.2 Data Flow Diagram (DFD)

Data Flow Diagram (DFD) merupakan model dari sistem untuk menggambarkan pembagian sistem ke modul yang lebih kecil. Data Flow Diagram berguna untuk memudahkan pemakai yang kurang menguasai bidang komputer untuk mengerti sistem yang akan dikerjakan. Data Flow Diagram juga merupakan alat perancangan sistem yang berorientasi pada alur data dengan konsep dekomposisi dapat digunakan untuk penggambaran analisa maupun rancangan sistem yang mudah dikomunikasikan oleh profesional sistem kepada pemakai maupun pembuat program [13].

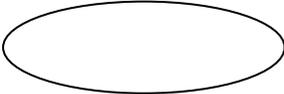
**Tabel 2.2 Simbol Data Flow Diagram**

Simbol	Keterangan
	Entitas : simbol suatu diluar lingkungan sistem yang akan menghasilkan input dan output
	Proses : simbol yang digunakan untuk melakukan pemrosesan data baik use maupun komputer.
	Data Store : simbol yang digunakan untuk mewakili suatu penyimpanan data
	Arus Data : simbol yang digunakan untuk menggambarkan arus data di dala sistem.

### 2.5.3 *Entity Relathionship Diagram (ERD)*

ERD merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. ERD untuk memodelkan struktur data dan hubungan antar data [13].

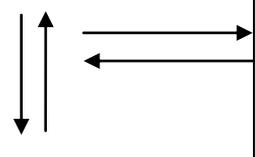
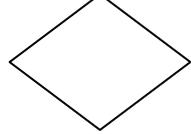
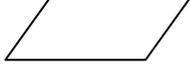
**Tabel 2.3 Simbol-simbol *Entity Relathionship Diagram***

Simbol	Keterangan
	Entitas suatu objek yang dapat diidentifikasi dalam suatu pemakai nyata.
	Atribut merupakan properti yang dimiliki setiap yang akan disimpan datanya. Entitas mempunyai elemen yang disebut atribut.
	Relasi menunjukkan adanya hubungan antar entitas
	Penghubung

### 2.5.4 *Flowchart*

Flowchart adalah penggambaran secara grafik dari langkah-langkah dan urutan dari suatu program. Flowchart menolong analisa dan programmer untuk memecahkan masalah kedalam segmen yang pengoperasian. Flowchart biasanya mempermudah penyelesaian suatu masalah khususnya masalah yang perlu di pelajari dan di eveluasi lebih lanjut [13].

**Tabel 2.4 Simbol-simbol *Flowchart***

Simbol	Nama	Deskripsi
	<i>Flow</i> <i>Direction</i>	Digunakan untuk menghubungkan antar simbol ( <i>connection</i> ).
	<i>Terminator</i>	Untuk memulai ( <i>start</i> ) atau akhir ( <i>and</i> ) dari sesuatu kegiatan.
	Proses	Simbol yang digunakan untuk pemrosesan suatu kegiatan.
	<i>Processing</i>	Pemilihan proses berdasarkan kondisi yang ada.
	<i>Input-output</i>	Simbol yang menyatakan <i>input</i> dan <i>output</i> data.
	Dokumen	Simbol yang menyatakan input dan output yang berasal dari dokumen / hardfile berupa lembaran.
	Database	Simbol yang menyatakan database sistem.

### 2.5.5 UML (*Unified Model Language*)

UML adalah salah satu bahasa visual untuk mempresentasikan dan mengkomunikasikan sistem melalui penggunaan diagram dan teks pendukung. Guna fungsi pemodelan visual ini, UML menggunakan empat (4) jenis diagram standar, yaitu:

### **1. Use Case**

*Use Case* digunakan pada saat pelaksanaan tahap *requirement* dalam pengembangan suatu sistem informasi. *Use Case* menggambarkan hubungan antara entitas yang biasa disebut aktor dengan suatu proses yang dapat dilakukannya.

### **2. Statistic Diagram / Class Diagram**

*Static Diagram* digunakan untuk menggambarkan stuktur kelas dan obyek yang akan digunakan dalam sistem yang akan dibangun. *Static Diagram* digunakan pada tahap analisa dan desain aplikasi.

### **3. Sequence Diagram**

*Sequence Diagram* digunakan untuk menjelaskan aliran pesan dari suatu *Class* ke *Class* lain secara *sequensial* (berurutan). *Sequence Diagram* digunakan pada tahap desain aplikasi.

### **4. Activity Diagram**

*Activity Diagram* digunakan untuk menjelaskan tanggung jawab elemen. *Activity Diagram* biasa dikolaburasikan dengan *Sequence Diagram* dalam pendiskripsian *visual* dari tahap desain aplikasi.

## **2.6 Konsep Dasar Gizi**

Gizi adalah ilmu yang mempelajari segala tentang makanan dan hubungan dengan kesehatan optimal. Gizi adalah suatu organisme menggunakan makanan yang dikonsumsi secara normal melalui proses digesti, absorpsi, transportasi, penyimpanan, metabolisme dan pengeluaran zat yang digunakan

untuk mempertahankan kehidupan, pertumbuhan dan fungsi normal dan organ serta menghasilkan energi [13]. Adapun fungsi gizi bagi tubuh adalah:

1. Memberi Energi
2. Pertumbuhan dan Pemeliharaan Jaringan Tubuh
3. Mengatur Proses Tubuh.

### **2.6.1 Konsep Dasar Status Gizi**

Status gizi adalah suatu ukuran mengenai kondisi tubuh seseorang yang dapat dilihat dari makanan yang dikonsumsi dan penggunaan zat-zat gizi di dalam tubuh. Perkembangan berat badan dalam keadaan normal, akan searah dengan pertumbuhan tinggi badan dengan kecepatan tertentu [13].

Status gizi adalah suatu ukuran mengenai kondisi tubuh seseorang yang dapat dilihat dari makanan yang dikonsumsi dan pengguna zat-zat gizi dalam tubuh. Status gizi secara umum dapat dibagi menjadi 3 kategori yaitu: gizi lebih, gizi baik dan gizi kurang [13].

Gizi normal merupakan suatu ukuran status gizi dimana terhadap keseimbangan antar jumlah energi yang masuk kedalam tubuh dan energi yang dikeluarkan dari luar tubuh sesuai dengan kebutuhan individu. Energi yang masuk kedalam tubuh dapat berupa karbohidrat, protein, lemak dan gizi-gizi lainnya [13].

Gizi kurang lebih sering disebut *imdernutrition* merupakan keadaan gizi seorang dimana jumlah energi yang masuk lebih sedikit dari energi yang dikeluarkan. Hal ini dapat terjadi karena jumlah energi yang masuk lebih sedikit dari anjuran kebutuhan hidup [13].

Gizi lebih atau *overnutrition* merupakan keadaan gizi seorang dimana jumlah energi yang masuk kedalam tubuh lebih sedikit dari jumlah energi yang dikeluarkan. Kelebihan energi yang dikonsumsi disimpan di dalam jaringan bentuk lemak. Gizi lebih dapat menyebabkan kegemukan atau obesitas [13].

## **2.7 Konsep Dasar Balita**

Balita menjadi golongan masyarakat yang paling rawan terhadap gangguan gizi. Masa balita yang dimulai dari 0 – 60 bulan merupakan periode emas pertumbuhan, khususnya saat 24 bulan pertama. Hal ini disebabkan oleh adanya perubahan dan cepatnya pertumbuhan balita pada usia 0 – 24 bulan, sehingga dibutuhkan nutrisi makanan yang relatif lebih banyak mengandung gizi dengan kualitas yang lebih tinggi. Bahkan pertumbuhan saat dewasa pun sangat bergantung pada bagaimana kondisi gizi yang diterimanya saat balita. Maka dari itu, kebutuhan gizi saat balita haruslah diperhatikan dengan baik [13].

## **2.8 Konsep Dasar Puskesmas**

Puskesmas merupakan suatu sarana pelayanan kesehatan masyarakat yang sangat penting yang bertanggung jawab menyelenggara pembangunan kesehatan di suatu wilayah. Pusat pembinaan peran serta masyarakat dalam bidang kesehatan serta pusat pelayanan kesehatan tingkat pertama yang menyelenggarakan kegiatan secara menyeluruh, yang berkesinambungan pada suatu masyarakat yang bertempat tinggal di suatu wilayah tersebut.

## BAB 3

### METODOLOGI PENELITIAN

#### 3.1 Pendahuluan

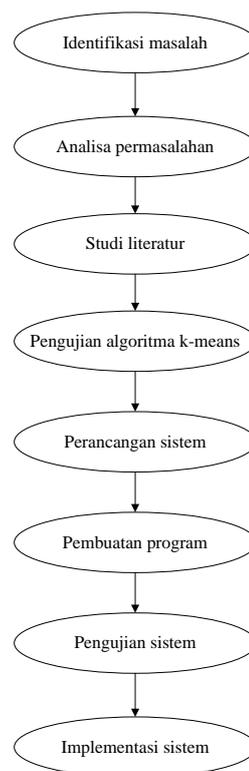
Metode Penelitian digunakan dengan cara mendatangi dan mewawancarai sejumlah karyawan dalam suatu populasi perusahaan, memberikan kuesioner atau angket melalui *fax* dan *e-mail* untuk menghemat waktu dan biaya [5]. Metode penelitian yang digunakan adalah metode analisa dengan pendekatan terstruktur (*Structured approach*) yang lengkap dengan alat (*tools*) dan teknik yang dibutuhkan dalam sistem sehingga hasil analisis dari sistem yang dikembangkan menghasilkan sistem yang strukturnya dapat didefinisikan dengan baik dan jelas .

Pada tahap ini juga digunakan notasi-notasi yang berlaku dalam prancangan sistem *data flow diagram* untuk menggambarkan arus data sistem sehingga dapat membantu dalam proses komunikasi dengan pemakai. *Data flow diagram* digunakan untuk menggambarkan sistem baru yang akan dikembangkan secara logika tanpa terlebih dahulu mempertimbangkan lingkungan fisik dimana data tersebut diproses.

#### 3.2 Kerangka Kerja Penelitian

Pada bab ini akan diuraikan metodologi penelitian dan kerangka kerja penelitian. Kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Metode pengembangan sistem yang digunakan mengadopsi dari metode *waterfall*. Metode *waterfall* merupakan salah satu metode dalam SDLC (*System Developmen Life Cycle*) yang mempunyai

ciri khas. Metode *waterfall* merupakan model klasik yang bersifat sistematis, berurutan dalam membangun sebuah *software*. Nama model ini sebenarnya adalah "*Linear Sequential Model*" . Model ini termasuk ke dalam *model generic* pada rekayasa perangkat lunak dan pertama kali model *waterfall* ini diperkenalkan oleh *Winston Royce* pada tahun 1970. Meskipun sering dianggap kuno, tetapi model *waterfall* ini merupakan model yang paling banyak dipakai dalam *Software Engineering (SE)*. Metode *Waterfall* ini melakukan pendekatan secara sistematis dan berurutan. Metode ini disebut dengan metode *Waterfall* karena tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya dan berjalan berurutan [10]. Adapun kerangka kerja penelitian ini dapat digambarkan pada gambar 3.1



**Gambar 3.1 Kerangka Kerja Penelitian**

Berdasarkan kerangka kerja pada gambar 3.1 maka masing-masing langkahnya dapat diuraikan seperti berikut ini :

### **3.2.1 Identifikasi Masalah**

Identifikasi masalah merupakan langkah pertama untuk melakukan analisis sistem. Masalah (*problem*) dapat didefinisikan sebagai suatu pertanyaan yang diinginkan untuk dipecahkan (*solusi*). Masalah inilah yang menyebabkan sasaran dari sistem tidak dapat dicapai. Pada tahap ini dilakukan wawancara dan observasi untuk memperoleh data yang dibutuhkan pada saat pengajuan tugas akhir ini. Wawancara dilakukan dengan ketua kader posyandu edelwies, sedangkan observasi dilakukan dengan datang langsung ke posyandu edelweis.

### **3.2.2 Analisa Permasalahan**

Pada tahap ini dilakukan proses analisa data yang difokuskan untuk pembuatan perangkat lunak. Untuk memahami sifat dasar dari perangkat lunak yang akan dibangun, seorang analisa sistem harus memahami sistem informasi, kinerja sistem dan tampilan menu (*interface*) yang diperlukan. Maka pada penelitian ini memerlukan data sebagai berikut :

#### **3.2.2.1 Analisa kebutuhan masukan**

*Input* atau masukan dari aplikasi datamining penentuan status gizi balita ini adalah banyak cluster dan banyak variabel yang berguna untuk mengelompokkan status gizi balita :

### 1. Variabel

Pada tahap ini, variabel adalah cara untuk menentukan nilai bobot dari *cluster* tinggi badan, berat badan dan usia.

### 2. Cluster

Pada tahap ini cluster adalah cara untuk mengelompokkan status gizi balita, yang mana dalam penelitian ini memiliki 3 cluster antara lain gizi baik, gizi kurang dan obesitas.

#### **3.2.2.2 Analisa kebutuhan proses**

Analisa Kebutuhan proses dalam aplikasi data mining penentuan status gizi balita antara lain :

1. Menentukan nilai  $k$  sebagai jumlah *cluster* yang ingin dibentuk.
2. Menentukan titik pusat cluster secara acak
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance*.
4. Mengelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya.
5. Melakukan iterasi, kemudian tentukan posisi *centroid* baru ( $C_k$ ) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

#### **3.2.2.3 Analisa kebutuhan keluaran**

Data keluaran yang dihasilkan oleh aplikasi data mining penentuan status gizi balita ini adalah pengelompokan jumlah balita yang termasuk gizi normal, gizi tidak normal dan obesitas.

#### **3.2.2.4 Analisa kebutuhan antarmuka**

Perancangan antarmuka merupakan program php merupakan pilihan yang tepat untuk mengimplementasikan aplikasi data mining penentuan status gizi balita menggunakan algoritma *k-means clustering*.

#### **3.2.2.5 Analisa kebutuhan perangkat lunak**

Perangkat keras komputer tidak berarti tanpa perangkat lunak begitu juga sebaliknya. Jadi perangkat lunak dan perangkat keras saling mendukung satu sama lain. Perangkat keras hanya berfungsi jika diberikan intruksi-intruksi kepada perangkat itu, intruksi-intruksi inilah disebut dengan perangkat lunak. Kebutuhan perangkat lunak minimal pada penelitian ini adalah :

1. Sistem operasi windows 7
2. *Microsoft word*
3. Bahasa pemogramam PHP
4. MySQL

#### **3.2.2.6 Analisa kebutuhan perangkat keras**

Perangkat keras yang digunakan untuk menggambarkan pengolahan data dari instansi terkait dalam perhitungan variabel dan *cluster*. Yang ditemukan dalam data mining dengan menggunakan algoritma *k-means clustering*. Kebutuhan perangkat keras minimal pada penelitian ini adalah:

1. Prosesor : Intel(R) celeron(R) CPU N3350 @ 1. 10GHz (2 CPUs)
2. Memory : 2048MB RAM
3. Page File : 3450MB used, 634MB availabe

### **3.2.3 Studi Literatur**

Setelah dapat mengatasi masalah yang ada. Setelah masalah di analisa, maka dipelajari literatur yang berhubungan dengan permasalahan. Kemudian literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian ini. Sumber literatur didapatkan dari jurnal, artikel, yang membahas tentang datamining penentuan status gizi balita menggunakan algoritma *k-means clustering*.

### **3.2.4 Pengumpulan Data**

Pada tahap pengumpulan data dalam penelitian ini dilakukan dengan beberapa cara:

#### **a. Pengumpulan Data Primer**

Penelitian ini dilakukan untuk mendapatkan data primer dengan cara mengambil sampel dari beberapa data dari variabel dan cluster untuk menentukan status gizi balita di Puskesmas Rambah Hilir 1. Tujuannya adalah mendapatkan data langsung dari objek atau sampel pengumpulan data sekunder.

#### **b. Pengumpulan Data Sekunder**

Penulis mengumpulkan data dan informasi melalui studi pustaka yang bersifat sekunder yaitu data yang diperoleh melalui buku referensi tentang data mining menggunakan algoritma *k-means clustering*.

### **3.2.5 Pengujian Algoritma K-means Clustering**

Metode atau teknik perancangan yang digunakan dalam perancangan data mining penentuan potensi status gizi balita menggunakan algoritma *k-means clustering*.

### **3.2.6 Perancangan Sistem**

Proses ini digunakan untuk mentransformasikan kebutuhan detail menjadi kebutuhan yang lengkap, dokumen desain fokus pada bagaimana dapat memenuhi fungsi yang dibutuhkan seperti berikut ini :

#### *a. Proses pemodelan Sistem*

Proses ini menggambarkan bagaimana perangkat lunak beroperasi dan mengilustrasikan aktifitas yang dilakukan. Cara yang digunakan adalah dengan menggunakan *Unified Model Language (UML)*.

#### *b. Pemodelan Data*

Pada tahap ini penulis akan merancang dan menggambarkan data yang digunakan dan diciptakan dalam suatu perangkat lunak yaitu dengan menggunakan *Class Diagram*.

#### *c. Desain Antar Muka*

Menggambarkan bagaimana pengguna memasukkan data dengan melakukan pemilihan menu, maupun mendapatkan *input*, proses dan *output* perangkat lunak.

### **3.2.7 Pembuatan Program**

Pada tahap pembuatan program ini dilakukan untuk membuat program sistem yang diperoleh perancangan program dari data yang ada. Tahap-tahap yang dilakukan untuk penelitian guna perancangan dan pembuatan program tersebut secara terstruktur. Adapun *software* yang dipakai untuk pembuatan kode adalah *Notepad++*, *PHP*, *Database* dan *MySQL*.

### **3.2.8 Pengujian Sistem**

Tahap ini adalah tahap pengujian dan tahap pendukung yang artinya sistem yang telah dibuat dari hasil analisis masalah yang telah melalui tahap-tahap desain, pengkodean barulah masuk kedalam pengujian sistem, sehingga akan dapat diketahui fungsional semua bagian yang sudah diuji, dan pengujian ini dilakukan dengan menggunakan metode *Black Box*, hal ini dilakukan untuk meminimalisir terjadinya kesalahan dan yang dihasilkan sesuai dengan yang diinginkan.

### **3.2.9 Implementasi Sistem**

Tahap ini dilakukan dimana analisis dan prancangan basisdata digunakan. Artinya dalam perjalanannya, akan ada banyak perubahan-perubahan dan kebutuhan-kebutuhan baru sesuai dengan perkembangan yang ada pada Puskesmas Rambah Hilir 1.