

## BAB II

### LANDASAN TEORI

#### 2.1 Defenisi Sistem

Menurut (Lilis Puspita, 2011) yang dikutip oleh (Febryanta, 2014) sistem adalah suatu kesatuan usaha yang terdiri dari bagian-bagian yang berkaitan satu sama lain yang berusaha mencapai suatu tujuan dalam lingkungan kompleks.

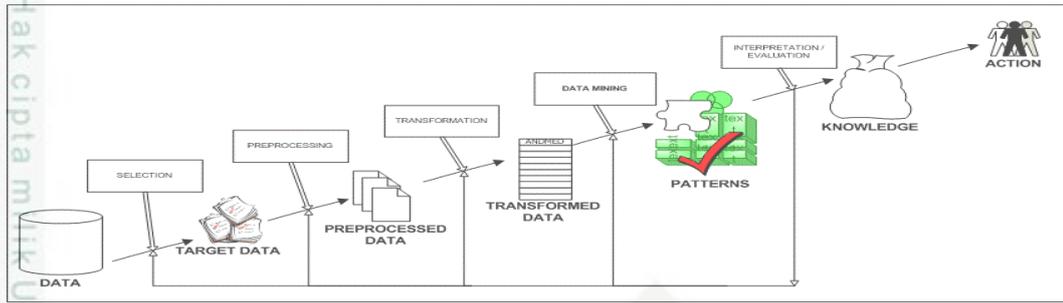
#### 2.2 Data Mining

Menurut (Burhan, Muktamar; 2013) *Data mining* adalah kegiatan yang 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 (Muktamar, 2013).

Sedangkan menurut (Sumanthi, Sivandham, 2009), *data mining* juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam *database* yang sering disebut dengan istilah *Knowledge Discovery in Database (KDD)*.

Teknik data mining didukung oleh tiga teknologi yaitu pengumpulan data secara besar, multiprocessor pada komputer dan algoritma data mining. Tugas dari data mining yaitu Deskriptif – menemukan gambaran pola yang menarik dari data dan Prediktif - memprediksi perilaku dari model berdasarkan data yang ada (Johanes, 2014).

Menurut (Hadi Suryono dkk, 2013) data mining adalah suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan basis pengetahuan. Tahap-tahap ini diilustrasikan di Gambar 2.1:



Gambar 2.1 Tahap-Tahap Proses KDD (Hadi Suryono dkk, 2013)

### 1. Data Cleaning

*Data cleaning* merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

### 2. Integrasi Data

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

### 3. Seleksi data

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

### 4. Transformasi data

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan klastering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

### 5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

### 6. Evaluasi

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.3 Klasifikasi

Klasifikasi merupakan suatu teknik data mining yang melihat sifat dari atribut dari kelompok data yang telah didefinisikan. Teknik ini dapat digunakan untuk memberi pengetahuan pada data baru dengan memanipulasi data yang ada yang telah diklasifikasikan dan dengan menggunakan hasilnya untuk memberikan pengetahuan atau sejumlah aturan. Aturan tersebut digunakan data baru untuk dapat diklasifikasikan terhadap suatu kategori atau kelas tertentu.

## 2.4 Naive Bayes Classifier

Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan *Bayes* dengan asumsi independensi yang kuat. Dengan kata lain, *Naive Bayes* merupakan model fitur independen. Maksud independensi yang kuat pada fitur dalam *Bayes* adalah sebuah fitur data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Yuda Septian Nugroho, 2017).

Prediksi *Bayes* didasarkan pada teorema *Bayes* dengan formula umum sebagai berikut (Yuda, 2017):

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \dots \dots \dots (2.1)$$

Keterangan :

- $P(H|E)$  : Probabilitas akhir bersyarat suatu hipotesis H terjadi jika bukti E terjadi.
- $P(E|H)$  : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.
- $P(H)$  : Probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun.
- $P(E)$  : Probabilitas awal bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan bukti (E) yang diamati. Ada

beberapa hal penting dari aturan *Bayes* tersebut, yaitu ( Yuda Septian Nugroho, 2017) :

1. Sebuah probabilitas awal H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah bukti diamati

## 2.5 Naive Bayes Untuk Klasifikasi Dan Prediksi

Kaitan antara *Naive Bayes* dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi adalah bahwa hipotesis dalam teorema *Bayes* merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukannya. Jika X adalah vektor masukkan yang berisi fitur dan Y adalah label kelas, maka dalam *Naive Bayes* dituliskan dengan notasi P(Y|X). Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y, sedangkan P(Y) disebut probabilitas awal (*prior probability*) Y (Yuda Septian Nugroho, 2017).

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir P(Y|X) pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dengan memaksimalkan nilai P(X'|Y') yang didapat ( Yuda Septian Nugroho, 2017).

Formula *Naive Bayes* untuk klasifikasi adalah :

$$P(Y|X) = \frac{P(Y) \times \prod_{i=1}^q P(X_i|Y)}{P(X)} \dots \dots \dots (2.2)$$

P(Y|X) adalah probabilitas data dengan vektor X pada kelas Y. P(Y) adalah probabilitas awal kelas Y.  $\prod_{i=1}^q P(X_i|Y)$  adalah probabilitas independen kelas Y dari semua fitur dalam vektor X. Nilai P(X) selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Probabilitas independen  $\prod_{i=1}^q P(X_i|Y)$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

merupakan pengaruh dari semua fitur data terhadap setiap kelas Y, yang dinotasikan sebagai berikut ( Yuda Septian Nugroho, 2017) :

$$P(X|Y = y) = \prod_{i=1}^q P(X_i|Y = y) \dots\dots\dots (2.3)$$

Setiap set fitur  $X = \{X_1, X_2, X_3, \dots, X_q\}$  terdiri atas q atribut.

*Naive Bayes* lebih mudah untuk menghitung fitur bertipe kategoris, namun untuk fitur bertipe numerik (kontinu) ada perlakuan khusus, yaitu:

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur tersebut dengan nilai interval diskret.
2. Menggunakan distribusi Gaussian untuk merepresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas  $P(X_i|Y)$ . Distribusi Gaussian dikarakteristikan dengan dua parameter yaitu *mean* ( $\mu$ ) dan *varian* ( $\sigma^2$ ). Untuk setiap kelas  $y_j$ , probabilitas bersyarat kelas  $y_j$  untuk fitur  $X_i$  adalah :

$$P(X_i = x_i|Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \dots\dots\dots (2.4)$$

Parameter  $\mu_{ij}$  bisa didapat dari *mean* sampel  $X_i$  dari semua data latih yang menjadi milik kelas  $y_j$ . Sedangkan  $\sigma_{ij}^2$  didapat dari *varian* sampel data latih.

Berdasarkan penjelasan sebelumnya maka dapat didapatkan formula-formula utama dalam proses klasifikasi dengan *Naive Bayes Classifier* yaitu:

1. Probabilitas atribut pada setiap kelas :

$$P(X = x|Y = y) = \frac{\text{jumlah kemunculan data dengan atribut } x \text{ pada kelas } y}{\text{jumlah data dengan kelas } y} \dots\dots\dots (2.5)$$

2. Probabilitas kelas :

$$P(Y) = \frac{\text{jumlah kemunculan data dengan kelas } y}{\text{jumlah keseluruhan data}} \dots\dots\dots (2.6)$$

3. Probabilitas akhir :

$$P(Y|X) = \text{Akumulasi probabilitas atribut pada setiap kelas} * \text{Probabilitas kelas} \dots\dots (2.7)$$

Proses perhitungan dalam penentuan kelas akhir pada *Naive Bayes Classifier* dimulai dengan mengetahui tipe dari fitur-fitur pada setiap kelas. Jika bertipe kategoris lakukan perhitungan probabilitasnya, namun jika bertipe

numerik maka gunakan distribusi *Gaussian* untuk mencari nilai probabilitasnya. Selanjutnya menghitung probabilitas masing-masing kelas. Probabilitas akhir didapatkan dengan mengalikan probabilitas setiap fitur dan probabilitas kelasnya. Label kelas dipilih berdasarkan nilai probabilitas akhir yang paling tinggi.

## 2.6 Normalisasi Data

Menurut (Elvianti, 2015) normalisasi pada penelitian ini digunakan untuk mempersempit range data latih. Normalisasi yang digunakan pada penelitian ini adalah min-max normalization yang merupakan proses transformasi nilai dari data yang dikumpulkan pada range value antara 0.0 dan 1.0, dimana nilai terkecil (min) adalah 0.0 dan nilai tertinggi (max) adalah 1.0, seperti yang ditunjukkan Persamaan 2.8

$$v^i = \frac{v - \min_a}{\max_a - \min_a} (new\_max_a - new\_min_a) + new\_min_a \dots\dots\dots(2.8)$$

Dimana :

$v^i$  : Data baru setelah normalisasi

$v$  : Data sebelum normalisasi

$new\_max_a$  : Batas nilai max baru adalah 1

$new\_min_a$  : Batas nilai min baru adalah 0

$max_a$  : Nilai maximum pada kolom

$min_a$  : Nilai minimum pada kolom

## 2.7 K-Nearest Neighbor (KNN)

k-Nearest Neighbor merupakan metode yang biasa digunakan pada klasifikasi data. Algoritma ini digunakan untuk mengklasifikasikan terhadap objek berdasarkan data pembelajaran yang jarak tetangganya paling dekat atau memiliki nilai selisih yang kecil dengan objek tersebut.

KNN merupakan suatu metode yang menggunakan algoritma supervised dengan hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini ialah mengklasifikasikan objek baru berdasarkan *attribut* dan *training sample*. Prinsip umum dari algoritma ini

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber;

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

adalah menemukan k data training untuk menentukan k-nearest neighbor berdasarkan ukuran jarak. Selanjutnya mayoritas dari k tetangga terdekat akan menjadi dasar untuk memutuskan kategori dari sample berikutnya. Selain itu algoritma ini sendiri sering digunakan untuk klasifikasi pada teknik data mining meskipun dapat digunakan untuk estimasi dan prediksi data.

Rumus yang biasa digunakan sebagai ukuran jarak untuk data numerik ini antara lain:

### 2.7.1 Manhattan Distance

Menurut (Bramer, 2007) yang dikutip (Henny, 2013) Manhattan Distance adalah formula untuk menghitung jarak antara dua titik. Perhitungan Manhattan Distance untuk mencari jarak minimal dari dua buah titik (x1,y1) dan (x2,y2) dapat dilakukan dengan menghitung  $|x2-x1|+|y2-y1|$  (Henny, 2013).

Manhattan Distance merupakan salah satu pengukuran yang paling banyak digunakan meliputi penggantian perbedaan kuadrat dengan menjumlahkan perbedaan absolute dari variabel-variabel. Fungsi ini hanya akan menjumlahkan selisih nilai x dan y dari dua buah titik, rumus *Manhattan*, seperti yang ditunjukkan pada Persamaan 2.9:

$$D(x, y) = \|x_2 - x_1\| = \sum_{i=1}^n |x_i - y_i| \dots \dots \dots (2.9)$$

- d : jarak antara titik pada data *training* x dan titik data *testing* y yang akan diklasifikasi, dimana  $x=x_1,x_2,\dots,x_i$  dan  $y=y_1,y_2,\dots,y_i$
- i : merepresentasikan nilai atribut
- n : merupakan dimensi atribut.

### 2.8 Modified k-Nearest Neighbor (MKNN)

Ide utama dari metode ini adalah hal pertama yang dilakukan adalah perhitungan validitas untuk semua data yang terdapat pada data latih. Selanjutnya, dilakukan perhitungan Weight Voting pada semua data uji menggunakan validitas data. (Hamid Parvin,dkk 2010).

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 2.8.1 Validitas Data Training

Validitas digunakan untuk menghitung jumlah titik dengan label yang sama untuk semua data pada data latih. Validitas setiap data tergantung pada setiap tetangga terdekatnya. Setelah dilakukan validasi data, selanjutnya data tersebut digunakan sebagai informasi lebih mengenai data tersebut (Parvin, 2008). Persamaan yang digunakan untuk menghitung validitas setiap data latih adalah (Parvin, 2008):

$$Validitas(x) = \frac{1}{K} \sum_{i=1}^K S(lbl(x), lbl(N_i(x))) \dots\dots\dots(2.10)$$

Dimana :

*K* : Jumlah titik terdekat

*Lbl(x)* : Kelas *x*

*N<sub>i</sub>(x)* : Label kelas titik terdekat *x*

Fungsi *S* digunakan untuk menghitung kesamaan antara titik *a* dan data ke-*b* tetangga terdekat. Persamaan untuk mendefinisikan fungsi *S* terdapat dalam Persamaan 3.4 di bawah ini (Parvin, 2008) :

$$S(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \dots\dots\dots(2.11)$$

Dimana :

*a* : Kelas *a* pada data training

*b* : Kelas lain selain *a* pada data training

### 2.8.2 Weight Voting

Dalam metode *MKNN*, pertama *weight* masing-masing tetangga dihitung dengan menggunakan  $1 / (d_e + 1)$ . Kemudian, validitas dari setiap data pada data latih dikalikan dengan *weight* berdasarkan pada jarak Manhattan. Sehingga metode *MKNN*, didapatkan persamaan *weight voting* tiap tetangga sebagai berikut (Parvin, 2008) :

$$W(x) = Validasi(x) \times \frac{1}{d_e + 0,5} \dots\dots\dots(2.12)$$

Dimana :

$W(i)$  : Perhitungan Weight Voting

$Validasi(x)$  : Nilai Validasi

$d_e$  : Jarak Manhattan

### 2.8.3 Akurasi Sistem

Menurut (Han dkk., 2012) yang dikutip (Elvianti, 2015) Performa dari suatu model kasifikasi dapat diukur dengan tingkat akurasinya berdasarkan *Confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. TP dan TN memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberikan informasi ketika *classifier* salah (Elvianti, 2015). Gambar 2.2 adalah contoh dari *confusion matrix*.

		Actual Class	
		Ya	Tidak
Predictive Class	Ya	TP	FN
	Tidak	FP	TN
Total		P'	N'

**Gambar 2.2** *Confusion Matrix* (Elvianti, 2015)

Akurasi merupakan persentase dari data yang diprediksi secara benar.

Perhitungan akurasi adalah :

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots \dots \dots (2.13)$$

Keterangan :

TP : *True positives*, merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.

TN : *True negatives*, merupakan jumlah data dengan kelas negatif yang diklasifikasikan negatif.

FP : *False positives*, merupakan jumlah data dengan kelas positif diklasifikasikan negatif.

FN : *False negatives*, merupakan jumlah data dengan kelas negatif diklasifikasikan positif.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.9 Status Gizi balita

Status gizi adalah keadaan tubuh yang merupakan hasil akhir dari keseimbangan antara zat gizi yang masuk ke dalam tubuh dan utilisasinya.

### 2.9.1 Defenisi Gizi

Gizi adalah suatu proses organisme menggunakan makanan yang dikonsumsi secara normal melalui proses pencernaan, absorpsi, transportasi, penyimpanan, metabolisme, dan pengeluaran zat-zat yang tidak digunakan untuk mempertahankan kehidupan, pertumbuhan dan fungsi normal dari organ-organ serta menghasilkan energi.

### 2.9.2 Defenisi Status Gizi

Status gizi adalah keadaan yang diakibatkan oleh status keseimbangan antara jumlah asupan (*intake*) zat gizi dan jumlah yang dibutuhkan oleh tubuh untuk berbagai fungsi biologis: (pertumbuhan fisik, perkembangan aktivitas, pemeliharaan kesehatan, dan lainnya).

Menurut (WHO, 2005) gizi balita ada empat kelas yaitu sangat kurus, kurus, normal dan obesitas. Balita berumur 6 bulan dengan berat badan 8 kg dapat dikategorikan bergizi normal, apabila berat badan mencapai 10 kg lebih dapat dikategorikan obesitas dan apabila berat badan balita 7 kg kebawah dapat dikategorikan kurus.

## 2.10 Penilaian Status Gizi Balita

Penilaian status gizi dapat digunakan untuk memberikan penilaian status gizi terhadap perorangan yang berfungsi untuk keperluan rujukan dari masyarakat atau puskesmas, juga dapat pula digunakan orang tua sebagai media untuk pemantauan pertumbuhan anak yang sangat dibutuhkan bagi perkembangan anak. Banyak cara untuk menilai status gizi salah satunya adalah dengan cara pengukuran tubuh manusia yang dikenal dengan istilah "*Anthropometri*".

## 2.10.1 Klasifikasi Status Gizi Balita

Dalam menentukan status gizi balita memiliki ukuran bakunya. Ukuran baku yang sekarang di gunakan di Indonesia adalah standar baku *World Health Organization-National Center for Health Statistics* (WHO-NCHS).

Menurut (WHO, 2005) yang dikutip dari (Kevin, dkk, 2016) penilaian status gizi berdasarkan Indeks BB/U (Berat Badan menurut Umur), TB/U (Tinggi Badan menurut Umur), BB/TB (Berat Badan menurut Tinggi Badan) dengan standar baku *Anthropometri* WHO-NCHS dapat digolongkan menjadi:

**Tabel 2.1 Penilaian Status Gizi (Kevin, dkk, 2016)**

No.	Indeks yang dipakai	Batas Pengelompokan	Sebutan Status Gizi
1	BB/U	<-3 SD -3 s/d <-2 SD -2 s/d +2 SD > +2 SD	Gizi Buruk Gizi Kurang Gizi Baik Gizi Lebih
2	TB/U	<-3 SD -3 s/d <-2 SD -2 s/d +2 SD > +2 SD	Sangat Pendek Pendek Normal Tinggi
3	BB/TB	<-3 SD -3 s/d <-2 SD -2 s/d +2 SD > +2 SD	Sangat Kurus Kurus Normal Gemuk

Dimana SD adalah Skor Simpangan Baku (Standar Deviation = Z)

Cara menghitung status gizi dengan menggunakan *Z-score* dapat ditentukan dengan menggunakan rumus:

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan})}{\text{Nilai Simbang Baku Rujukan}} \dots\dots\dots(2.14)$$

Dimana terdapat dua kategori dalam menghitung status gizi balita menggunakan *Z-score*, yaitu:

Bila “Nilai Riel Preorangan” hasil pengukuran  $\geq$  “Nilai Median Acuan” BB/U, TB/U, BB/TB, maka rumusnya

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan})}{SD \text{ Upper}}$$

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan})}{+1 SD - \text{Median}}$$

Bila “Nilai Riel Preorangan” hasil pengukuran  $\leq$  “Nilai Median Acuan”  
 BB/U, TB/U, BB/TB, maka rumusnya

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan})}{SD \text{ Lower}}$$

$$Zscore = \frac{(\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan})}{\text{Median} - -1 SD}$$

Keterangan: Nilai Riel itu berat badan sebenarnya (aktual).

Nilai Median itu diambil dari nilai tabel Baku Rujukan WHO-NCHS

Nilai (-1SD) itu juga dapat dilihat pada tabel WHO-NCHS

- Jika nilai riel lebih kecil dari pada nilai median berarti yang digunakan sebagai pembagi adalah nilai -1SD
- Jika nilai riel lebih besar dari pada nilai median berarti yang digunakan sebagai pembagi adalah nilai +1SD

Menurut (Arisman, 2008) yang dikutip dari (Febreality, 2011) interpretasi

status gizi balita berdasarkan tiga indeks *Anthropometri* (BB/U, TB/U, BB/TB) adalah sebagai berikut:

**Tabel 2.2 Interpretasi Status Gizi (Febreality, 2011)**

No	Indeks yang digunakan			Interpretasi
	BB/U	TB/U	BB/TB	
1	Rendah	Rendah	Normal	Normal, dulu kurang gizi
	Rendah	Tinggi	Rendah	Sekarang kurang + +
	Rendah	Normal	Rendah	Sekarang kurang +
2	Normal	Normal	Normal	Normal
	Normal	Tinggi	Rendah	Sekarang kurang
	Normal	Rendah	Tinggi	Sekarang lebih, dulu kurang
3	Tinggi	Tinggi	Normal	Tinggi, normal
	Tinggi	Rendah	Tinggi	Obesitas
	Tinggi	Normal	Tinggi	Sekarang lebih, belum

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Indeks yang digunakan			Interpretasi
	BB/U	TB/U	BB/TB	
				obesitas

## 2.10.2 Anthropometri

Ada beberapa macam *Anthropometri* yang telah digunakan yaitu Umur (U), Berat Badan (BB), Tinggi Badan (TB), Lingkar Perut (LP), Lingkar Lengan Atas (LLA), Lingkar Dada (LD), dan Lapisan Lemak Bawah Kulit (LLBK)

Jenis *Anthropometri* yang digunakan untuk menentukan status gizi balita adalah (Proverawati, 2009):

### 1. Umur (U)

Umur memegang peranan dalam penentuan status gizi, kesalahan penentuan akan menyebabkan interpretasi status gizi balita yang salah.

### 2. Berat Badan (BB)

Berat badan adalah salah satu ukuran yang memberikan gambaran massa jaringan, termasuk cairan tubuh. Berat badan sangat peka terhadap perubahan yang mendadak baik karena penyakit infeksi maupun konsumsi makanan yang menurun.

### 3. Tinggi Badan (TB)

Tinggi badan dapat memberikan gambaran fungsi dari pertumbuhan yang dapat dilihat dari keadaan kurus kering dan kecil pendek. Tinggi badan sangat baik untuk melihat akan keadaan gizi pada masa lalu

### 4. Lingkar Kepala (LK)

Lingkar kepala sering digunakan sebagai ukuran status gizi untuk menggambarkan perkembangan otak. Otak tumbuh pada masa balita yang dimulai pada fase janin usia 30 minggu sampai bayi usia 18 bulan. Lingkar kepala pada waktu lahir rata-rata 34-35 cm. Ukuran lingkar kepala akan bertambah 2 cm perbulan pada usia 0 – 3 bulan, selanjutnya usia 4 – 6 bulan akan bertambah 1 cm perbulan, dan pada usia 6 – 12 bulan akan bertambah 0,5cm perbulan. Standar baku ukuran lingkar kepala belum ada. Ukuran otak

dan lapisan tulang kepala dan tengkorak dapat bervariasi sesuai dengan keadaan gizi.

Keterangan untuk ketiga indeks *Anthropometri* (BB/U, TB/U, BB/TB):

Rendah: <-2 SD Standar baku *Anthropometri* WHO-NCHS

Normal: -2 s/d +2 SD Standar baku *Anthropometri* WHO-NCHS

Tinggi : >+2SD Standar baku *Anthropometri* WHO-NCHS

Contoh kasus perhitungan status gizi balita dengan menggunakan *Z-score*:

Sebutkan status gizi seorang balita, jika balita tersebut laki-laki dengan umur 9 bulan, berat badan 12 kg dan tinggi badan 70 cm

Jawab:

a. Untuk Indeks BB/U

$$Zscore = \frac{(Nilai Riel Perorangan - Nilai Median Acuan)}{+1 SD - Median}$$

$$Zscore = \frac{(12 - 8,9)}{9,9 - 8,9} = \frac{(3,1)}{1} = +3,1 SD$$

Status Gizi Berdasarkan BB/U adalah Gizi Lebih

b. Untuk Indeks TB/U

$$Zscore = \frac{(Nilai Riel Perorangan - Nilai Median Acuan)}{Median - -1 SD}$$

$$Zscore = \frac{(70 - 72)}{72 - 69,7} = \frac{(-2)}{2,3} = -0,86 SD$$

Status Gizi Berdasarkan TB/U adalah Normal

c. Untuk Indeks BB/TB

$$Zscore = \frac{(Nilai Riel Perorangan - Nilai Median Acuan)}{+1 SD - Median}$$

$$Zscore = \frac{(12 - 8,6)}{9,3 - 8,6} = \frac{(3,4)}{0,7} = +4,85 SD$$

Status Gizi Berdasarkan BB/TB adalah Gemuk

Dari ketiga indeks status gizi diatas (BB/U, TB/U, BB/TB), maka status gizi tersebut dapat diinterpretasikan berdasarkan tabel 2.2, yaitu:

BB/U = +3,1 SD = Tinggi

TB/U = -0,86 SD = Normal

BB/TB = +4,85 SD = Tinggi

Berdasarkan tabel interpretasi status gizi, Tinggi, Normal, Tinggi, maka status gizinya adalah gizi lebih tapi belum obesitas.

## 2.11 Penelitian Terkait

Berikut adalah beberapa penelitian sebelumnya yang menjadi acuan dan dasar penelitian ini

**Tabel 2.3 Penelitian Terkait**

No	Tahun	Peneliti	Judul	Akurasi	Perbedaannya
1	2015	Mega Kartika Sari, Ernawati, dan Pranowo	Kombinasi metode <i>k-Nearest Neighbor</i> dan <i>Naive Bayes</i>	Tingkat akurasi adalah 79,92%	Pada penelitian ini menggunakan metode <i>k-Nearest Neighbor</i> dan <i>Naive Bayes Classifie</i> , sedangkan pada penelitian penulis menggunakan metode <i>Naive Bayes</i> dan <i>Modified k-Nearest Neighbor</i> metode pengembangan dari <i>k-Nearest Neighbor</i> untuk kasus status gizi balita.
2	2016	Kevin Martha Rasepta	Klasifikasi status gizi balita menggunakan metode <i>Modified K-Nearest Neighbor</i>	dengan tingkat akurasi sebesar 82,057%	Pada penelitian sebelumnya menggunakan metode <i>Modified K-Nearest Neighbor</i> , sedangkan pada penelitian penulis menggunakan kombinasi metode <i>Naive Bayes dan Classifier Modified k-nearest neighbor</i>



- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, pen-  
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin

No	Tahun	Peneliti	Judul	Akurasi	Perbedaannya
3	2014	Tri Halomoan Simanjuntak, Wayan Firdaus Mahmudy, dan Sutrisno	Implementasi <i>Modified k-nearest neighbor</i> dengan Otomatisasi Nilai K pada Pengklasifikasian Penyakit Tanaman Kedelai	Tingkat akurasi adalah 98,83%	Objek penelitian ini adalah Pengklasifikasian Penyakit Tanaman Kedelai, sedangkan objek penelitian penulis adalah pengklasifikasian gizi balita. Sedangkan pada metodenya menggunakan metode <i>Modified K-Nearest Neighbor</i> dan <i>Naive Bayes Classifier</i> ,
4	2016	Salina monalisa rambe	Penerapan Metode Naive Bayes Classifier untuk Prediksi Perkara Pertanahan <i>Naive Bayes</i>	Tingkat akurasi adalah 83,213%	Pada penelitian ini menggunakan metode <i>Naive Bayes Classifier</i> , sedangkan pada penelitian penulis menggunakan metode <i>Naive Bayes Classifier</i> dan <i>Modified k-nearest neighbor</i>
5	2014	Johanes Widagdho Yodha, dan Achmad Wahid Kurniawan	Pengenalan motif batik menggunakan Deteksi Tepi Canny dan k-Nearest Neighbor	Tingkat akurasi maksimalnya adalah 100% dan minimalnya adalah 56,67%	Pada penelitian ini disebutkan penggunaan Manhattan Distance memiliki tingkat akurasi yang lebih baik dari pada Euclidean Distance

No	Tahun	Peneliti	Judul	Akurasi	Perbedaannya
6	2015	Rika Anggraini	Perbandingan penerapan, dan <i>Euclidean Distance</i> , <i>Manhattan Distancce</i> , dan <i>Adaptive Distance Measure</i> dalam k-Neighbor pada identifikasi penyakit kulit	Tingkat akurasi maksimal 98,15% dan minimalnya 97,22%	Pada penelitian ini disebutkan penggunaan Manhattan Distance memiliki tingkat akurasi yang lebih baik dari pada Euclidean Distance

