

BAB II

KAJIAN PUSTAKA

Dalam bab ini akan membahas tentang kajian pustaka yang digunakan di dalam penelitian ini, seperti tentang sistem, klasifikasi, *naïve bayes*, status gizi balita, *cross validation* dan *confusion matrix*.

2.1 Status Gizi

2.1.1 Definisi Status Gizi

Masa balita merupakan tahap pertumbuhan yang sangat pesat, dimana sangat membutuhkan kasih sayang dan perhatian dari orang tua. Selain itu balita juga sangat membutuhkan zat gizi yang seimbang agar tumbuh kembang anak tersebut tidak terhambat karena anak balita rentang dalam menderita kekurangan gizi (Santoso & Ranti, 2004) .

Status gizi adalah kondisi tubuh yang berakibat dari konsumsi makanan dan penggunaan zat-zat gizi yang dibedakan kedalam 4 kategori diantaranya status gizi buruk, status gizi kurang, status gizi baik, status gizi lebih (Supariasa, Bakri, & Fajar, 2002) .

2.1.2 Penilaian Status Gizi Balita

Di Indonesia cara yang paling umum dan sering digunakan untuk penilaian status gizi adalah dengan menggunakan *antropometri*, karena lebih praktis dan mudah dilakukan. Untuk mengetahui status gizi seseorang dapat digunakan berbagai cara, salah satunya yaitu dengan cara *Z-Skor* terhadap nilai *median*. *Z-Skor* merupakan index *antropometri* yang di gunakan secara internasional untuk penentuan status gizi dan pertumbuhan, yang diekspresikan sebagai satuan standar deviasi (SD) populasi. *Z-Skor* digunakan untuk menghitung status gizi secara *antropometri* pada berat badan terhadap umur (BB/U), tinggi badan terhadap umur (TB/U), dan berat badan terhadap tinggi badan (BB/TB) (Fitri, Setyawati, & Rahadi S, 2014) .

Menurut Supariasa, Bakri, & Fajar (2002) beberapa indeks antropometri yang sering digunakan yaitu Berat Badan menurut Umur (BB/U), Tinggi Badan menurut Umur (TB/U) dan Tinggi Badan menurut Berat Badan (TB/BB). Berikut merupakan pengertian, kelebihan dan kekurangan dari indeks *antropometri* yang sering digunakan yaitu menurut (Supariasa et al., 2002):

Tabel 2.1. Pengertian, Kelebihan dan Kekurangan dari Indeks Antropometri (Supariasa et al., 2002)

Jenis	Pengertian	Kelebihan	Kelemahan
BB/U	Berat badan adalah salah satu parameter yang memberikan gambaran massa tubuh. Berat badan adalah parameter antropometri yang sangat labil.	<ul style="list-style-type: none"> 1. Lebih mudah dan cepat dimengerti oleh masyarakat umum 2. Baik untuk mengukur status gizi akut atau kronis 3. Berat badan dapat berfluktuasi 4. Sangat sensitive terhadap perubahan-perubahan kecil 5. Dapat mendeteksi kegemukan 	<ul style="list-style-type: none"> 1. Dapat mengakibatkan interpretasi status gizi yang keliru bila terdapat edema maupun asites 2. Di daerah pedesaan yang masih terpencil dan tradisional, umur sering sulit ditaksir secara tepat karena pencatatan umur yang belum baik 3. Memerlukan data umur yang akurat, terutama untuk anak dibawah usia lima tahun 4. Sering terjadi kesalahan dalam pengukuran, seperti pengaruh pakaian atau gerakan anak pada saat penimbangan

Jenis	Pengertian	Kelebihan	Kelemahan
			5. Secara operasional sering mengalami hambatan karena masalah sosial budaya setempat. Dalam hal ini orang tua tidak mau menimba anaknya, karena dianggap barang dagangan dan sebagainya
TB/U	Tinggi badan merupakan antropometri yang menggambarkan keadaan pertumbuhan skeletal	1. Baik untuk menilai status gizi masa lampau 2. Ukuran panjang dapat dibuat sendiri, murah dan mudah dibawa	1. Tinggi badan tidak cepat naik, bahkan tidak mungkin turun 2. Pengukuran relative sulit dilakukan karena anak harus berdiri tegak, sehingga diperlukan 2 orang untuk melakukannya 3. Ketepatan umur sulit didapat
BB/TB	Indeks BB/TB merupakan indikator yang baik untuk menilai status	1. Tidak memerlukan data umur 2. Dapat membedakan proporsi badan	1. Tidak dapat memberikan gambaran, apakah anak tersebut pendek, cukup

Jenis	Pengertian	Kelebihan	Kelemahan
	gizi saat kini (sekarang).	(gemuk, normal dan kurus)	<p>tinggi badan dan kelebihan tinggi badan menurut umurnya, karena faktor umur tidak dipertimbangkan</p> <p>2. Dalam praktek sering mengala mi kesulitan dalam melakukan pengukuran panjang/tinggi badan pada kelompok balita</p> <p>3. Membutuhkan dua macam alat ukur</p> <p>4. Pengukuran relative lebih lama</p> <p>5. Membutuhkan dua orang untuk melakukannya</p> <p>6. Sering terjadi kesalahan dalam pembacaan hasil pengukuran, terutama bila dilakukan oleh kelompok non-profesional</p>

Penimbangan balita sebagai faktor dominan terhadap kejadian balita gagal tumbuh setelah dikontrol pendidikan ibu. Balita yang tidak ditimbang secara teratur berisiko 1,5 kali mengalami gagal tumbuh dibandingkan yang ditimbang secara teratur (Rahmadini, Sudiarti, & Utari, 2013).

Di Indonesia yang menjadi standar baku ukuran untuk menentukan status gizi balita yaitu dengan menggunakan baku *World Health Organization National Center for Health Statistics* (WHO-NCHS). WHO-NCHS dalam penilaian status gizi balita berdasarkan Indeks BB/U , TB/U , BB/TB. Dimana penilaian status gizi balita dibedakan antara anak laki-laki dan anak perempuan.

Beberapa indeks *antropometri* yang sering digunakan berdasarkan Keputusan Menteri Kesehatan Republik Indonesia (Nomor:1995/MENKES/SK/XII/2010) yang dapat diunduh pada situs (<http://gizi.depkes.go.id/>) tentang Standar *Antropometri* Penilaian Status Gizi Anak.

Terdapat pengklasifikasian atau penilaian status gizi balita dengan menggunakan standar *antropometri* berdasarkan Keputusan Menteri Kesehatan Republik Indonesia (Nomor : 1995/MENKES/SK/XII/ 2010) yaitu :

Tabel 2.2. Standar Baku Antropometri

WHO-NCHS (Departemen Kesehatan, 2011)

Indeks	Kategori Status Gizi	Ambang Batas (Z-Score)
Berat Badan menurut Umur (BB/U) Anak Umur 0-60 Bulan	Gizi Buruk	< -3 SD
	Gizi Kurang	-3 SD sampai dengan <-2 SD
	Gizi Baik	-2 SD sampai dengan 2 SD
	Gizi Lebih	>2 SD
Panjang Badan menurut Umur (PB/U) atau Tinggi Badan menurut Umur (TB/U) Anak Umur 0-60 Bulan	Sangat Pendek	< -3 SD
	Pendek	-3 SD sampai dengan <-2 SD
	Normal	-2 SD sampai dengan 2 SD
	Tinggi	>2 SD
Berat Badan menurut Panjang Badan (BB/PB) atau Berat Badan menurut Tinggi Badan (BB/TB) Anak Umur 0-60 Bulan	Sangat kurus	< -3 SD
	Kurus	-3 SD sampai dengan <-2 SD
	Normal	-2 SD sampai dengan 2 SD
	Gemuk	>2 SD
	Sangat kurus	< -3 SD
	Kurus	-3 SD sampai dengan <-2 SD

Indeks	Kategori Status Gizi	Ambang Batas (Z-Score)
Indeks Massa Tubuh menurut Umur (IMT/U) Anak Umur 0-60 Bulan	Normal Gemuk	-2 SD sampai dengan 2 SD >2 SD

Dimana SD merupakan Skor Simpangan Baku (*Standar Deviation Score = Z*). Untuk menghitung status gizi yaitu dengan menggunakan rumus *Z-score*. Rumus perhitungan Z-Score:

$$\text{Rumus perhitungan Z-Score:} \\ \frac{\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan}}{\text{BB/U, TB/U, BB/TB}} \quad (\text{Supariasa et al., 2002})$$

Menurut (Febrealti, 2011) dari rumus diatas terdapat 2 kategori dalam menghitung status gizi balita yaitu:

- Apabila “Nilai *Riel Perorangan*” hasil pengukuran \geq “Nilai Median Acuan” BB/U, TB/U, BB/TB, maka rumusnya :

$$\frac{\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan}}{\text{BB/U, TB/U, BB/TB}} \quad (1)$$

Atau

$$\frac{\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan}}{\text{BB/U, TB/U, BB/TB}} \quad (2)$$

- Apabila “Nilai *Riel Perorangan*” hasil pengukuran \leq “Nilai Median Acuan” BB/U, TB/U, BB/TB, maka rumusnya :

$$\frac{\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan}}{\text{BB/U, TB/U, BB/TB}} \quad (3)$$

Atau

$$\frac{\text{Nilai Riel Perorangan} - \text{Nilai Median Acuan}}{\text{BB/U, TB/U, BB/TB}} \quad (4)$$

Contoh perhitungan status gizi pada balita:

Seorang bayi berjenis kelamin laki-laki berumur 3 bulan dengan berat badan 6 kg dan tinggi badan atau panjang badan 63 cm. Bagaimana status gizi BB/U, PB/U dan BB/PB bayi tersebut ?

Jawab:

- Indeks BB/U

$$- = \frac{6 - 6.4}{6.4 - 5.7} = \frac{0.4}{0.7} = ,$$

Jadi, Status gizi berdasarkan BB/U yaitu Gizi Baik karena berada pada batas pengelompokan -2 s/d +2SD

b. Indeks TB/U

$$- = \frac{63 - 61.4}{63.5 - 61.4} = \frac{1.4}{2.1} = .$$

Jadi, Status gizi berdasarkan TB/U yaitu Pendek karena berada pada batas pengelompokan -2 s/d +2SD

c. Indeks BB/TB

$$- = -(-1) =$$

$$- = \frac{6 - 6.8}{6.8 - 6.2} = \frac{0.8}{0.6} = 1.3$$

Jadi, status gizi berdasarkan BB/TB yaitu Normal karena berada pada batas pengelompokan -2 s/d +2SD

Menurut (Direktorat Bina Gizi, 2011) penting untuk mengkoreksi hasil pengukuran apabila terdapat balita yang diukur tidak sesuai dengan kelompok umur:

1. Jika seorang anak berumur kurang dari 2 tahun diukur tingginya (berdiri) maka ditambah 0.7 cm untuk mengkonversi menjadi panjang badan.
2. Jika seorang anak berumur 2 tahun atau lebih dan diukur panjangnya (terlentang) maka dikurangi 0.7 cm untuk mengkorvensi menjadi tinggi badan.

2.2 Sistem

Sistem merupakan suatu kumpulan dari beberapa sub sistem yang saling berhubungan satu dengan lainnya atau bekerja sama untuk mencapai tujuan

bersama dengan menerima input dan output dengan melalui proses yang teratur (Miranti, 2016) .

Menurut (Setiawan, 2009) Klasifikasi sistem dari berbagai sudut pandang diklasifikasi ke dalam beberapa bagian, diantaranya:

1. Sistem Abstrak dan Sistem Fisik

Sistem abstrak yaitu sistem yang berupa ide yang secara fisik tidak tampak, sedangkan sistem fisik yaitu sistem yang ada tampak secara fisik

2. Sistem Alamiah dan Sistem Buatan Manusia

Sistem alamiah adalah sistem yang secara alami di proses oleh alam, sedangkan sistem buatan adalah sistem yang dirancang oleh manusia.

3. Sistem Tertentu dan Sistem Tak Tentu

Sistem tertentu adalah sistem yang kondisi masa depannya dapat diprediksi, sedangkan sistem tak tentu adalah sistem yang kondisi masa depannya tidak dapat diprediksi karena ada unsur probabilitas.

4. Sistem Tertutup dan Sistem Terbuka

Sistem tertutup adalah sistem yang tidak ada hubungannya atau tidak dipengaruhi oleh lingkungan luar, sedangkan sistem terbuka adalah yang ada hubungannya atau dipengaruhi oleh lingkungan luar.

2.3 Klasifikasi

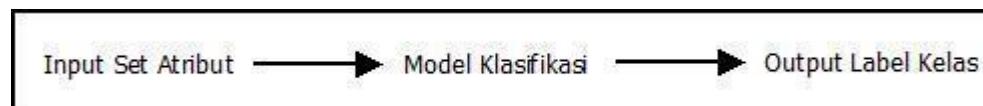
Klasifikasi adalah suatu *fungsionalitas data mining* yang akan menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek di dalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian (Amri, 2013).

Pada klasifikasi biasanya diberikan sekumpulan kategori atau label dan beberapa data yang dijadikan sebagai data latih yang akan digunakan untuk membangun sebuah model. Kemudian model tersebut akan diuji dengan data set yang akan ditebak katagori atau label pada setiap datanya (Rahmawati, 2016).

Menurut (Darmawan, 2012) menyatakan bahwa pendekatan umum yang digunakan dalam masalah klasifikasi adalah pertama, *training* data berisi *record* yang mempunyai label *class* yang telah diketahui. *Training* data digunakan untuk membangun model klasifikasi yang kemudian diaplikasikan

ke *testing* data yang berisi *record-record* dengan label *class* yang tidak diketahui.

Definisi lain memaparkan bahwa klasifikasi merupakan metode pembelajaran yang dijadikan acuan untuk mengekstak model untuk menggambarkan atau memprediksi kelas dimasa depan (Soni, Ansari, Sharma, & Soni, 2011).



Gambar 2.1. Model Klasifikasi

Terdapat model beberapa model klasifikasi menurut (Gorunescu, 2011) diantaranya:

- a. *Decision/classification trees*
- b. *Bayesian classifiers/Naive Bayes classifiers*
- c. *Neural networks*
- d. *Statistical analysis*
- e. *Genetic algorithms*
- f. *Rough sets*
- g. *K-nearest neighbor classifier*
- h. *Rule-based methods*
- i. *Memory based reasoning*
- j. *Support vector machines.*

2.4 Naïve Bayes

Algoritma *Naïve Bayes* merupakan algoritma yang terdapat pada klasifikasi dengan metode probabilitas dan statistic yang dikemukakan oleh ilmuan Inggris *Thomas Bayes*, yaitu memprediksi peluang masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*, teorema tersebut dikombinasikan dengan *Naïve* dimana diasumsikan kondisi antar atribut saling bebas (Bustami, 2013).

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Dimitoglou, Adams, & Jim, 2012).

Definisi lain mengatakan bahwa *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Rahmawati, 2016).

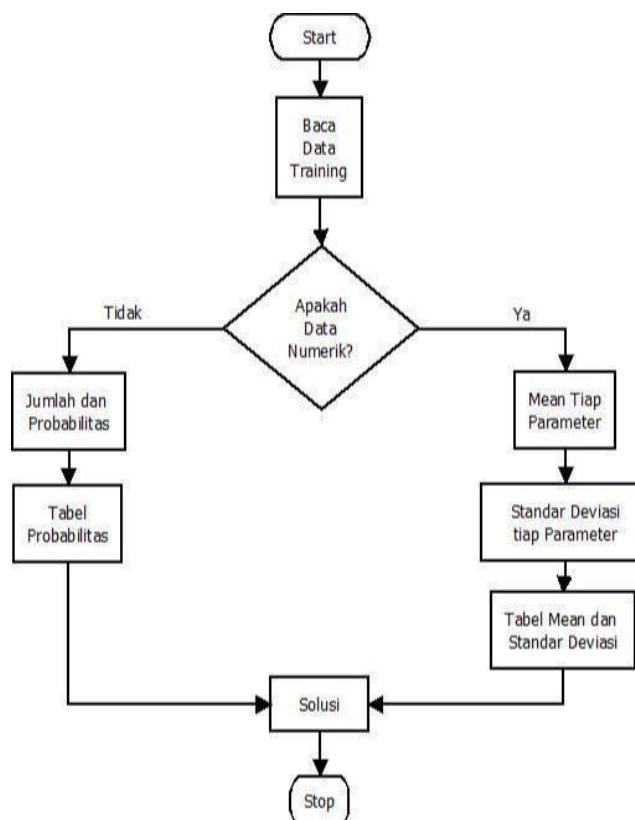
Persamaan Teorema Bayes :

$$(1) = \frac{(n)}{o} \quad (5)$$

Keterangan :

- $(|)$: Data dengan kelas yang belum diketahui
- (n) : Hipotesis data merupakan suatu kelas spesifik
- $: Probabilitas hipotesis A berdasarkan kondisi B$
- $: Probabilitas hipotesis A dengan syarat kondisi B$
- $: Probabilitas B$

Adapun alur dari metode *Naïve Bayes* sebagai berikut:



Gambar

2.2 Alur Metode atau Flowchart *Naïve Bayes*

Dari gambar diatas, menjelaskan bagaimana cara menghitung dengan menggunakan *Naïve Bayes*:

1. Baca *data training*
2. Hitung jumlah dan probabilitas, tetapi jika data tersebut merupakan data numerik maka :
 - a. Jika *data training* bersifat data numerik atau data kontinu maka harus mencari nilai rata-rata (*mean*) dari setiap parameter. Kemudian setelah mencari nilai *mean* selanjutnya mencari nilai Standar Deviasi dari setiap parameter tersebut. Setelah hasil Standar Deviasi didapatkan maka nilai Standar Deviasi yang paling tinggi yang akan dijadikan sebagai solusi.

Rumus mencari nilai Mean:

$$\bar{x} = \frac{\sum x_i}{n} \quad (6)$$

Dimana:

- = rata-rata hitung (*mean*)
- = nilai sample ke-i
- = jumlah sample

Rumus mencari nilai Standar Deviasi:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (7)$$

Dimana:

- = standar deviasi
- = nilai sample ke-i
- = rata-rata hitung (*mean*)
- = jumlah sample

Jika *data training* bukan data numerik maka dilakukan perhitungan jumlah dan probabilitas dari setiap parameter yang di-kan. Selanjutnya apabila nilai probabilitas sudah ditentukan

maka probabilitas yang paling tinggi yang akan dijadikan sebagai solusi.

3. Mendapatkan nilai mean, standar deviasi dan probabilitas
4. Solusi kemudian dihasilkan

Untuk klasifikasi dengan data kontinu digunakan rumus Densitas Gaus sebagai berikut:

$$\frac{(-)^2}{\sigma^2} \quad (8)$$

Dimana:

Ω = peluang

x = nilai sample

$\mu = 3,14$

$\sigma = 2,718282$

$\Omega = \text{rata-rata hitung (mean)}$

$\sigma = \text{standar deviasi}$

Selanjutnya setelah diketahui nilai *likelihood* semua kelas yaitu menormalisasi nilai probabilitas setiap kelas dengan cara mengalikan semua hasil variable. Setelah dilakukan pengalian kemudian dicari nilai tertinggi sebagai keputusan atau nilai akhir.

Contoh kasus *data training* nominal penyelesaian dengan menggunakan *Naïve Bayes*:

Tabel 2.3. Kondisi Lapangan (Mahindra, 2016)

No	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes

No	Outlook	Temperature	Humidity	Windy	Play
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Apabila kita akan memprediksi apakah hari selanjutnya diperbolehkan bermain apa tidak dengan menggunakan dataset sebagai berikut :

Tabel 2.4. Data Testing Prediksi Bermain

Outlook	Temperatute	Humidity	Windy	Play
Sunny	Mild	High	FALSE	?

1. Tahap pertama yaitu menghitung jumlah dan probabilitas *class/label*

Tabel 2.5. Probabilitas Setiap Outlook untuk Setiap Kategori pada Play

Jumlah Kelas Play		Probabilitas Kelas Play	
No	Yes	No	Yes
4	10	4/14	10/14

2. Tahap kedua menghitung jumlah kasus yang sama dengan kelas yang sama

Tabel 2.6. Probabilitas Setiap Outlook untuk Setiap Kategori pada Play

Outlook	Jumlah Kelas Play		Probabilitas Kelas Play	
	No	Yes	No	Yes
Sunny	3	2	3/4	2/10
Cloudy	0	4	0/4	4/10
Rainy	1	4	1/4	4/10

Tabel 2.7. Probabilitas Setiap Temperature untuk Setiap Kategori pada Play

Temperature	Jumlah Kelas Play		Probabilitas Kelas Play	
	No	Yes	No	Yes
Hot	2	2	2/4	2/10
Mild	2	4	2/4	4/10
Cool	0	4	0/4	4/10

Tabel 2.8 Probabilitas Setiap Humidity untuk Setiap Kategori pada Play

Humidity	Jumlah Kelas Play		Probabilitas Kelas Play	
	No	Yes	No	Yes
High	4	3	4/4	3/10
Normal	0	7	0/4	7/10

Tabel 2.9. Probabilitas Setiap Windy untuk Setiap Kategori pada Play

Windy	Jumlah Kelas Play		Probabilitas Kelas Play	
	No	Yes	No	Yes
FALSE	2	6	2/4	6/10
TRUE	2	4	2/4	4/10

3. Tahap ketiga kalikan semua variable Yes dan No

a. Probabilitas Yes :

$$\begin{array}{ccccccc}
 & 2 & & 4 & & 3 & \\
 \hline
 = & \overline{10} & & \overline{10} & & \overline{10} & \\
 & 10 & & 10 & & 10 & \\
 & & & & & & 14
 \end{array}
 = 0.010286$$

b. Probabilitas No:

$$\begin{array}{ccccccc}
 & 3 & & 2 & & 4 & \\
 \hline
 = & \overline{4} & & \overline{4} & & \overline{4} & \\
 & 4 & & 4 & & 4 & \\
 & & & & & & 14
 \end{array}
 = 0.05357$$

Setelah sudah diketahui nilai probabilitas Yes dan No, karena nilai hasil No lebih besar dibandingkan dengan Yes maka keputusannya adalah No.

Contoh kasus *data training* nominal dan numerik penyelesaian dengan menggunakan *Naïve Bayes* :

Tabel 2.10. Sampel Data (Yudistira, 2014)

Data ke	Usia (thn)	TB (cm)	BB (Kg)	JK	LPT (cm)	LP (cm)	Status Gizi
1	19	169	51	L	15	63	KTR
2	18	172	50	L	14.5	60	KTB
3	18	161	50	L	12	72	N
4	21	165	57	L	15	71	N
5	24	170	74	L	16	85	OTR

Data ke	Usia (thn)	TB (cm)	BB (Kg)	JK	LPT (cm)	LP (cm)	Status Gizi
6	20	171	126	L	20	113	OTB
7	18	168	47	L	13	59	KTB
8	20	170	75	L	20	79	OTR
9	18	173	76	L	17	80	OTR
10	19	168	51	L	14	68	KTR
11	24	160	59	L	15	71	N
12	19	167	75	L	16	91	OTR
13	25	166	74	L	19	84	OTR
14	22	170	72	L	16	79	N
15	18	167	50	L	14	67	KTR
16	24	169	55	L	14	76	N
17	20	162	56	L	17	75	N
18	26	173	56	L	14	72	N
19	18	163	55	L	16	71	N
20	19	169	80	L	18	84	OTB
21	20	164	53	L	15	71	N
22	21	151	58	P	15.5	79	OTR
23	23	151	51	P	14.5	70	N
24	19	156	64	P	16	77	OTR
25	20	159	49	P	14	65	N
26	22	158	48	P	13	71	N
27	21	155	53	P	13	73	N
28	25	153	51	P	13	72	N
29	20	152	52	P	12	64	N
30	23	150	49	P	14	66.5	N
31	25	159	69	P	15	77	OTB
32	21	155	47.5	P	14	66	N
33	20	148	46	P	14	67	N
34	18	148	48	P	14	65.5	N

Data ke	Usia (thn)	TB (cm)	BB (Kg)	JK	LPT (cm)	LP (cm)	Status Gizi
35	20	151	58	P	15.5	77	OTR
36	22	152	43	P	13	60	N
37	23	161	50	P	15	62.5	N
38	22	153	60	P	17	78	OTR
39	21	162	48	P	13.5	63	KTR
40	24	154	46	P	13	64	N

Keterangan :

- KTB = Kurus Tingkat Berat
- KTR = Kurus Tingkar Ringan
- N = Normal
- OTR = Obesitas Tingkat Ringan
- OTB = Obesitas Tingkat Berat
- LPT = Lingkar Pergelangan Tangan
- LP = Lingkar Perut

A. Pengelompokkan Variabel

Data Numerik :

1. Usia (tahun)
2. Tinggi Badan (cm)
3. Berat Badan (kg)
4. Lingkar Pergelangan Tangan (cm)
5. Lingkar Perut (cm)

1. Status Gizi
2. Jenis Kelamin

B. Menghitung nilai Mean dan Standar Deviasi

Tabel 2.11. Mean dan Standar Deviasi Variabel Usia pada Setiap Kategori

Data ke-	KTB			KTR			N			OBR			OTB			
	μ	($-\mu$)		μ	($-\mu$)		μ	($-\mu$)		μ	($-\mu$)		μ	($-\mu$)		
1	18	18	0	0	19	19.25	-0.25	0.0625	18	21.591	-3.591	12.895	24	20.889	3.111	9.679
2	18	18	0	0	19	19.25	-0.25	0.0625	21	21.591	-0.591	0.349	20	20.889	-0.889	0.790
3				18	19.25	-1.25	1.5625		24	21.591	2.409	5.804	18	20.889	-2.889	8.346
4				21	19.25	1.75	3.0625		22	21.591	0.409	0.167	19	20.889	-1.889	3.568
5									24	21.591	2.409	5.804	25	20.889	4.111	16.901
6									20	21.591	-1.591	2.531	21	20.889	0.111	0.012
7									26	21.591	4.409	19.440	19	20.889	-1.889	3.568
8									18	21.591	-3.591	12.895	20	20.889	-0.889	0.790
9									20	21.591	-1.591	2.531	22	20.889	1.111	1.235
10									23	21.591	1.409	1.986				
11									20	21.591	-1.591	2.531				
12									22	21.591	0.409	0.167				
13									21	21.591	-0.591	0.349				
14									25	21.591	3.409	11.622				
15									20	21.591	-1.591	2.531				
16									23	21.591	1.409	1.986				
17									21	21.591	-0.591	0.349				
18									20	21.591	-1.591	2.531				
19									18	21.591	-3.591	12.895				
20									22	21.591	0.409	0.167				
21									23	21.591	1.409	1.986				
22									24	21.591	2.409	5.804				
($-\mu$)	0.000			4.750			107.318			44.889			20.667			
n	2			4			22			9			3			
Varian	0.000			1.583			5.110			5.611			10.333			
SD	0.000			1.258			2.261			2.369			3.215			

Tabel 2.12. Mean dan Standar Deviasi Variabel TB pada Setiap Kategori

Data ke-	KTB			KTR			N			OBR			OTB			
		μ	(- μ)			μ	(- μ)			μ	(- μ)			μ	(- μ)	
1	172	170	2	4	169	166.5	2.5	6.25	161	158.3182	2.681818	7.192149	170	161.8889	8.111111	65.79012
2	168	170	-2	4	168	166.5	1.5	2.25	165	158.3182	6.681818	44.64669	170	161.8889	8.111111	65.79012
3					167	166.5	0.5	0.25	160	158.3182	1.681818	2.828512	173	161.8889	11.111111	123.4568
4					162	166.5	-4.5	20.25	170	158.3182	11.68182	136.4649	167	161.8889	5.111111	26.12346
5									169	158.3182	10.68182	114.1012	166	161.8889	4.111111	16.90123
6									162	158.3182	3.681818	13.55579	151	161.8889	-10.8889	118.5679
7									173	158.3182	14.68182	215.5558	156	161.8889	-5.88889	34.67901
8									163	158.3182	4.681818	21.91942	151	161.8889	-10.8889	118.5679
9									164	158.3182	5.681818	32.28306	153	161.8889	-8.88889	79.01235
10									151	158.3182	-7.31818	53.55579				
11									159	158.3182	0.681818	0.464876				
12									158	158.3182	-0.31818	0.10124				
13									155	158.3182	-3.31818	11.01033				
14									153	158.3182	-5.31818	28.28306				
15									152	158.3182	-6.31818	39.91942				
16									150	158.3182	-8.31818	69.19215				
17									155	158.3182	-3.31818	11.01033				
18									148	158.3182	-10.3182	106.4649				
19									148	158.3182	-10.3182	106.4649				
20									152	158.3182	-6.31818	39.91942				
21									161	158.3182	2.681818	7.192149				
22									154	158.3182	-4.31818	18.64669				
(-)	8.000			29.000			1080.773			648.889			82.667			
n	2			4			22			9			3			
Varian	8.000			9.667			51.465			81.111			41.333			
SD	2.828			3.109			7.174			9.006			6.429			

Tabel 2.13. Mean dan Standar Deviasi Variabel BB pada Setiap Kategori

Data ke-	KTB			KTR			N			OTR			OTB		
		μ	(- μ)			μ	(- μ)			μ	(- μ)			μ	(- μ)
1	50	48.5	1.5	2.25	51	50	1	1	50	52.11364	-2.11364	4.467459	74	68.22222	5.777778
2	47	48.5	-1.5	2.25	51	50	1	1	57	52.11364	4.886364	23.87655	75	68.22222	6.777778
3					50	50	0	0	59	52.11364	6.886364	47.422	76	68.22222	7.777778
4					48	50	-2	4	72	52.11364	19.88636	395.4675	75	68.22222	6.777778
5									55	52.11364	2.886364	8.331095	74	68.22222	5.777778
6									56	52.11364	3.886364	15.10382	58	68.22222	-10.2222
7									56	52.11364	3.886364	15.10382	64	68.22222	-4.2222
8									55	52.11364	2.886364	8.331095	58	68.22222	-10.2222
9									53	52.11364	0.886364	0.78564	60	68.22222	-8.2222
10									51	52.11364	-1.11364	1.240186			
11									49	52.11364	-3.11364	9.694731			
12									48	52.11364	-4.11364	16.922			
13									53	52.11364	0.886364	0.78564			
14									51	52.11364	-1.11364	1.240186			
15									52	52.11364	-0.11364	0.012913			
16									49	52.11364	-3.11364	9.694731			
17									47.5	52.11364	-4.61364	21.28564			
18									46	52.11364	-6.11364	37.37655			
19									48	52.11364	-4.11364	16.922			
20									43	52.11364	-9.11364	83.05837			
21									50	52.11364	-2.11364	4.467459			
22									46	52.11364	-6.11364	37.37655			
(- μ) ²	4.500			6.000			758.966			513.556			1828.667		
n	2			4			22			9			3		
Varian	4.500			2.000			36.141			64.194			914.333		
SD	2.121			1.414			6.012			8.012			30.238		

Tabel 2.14. Mean dan Standar Deviasi Variabel LPT pada Setiap Kategori

Data ke-	KTB			KTR			N			OTR			OTB			
		μ	(- μ)			μ	(- μ)			μ	(- μ)			μ	(- μ)	
1	14.5	13.75	0.75	0.5625	15	14.125	0.875	0.765625	12	14.11364	-2.11364	4.467459	16	16.88889	-0.88889	0.790123
2	13	13.75	-0.75	0.5625	14	14.125	-0.125	0.015625	15	14.11364	0.886364	0.78564	20	16.88889	3.111111	9.679012
3					14	14.125	-0.125	0.015625	16	14.11364	1.886364	3.558368	17	16.88889	0.111111	0.012346
4					13.5	14.125	-0.625	0.390625	15	14.11364	0.886364	0.78564	16	16.88889	-0.88889	0.790123
5									14	14.11364	-0.11364	0.012913	19	16.88889	2.111111	4.45679
6									17	14.11364	2.886364	8.331095	15.5	16.88889	-1.38889	1.929012
7									14	14.11364	-0.11364	0.012913	16	16.88889	-0.88889	0.790123
8									16	14.11364	1.886364	3.558368	15.5	16.88889	-1.38889	1.929012
9									15	14.11364	0.886364	0.78564	17	16.88889	0.111111	0.012346
10									14.5	14.11364	0.386364	0.149277				
11									14	14.11364	-0.11364	0.012913				
12									13	14.11364	-1.11364	1.240186				
13									13	14.11364	-1.11364	1.240186				
14									13	14.11364	-1.11364	1.240186				
15									12	14.11364	-2.11364	4.467459				
16									14	14.11364	-0.11364	0.012913				
17									14	14.11364	-0.11364	0.012913				
18									14	14.11364	-0.11364	0.012913				
19									14	14.11364	-0.11364	0.012913				
20									13	14.11364	-1.11364	1.240186				
21									15	14.11364	0.886364	0.78564				
22									13	14.11364	-1.11364	1.240186				
(- μ)	1.125			1.188			33.966			20.389			12.667			
n	2			4			22			9			3			
Varian	1.125			0.396			1.617			2.549			6.333			
SD	1.061			0.629			1.272			1.596			2.517			

Tabel 2.15. Mean dan Standar Deviasi Variabel LP pada Setiap Kategori

Data ke-	KTB			KTR			N			OTR			OTB			
		μ	(- μ)			μ	(- μ)			μ	(- μ)			μ	(- μ)	
1	60	59.5	0.5	0.25	63	65.25	-2.25	5.0625	72	69.29545	2.704545	7.314566	85	81.11111	3.888889	15.12346
2	59	59.5	-0.5	0.25	68	65.25	2.75	7.5625	71	69.29545	1.704545	2.905475	79	81.11111	-2.11111	4.45679
3					67	65.25	1.75	3.0625	71	69.29545	1.704545	2.905475	80	81.11111	-1.11111	1.234568
4					63	65.25	-2.25	5.0625	79	69.29545	9.704545	94.1782	91	81.11111	9.888889	97.79012
5									76	69.29545	6.704545	44.95093	84	81.11111	2.888889	8.345679
6									75	69.29545	5.704545	32.54184	79	81.11111	-2.11111	4.45679
7									72	69.29545	2.704545	7.314566	77	81.11111	-4.11111	16.90123
8									71	69.29545	1.704545	2.905475	77	81.11111	-4.11111	16.90123
9									71	69.29545	1.704545	2.905475	78	81.11111	-3.11111	9.679012
10									70	69.29545	0.704545	0.496384				
11									65	69.29545	-4.29545	18.45093				
12									71	69.29545	1.704545	2.905475				
13									73	69.29545	3.704545	13.72366				
14									72	69.29545	2.704545	7.314566				
15									64	69.29545	-5.29545	28.04184				
16									66.5	69.29545	-2.79545	7.814566				
17									66	69.29545	-3.29545	10.86002				
18									67	69.29545	-2.29545	5.269112				
19									65.5	69.29545	-3.79545	14.40548				
20									60	69.29545	-9.29545	86.40548				
21									62.5	69.29545	-6.79545	46.1782				
22									64	69.29545	-5.29545	28.04184				
	0.500			20.750			467.830			174.889			728.667			
n	2			4			22			9			3			
Varian	0.500			6.917			22.278			21.861			364.333			
SD	0.707			2.630			4.720			4.676			19.088			

C. Perhitungan Probabilitas pada data Nominal

Probabilitas kemunculan setiap nilai untuk atribut Status Gizi :

Tabel 2.16. Probabilitas Setiap Kategori pada Status Gizi

	Jumlah Kategori Status Gizi					Probabilitas Status Gizi				
	KTB	KTR	N	OTR	OTB	KTB	KTR	N	OTR	OTB
Jumlah	2	4	22	9	3	0.05	0.1	0.55	0.225	0.075

Probabilitas kemunculan setiap nilai untuk atribut Jenis Kelamin pada setiap kategori Status Gizi:

Tabel 2.17. Probabilitas Setiap JK untuk Setiap Kategori pada Status Gizi

JK	Jumlah Kategori Status Gizi					Probabilitas Status Gizi				
	KTB	KTR	N	OTR	OTB	KTB	KTR	N	OTR	OTB
L	2	3	9	5	2	2/2	3/4	9/22	5/9	2/3
P	0	1	13	4	1	0/2	1/4	13/22	4/9	1/3
Jumlah	2	4	22	9	3	1	1	1	1	1

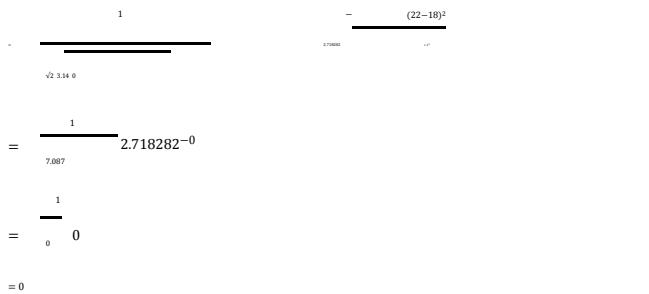
Selanjutnya untuk mengklasifikasi penentu status gizi orang dewasa, sebagai contoh jika diketahui jenis kelamin laki-laki, tinggi badan 170 cm, berat badan 82 cm, usia 22 tahun, lingkar pergelangan tangan 17 dan lingkar perut 100, maka status gizinya ?

D. Perhitungan prediksi dengan *Naïve Bayes* dengan Fungsi *Densitas Gaus*

Setelah dilakukan perhitungan nilai mean dan standar deviasi dari setiap variable yang termasuk kedalam data kontinu maka selanjutnya akan dihitung menggunakan metode *Naïve Bayes* dengan menerapkan rumus Densitas Gaus :

1. Usia

$$\text{a. Usia} = 22 \mid \text{Status Gizi} = \text{KTB}$$



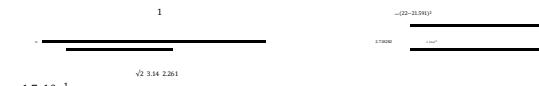
$$\text{b. Usia} = 22 \mid \text{Status Gizi} = \text{KTR}$$



$$= \frac{1}{3.152548} 0.0917$$

$$= 2.9 \cdot 10^{-2}$$

c. Usia = 22 | Status Gizi = N



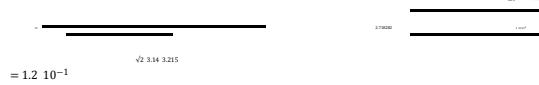
$$= 1.7 \cdot 10^{-1}$$

d. Usia = 22 | Status Gizi = OTR



$$= 1.5 \cdot 10^{-1}$$

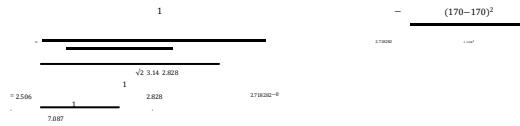
e. Usia = 22 | Status Gizi = OTB



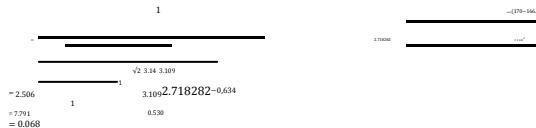
$$= 1.2 \cdot 10^{-1}$$

2. Tinggi Badan

a. Tinggi Badan = 170 | Status Gizi = KTB



b. Tinggi Badan = 170 | Status Gizi = KTR



c. Tinggi Badan = 170 | Status Gizi = N



$$= 0.015$$

d. Tinggi Badan = 170 | Status Gizi = OTR

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 9.006} = 0.030$$

e. Tinggi Badan = 170 | Status Gizi = OTB

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 6.429} = 0.053$$

3. Berat Badan

a. Berat Badan = 82 | Status Gizi = KTB

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 2.121} = 2.506 \cdot 2.121 \\ = 5.315 \cdot 10^{-55} = 1.3$$

b. Berat Badan = 82 | Status Gizi = KTR

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 1.414} = 2.506 \cdot 1.414 \\ = 3.543 \cdot 10^{-112} = 1.7$$

c. Berat Badan = 82 | Status Gizi = N

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 6.012} = 2.506 \cdot 1.414 \\ = 3.543 \cdot 10^{-112} = 1.7$$

d. Berat Badan = 82 | Status Gizi = OTR

$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 8.012} = 1.1 \cdot 10^{-2}$$

e. Berat Badan = 82 | Status Gizi = OTB

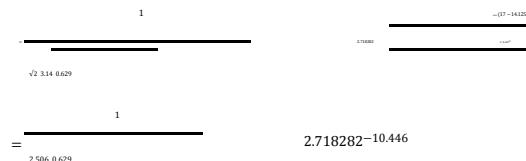
$$\frac{1}{\sqrt{2} \cdot 3.14 \cdot 30.238} = 1.2 \cdot 10^{-2}$$

4. Lingkar Pergelangan Tangan

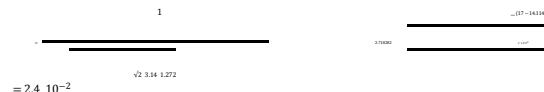
a. Lingkar Pergelangan Tangan = 17 | Status Gizi = KTB



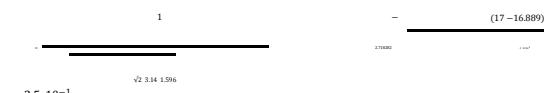
b. Lingkar Pergelangan Tangan = 17 | Status Gizi = KTR



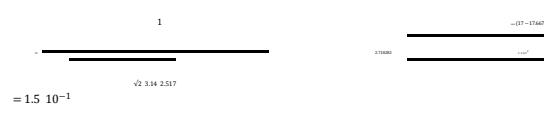
c. Lingkar Pergelangan Tangan = 17 | Status Gizi = N



d. Lingkar Pergelangan Tangan = 17 | Status Gizi = OTR

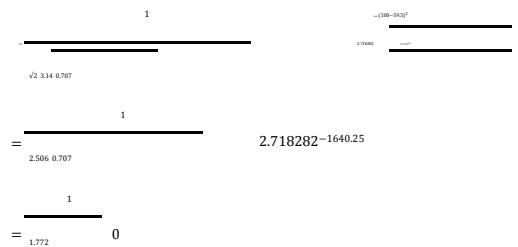


e. Lingkar Pergelangan Tangan = 17 | Status Gizi = OTB



5. Lingkar Perut

a. Lingkar Perut = 100 | Status Gizi = KTB

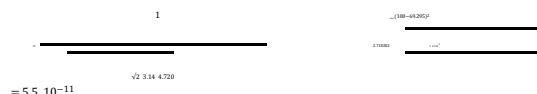


$$= 0$$

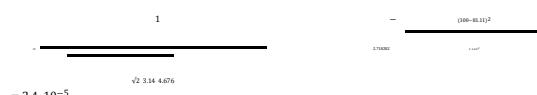
b. Lingkar Perut = 100 | Status Gizi = KTR



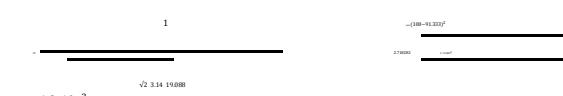
c. Lingkar Perut = 100 | Status Gizi = N



d. Lingkar Perut = 100 | Status Gizi = OTR



e. Lingkar Perut = 100 | Status Gizi = OTB



E. Mengitung Nilai Likelihood

Selanjutnya yaitu melakukan perhitungan rumus likelihood dengan berdasarkan metode Naïve Bayes:

Likelihood KTB $= 0 * 0.141 * 1.3 \cdot 10^{-55} * 3.4 \cdot 10^{-3} * 0 * 0.05$

$$\dots * 1$$

$$= 0$$

Likelihood KTR $= 0.1 * 0.75 * 2.9 \cdot 10^{-2} * 0.068 * 1.7 \cdot 10^{-112}$

$$\dots * 1.8 \cdot 10^{-5} * 1.9 \cdot 10^{-39}$$

$$= 8.6 \cdot 10^{-160}$$

Likelihood N $= 0.55 * 0.409 * 1.7 \cdot 10^{-1} * 0.015 * 2.8 \cdot 10^{-2}$

$$\dots * 2.4 \cdot 10^{-2} * 5.5 \cdot 10^{-11}$$

$$= 2.1 \cdot 10^{-22}$$

Likelihood OTR $= 0.556 * 0.225 * 1.5 \cdot 10^{-1} * 0.030 * 1.1 \cdot 10^{-2}$

$$\dots * 2.4 \cdot 10^{-5} * 2.5 \cdot 10^{-1}$$

Likelihood

$$= 0.075 * 0.667 * 1.3 \cdot 10^{-1} * 0.053 * 1.2 \cdot 10^{-2} * 1.5 \cdot 10^{-1} * 1.9 \cdot 10^{-2}$$

$\approx 3.7 \cdot 10^{-11}$

F. Normalisasi Nilai

Selanjunya yaitu mencari nilai probabilitas dapat dihitung dengan melakukan normalisasi pada setiap kelas KTB, KTR, N, OTR dan OTB.

Probabilitas

$$= \frac{(0 + (8.6 \cdot 10^{-160}) + (2.1 \cdot 10^{-22}) + 3.7 \cdot 10^{-11} + (1.2 \cdot 10^{-8}))}{12 \cdot 10^{-8}}$$

≈ 1

Probabilitas

$$= \frac{(0 + (8.6 \cdot 10^{-160}) + (2.1 \cdot 10^{-22}) + 3.7 \cdot 10^{-11} + (1.2 \cdot 10^{-8}))}{12 \cdot 10^{-8}}$$

$\approx 6.6 \cdot 10^{-140}$

Probabilit

$$= \frac{(0 + (8.6 \cdot 10^{-160}) + (2.1 \cdot 10^{-22}) + 3.7 \cdot 10^{-11} + (1.2 \cdot 10^{-8}))}{12 \cdot 10^{-8}}$$

$\approx 2.1 \cdot 10^{-11}$

Probabilitas

$$= \frac{(0 + (8.6 \cdot 10^{-160}) + (2.1 \cdot 10^{-22}) + 3.7 \cdot 10^{-11} + (1.2 \cdot 10^{-8}))}{12 \cdot 10^{-8}}$$

$\approx 1.7 \cdot 10^{-14}$

Probabilitas

$$= \frac{(0 + (8.6 \cdot 10^{-160}) + (2.1 \cdot 10^{-22}) + 3.7 \cdot 10^{-11} + (1.2 \cdot 10^{-8}))}{12 \cdot 10^{-8}}$$

$\approx 3.07 \cdot 10^{-8}$

≈ 1

Untuk mengetahui apakah orang tersebut termasuk dalam kategori KTB, KTR, N, OTR atau OTB dengan cara melihat hasil nilai akhir probabilitas tertinggi. Dari hasil perhitungan normalisasi nilai probabilitas KTB = 0, KTR = $7.14 \cdot 10^{-152}$, N = $1.7 \cdot 10^{-14}$, OTR = $3.07 \cdot 10^{-3}$ dan OTB = 1 , maka orang dewas tersebut memiliki status gizi Obesitas Tingkat Berat (OTB).

2.5 Diskritisasi

Diskritisasi merupakan perubahan data kontinu menjadi data kategori (Gorunescu, 2011). Pendiskritan atribut merupakan teknik untuk mengubah sebuah fungsi atau nilai kontinu kedalam bentuk diskrit (Wirawan & Eksistyanto, 2015). Diskritisasi atribut kontinyu merupakan salah satu langkah pengolahan data penting dari ekstraksi pengetahuan (Mirqotussa'adah, Muslim, Sugiharti, Prasetyo, & Alimah, 2017).

Menurut (Rahmawati, 2016) terdapat dua cara untuk melakukan diskritisasi atribut yaitu :

1. *Scanning*

Penelusuran (*scanning*) seluruh dataset untuk semua nilai kontinu yang muncul, kemudian menggunakan nilai tersebut sebagai dominan dari atribut nominal. Teknik ini membuat domain nilai menjadi himpunan tertutup dari nilai yang mungkin muncul, sehingga dataset menjadi '*sparse*'.

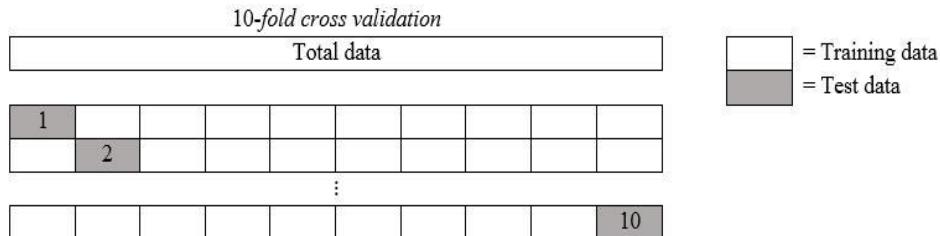
2. *Binning*

Mendefinisikan kumpulan kelas nominal untuk setiap atribut, kemudian menetapkan setiap nilai atribut ke dalam salah satu kelas. Misalnya, jika domain atribut numerik mempunyai nilai dari 0 sampai dengan 100, domain tersebut dapat dibagi menjadi empat bin (0...24, 25...49, 50,,74, 75...100).

2.6 Evaluasi dan Validasi

2.6.1 Cross Validation

Evaluasi dilakukan setelah pembuatan model. Evaluasi merupakan langkah untuk menguji persentase keakuratan model yang telah dibuat. Salah satu metode yang paling populer adalah *cross validation*. Metode ini membagi data menjadi 2 bagian, yaitu data pelatihan dan data pengujian. Selanjutnya, setelah data diuji dilakukan proses silang dimana data pengujian lantas dijadikan data pelatihan ataupun sebaliknya, data pelatihan sebelumnya dijadikan kini menjadi data pengujian. Metode tersebut juga lebih dikenal dengan *k-fold cross validation*. Berikut adalah visualisasi dari sebanyak 10 *fold* (Poetra, 2016):

**Gambar****2.2. Visualisasi dari *k-fold cross validation* sebanyak 10 fold**

Menurut (Rahmi, 2016) pada dasarnya terdapat dua tujuan menggunakan *cross validation* yaitu sebagai berikut:

1. Untuk mengukur kinerja model hasil *learning*. Kendala pada model yang dapat ditemukan dengan *cross validation* diantaranya:
 - a. *Overfitting*, kondisi dimana akurasi untuk data *training* baik namun untuk *data testing* akurasi buruk. Penyebabnya dapat diakibatkan karena model terlalu kompleks, terlalu banyak atribut atau *variable*, data outlier.
 - b. *Underfitting*, kondisi dimana akurasi untuk data *training* buruk namun *data testing* memberi akurasi yang baik. Penyebabnya karena banyak data *training* yang tidak terpetakan dengan baik atau terlalu sedikitnya atribut atau *variable*
2. Untuk membandingkan kinerja model hasil *learning* dari dua atau lebih algoritma dan menemukan algoritma terbaik berdasarkan data yang tersedia.

2.6.2 Confusion Matrix

Confusion matrix merupakan tabel yang terdiri atas banyaknya baris data yang diuji yang diprediksi benar dan tidak benar oleh model klasifikasi (Poetra, 2016).

Menurut (Rahmawati, 2016) model dari *confusion matrix* sebagai berikut:

Tabel 2.18. Confusion Matrix

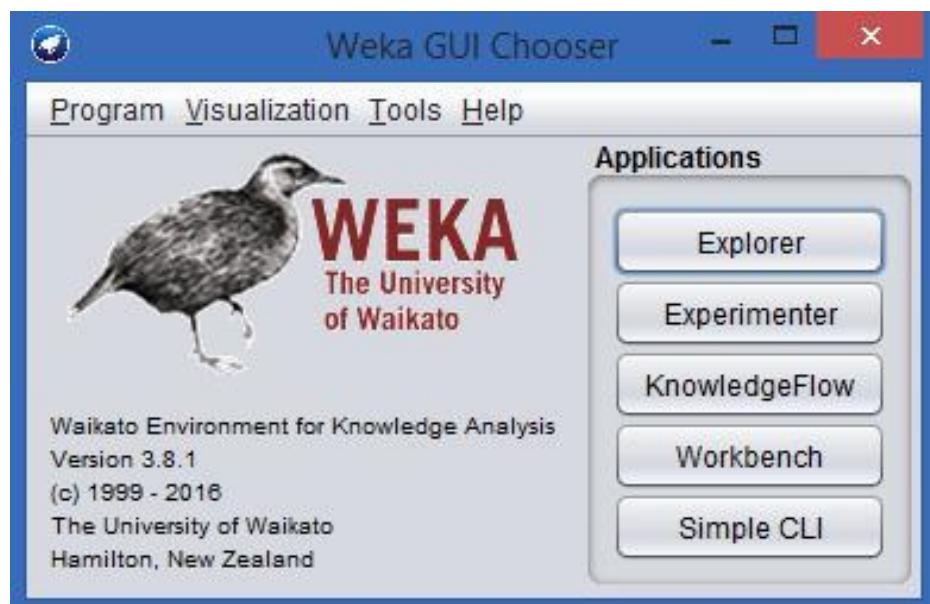
	<i>Classified Positive</i>	<i>Classified Negative</i>
<i>Actual Positive</i>	<i>TP</i>	<i>FN</i>
<i>Actual Negative</i>	<i>FP</i>	<i>TN</i>

Menurut (Goyal & Mehta, 2012) *confusion matrix* merupakan informasi tentang klasifikasi *actual* dan hasil klasifikasi prediksi yang dilakukan oleh sistem.

Akurasi, presentasi jumlah benar dari seluruh data. Untuk menghitung akurasi adalah sebagai berikut:

$$= \frac{\text{jumlah benar}}{\text{jumlah keseluruhan}} \quad (5)$$

2.7 WEKA



Gambar 2.3. Aplikasi WEKA 3.8.1

Waikato Environment for Knowledge Analysis (WEKA) merupakan tool mesin pembelajaran yang diperkenalkan oleh Universitas Waikato, New Zealand yang bersifat *open source* yang dibuat menggunakan dengan Java (GNU Public License) dan bisa digunakan untuk penelitian, pendidikan dan proyek. Weka ini bisa dijalankan di Windows, Linux dan Mac yang terdiri dari beberapa kumpulan algoritma pembelajaran (T. C. Sharma & Jain, 2013).

WEKA diimplementasikan pertama kali pada tahun 1997 dan mulai menjadi *open source* pada tahun 1999 (Mardiana & Nyoto, 2015). Menurut (N. Sharma, Bajpai, & Litoriya, 2012) WEKA merupakan tools data mining yang paling sederhana untuk pengklasifikasian berbagai jenis data dan merupakan model pertama yang menyediakan *graphical user interface*.

Weka telah membuktikan bahwa aplikasi tersebut merupakan alat yang berguna bahkan penting dalam menganalisis suatu *dataset* (Garner, 1995). Perangkat lunak ini memiliki fasilitas untuk melakukan *preprocessing data, classification, regression, clustering, association rules, dan visualization*. (Kristiawan, 2016).

Menurut (Kristiawan, 2016) WEKA memiliki 4 jenis *test option* yang dapat digunakan untuk melakukan proses klasifikasi. Proses yang ditangani adalah proses pelatihan dan pengujian. Keempat jenis test option tersebut yaitu :

a. *Use training set*

Klasifikasi ini menggunakan satu data untuk melakukan pelatihan. Lalu dari seluruh data yang telah dilatih sebelumnya juga digunakan untuk proses pengujian.

b. *Supplied test set*

Klasifikasi ini dilakukan evaluasi dengan cara memprediksi seberapa baik satu dataset yang diambil dari sebuah data tertentu yang memang sudah disediakan untuk pengujian. Proses pelatihan akan dilakukan terlebih dahulu dengan data latih kemudian proses pengujian akan dilakukan dengan data uji yang berbeda dengan data yang dilatih pada klasifikasi.

c. *Cross-validation*

Klasifikasi ini dilakukan evaluasi dengan *cross validation* dan menggunakan jumlah fold yang tertentu yang dapat diinputkan manual. Pada *cross validation* akan ada pilihan beberapa fold yang akan digunakan. Nilai fold default aplikasi yang diberikan adalah 10. Proses pengujian akan dilakukan sebanyak nilai *fold* yang diberikan serta akan dibentuk subset sebanyak nilai fold. Kemudian proses pengujian akan dilakukan menggunakan sebuah subset yang terbentuk dan sisanya akan digunakan untuk proses pelatihannya.

d. *Percentage split*

Klasifikasi ini dilakukan evaluasi dengan melakukan pembagian data antara data uji dan data latih pada satu *dataset* dengan menggunakan persentase. Persentase yang diinputkan akan digunakan untuk proses

pelatihan dan sisanya akan digunakan untuk proses pengujian. Proses ini biasanya dilakukan untuk dengan perbandingan 2/3 data untuk pelatihan dan 1/3 data untuk proses pengujian atau nilai $k = 66\%$.