

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Distribusi**

##### **2.1.1. Pengertian Distribusi**

Distribusi adalah salah satu elemen dari *marketing mix*. Distribusi dapat diartikan sebagai kegiatan pemasaran yang berusaha memperlancar dan mempermudah penyampaian barang dan jasa dari produsen kepada konsumen. Menurut The American Marketing Association dalam buku Manajemen Pemasaran Modern oleh Basu Swastha, (2005:285) saluran merupakan suatu struktur unit organisasi dalam perusahaan dan luar perusahaan yang terdiri atas agen, dealer, pedagang besar dan pengecer, melalui mana sebuah komoditi, produk, atau jasa dipasarkan. Sedangkan saluran distribusi menurut Fandy Tjiptono, Gregorius Chandra dan Dadi Andriana (2008 : 588) didefinisikan sebagai berikut, bahwa “saluran distribusi merupakan serangkaian partisipasi organisasional yang melakukan semua fungsi dibutuhkan untuk menyampaikan produk atau jasa dari penjual ke pembeli akhir.

Menurut Kotler (1985:3) mendefinisikan saluran distribusi sebagai himpunan perusahaan dan perorangan yang mengambil alih hak, atau membantu dalam pengalihan hak atas barang atau jasa tertentu selama barang atau jasa tersebut berpindah dari produsen ke konsumen. Sedangkan Basu Swastha DH (2009:190) mendefinisikan saluran distribusi untuk suatu barang adalah saluran yang digunakan oleh produsen untuk menyalurkan barang tersebut dari produsen sampai ke konsumen atau pemakai industri. Saluran distribusi ini merupakan suatu struktur yang menggambarkan alternatif saluran yang dipilih, dan

menggambarkan situasi pemasaran yang berbeda oleh berbagai macam perusahaan atau lembaga usaha.

Pendistribusian yang optimal bergantung pada kemudahan transportasi produk yaitu semakin mudah ditransportasikan, semakin besar kemungkinan digunakannya perantara. Pendistribusian juga bergantung pada tingkat standardisasi produk yaitu semakin terstandardisasi produk, semakin besar kemungkinan digunakannya perantara.

### **2.1.2. Penentuan Pendistribusian Semen**

Keputusan dalam menentukan pendistribusian semen merupakan salah satu upaya perusahaan dalam menentukan cara bagaimana produk-produknya mudah diakses oleh para pelanggan. Perusahaan harus mengembangkan strateginya untuk memastikan agar produk-produknya didistribusikan ke para pelanggan melalui tempat-tempat yang memberi kemudahan bagi mereka. Dalam hal ini pihak Departemen *Supply Chain Management* (SCM) di PT Semen Indonesia, Tbk merupakan departemen yang mengurus tentang penentuan pendistribusian semen di PT Semen Indonesia, Tbk. H-3 di setiap akhir bulan pihak SCM melakukan proses penentuan pendistribusian semen maka dari itu pihak gudang distributor sebelum H-3 di setiap akhir bulan harus sudah mengupdate segala kebutuhan data dengan yang terbaru. Pihak SCM dalam memantau persediaan stok, informasi harga dan persebaran pendistribusian semen saat ini sudah menggunakan Sistem Informasi yang bernama *Dashboard Supply Chain Management* yang berbasis web. Didalam sistem ini terdapat salah satu menu yaitu menu yang menentukan status pendistribusian semen pada gudang distributor.

Adapun kebutuhan *update* data secara reguler dalam menentukan pendistribusian semen di gudang distributor, yaitu :

1. *Stock on hand* (ton), yakni stok yang ada di tempat tujuan, dilaporkan tiap sore oleh distributor atau LT.

2. *Stok in-transit* (ton), yakni stok yang sedang diangkut tetapi belum dibongkar di tempat tujuan, direcord di tempat pemberangkatan di Tuban. Setiap kali truk selesai dibongkar, distributor atau LT wajib melaporkan agar status truk *in transit* bisa diubah.
3. Kecepatan penyerapan pasar (ton atau hari). Tiap sore distributor / LT melaporkan ke SI. Kecepatan penyerapan pasar dilihat dari rata-rata penjualan dalam 1 minggu terakhir.
4. Kapasitas gudang yang ada (ton), *diupdate* tiap ada perubahan.
5. Jumlah *truck in transit* (truk), adalah jumlah truk yang sudah dilepas dari Tuban, tapi belum selesai melakukan *unloading* di tempat tujuan.
6. Kecepatan bongkar (truk / hari), *diupdate* bila ada perubahan.
7. Waktu kirim dari Tuban ke tempat tujuan (hari), diambil sebagai waktu rata-rata.

Menurut pihak SCM terdapat 3 indikator urgensi dalam menentukan pendistribusian semen di gudang distributor, yaitu :

1. *Stock to Demand Ratio* (SDR) yang mencerminkan seberapa kritis posisi stok relatif terhadap permintaan. Tujuan yang diprioritaskan adalah yang nilainya rendah.
2. *Truck to Unloading Ratio* (TUR) yang mencerminkan banyaknya truk yang antri dibandingkan kecepatan bongkar. Nilai TUR rendah mencerminkan bahwa truk yang datang akan langsung atau menunggu sebentar untuk dibongkar.
3. *Stock to Capacity Ratio* (SCR) yang mencerminkan rasio antara beban dengan kapasitas yang tersedia. Nilai SCR rendah berarti masih tersedia tempat yang cukup di gudang untuk menerima tambahan semen.

**Tabel 2.1** Indikator Urgensi Dalam Penentuan Pendistribusian

Indikator	Makna	Formula
Stock to Demand Ratio (SDR)	Rasio antara stok dengan penyerapan pasar. Nilai rendah mengindikasikan perlu ada pengiriman.	$SDR = \frac{[\text{on hand} + \text{in transit}]}{[LT * Demand]}$
Truck to Unloading Ratio (TUR)	Rasio antara jumlah truk in transit dengan kecepatan bongkar. Nilai rendah mengindikasikan tidak banyak truk yang antri bongkar.	$TUR = \frac{[\text{truck in transit}]}{LT * \text{Unloading Rate}}$
Stock to Capacity Ratio (SCR)	Rasio antara posisi stok dengan kapasitas gudang. Nilai rendah mengindikasikan tersedianya tempat untuk menampung stok. SCR minus diberi nilai 0.	$SCR = \frac{[OH + \text{in transit} - D * LT]}{\text{WHS Capacity}}$

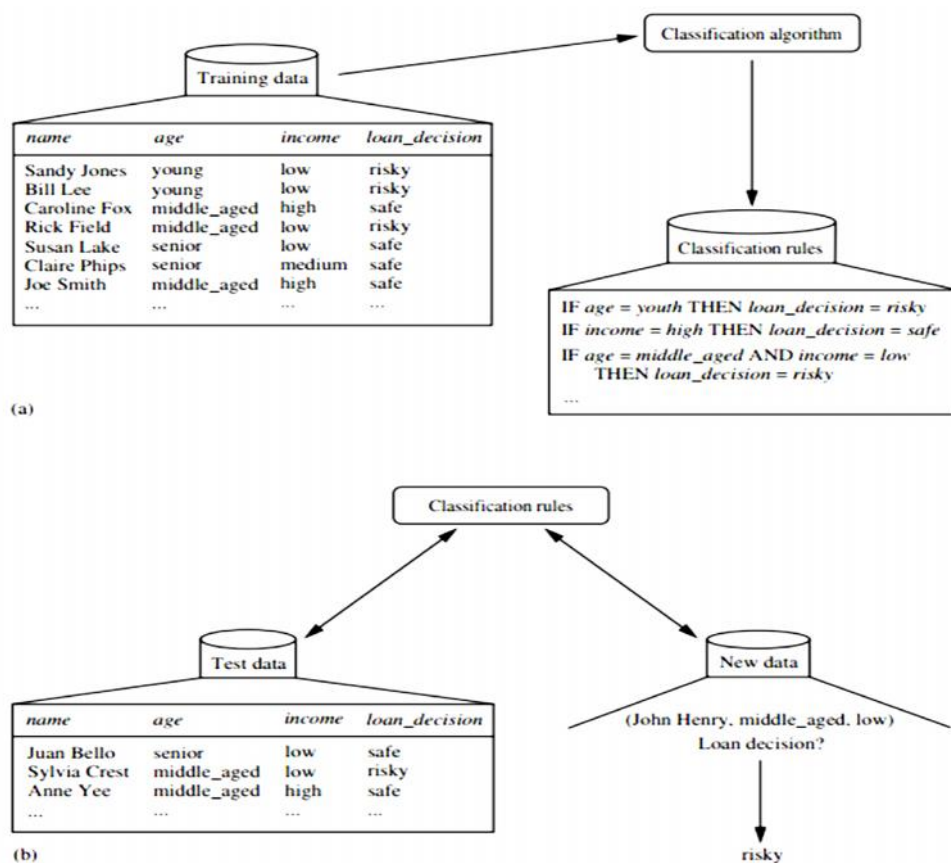
### 2.1.3. Fungsi Saluran Distribusi

1. Dengan fungsi informasi, saluran distribusi dapat mengenali pelanggan, pesaing serta pemasok.
2. Dengan fungsi promosi, saluran distribusi dapat memperkenalkan produk kepada pelanggan sekaligus membujuk mereka untuk membelinya.
3. Dengan fungsi negosiasi, saluran distribusi dapat mencapai persetujuan akhir mengenai harga dan hal-hal lain sehubungan dengan penawaran, sehingga perpindahan hak kepemilikan bisa dilaksanakan.
4. Dengan fungsi pemesanan, saluran distribusi dapat langsung merespon kebutuhan pelanggan, dengan melakukan pemesanan ke gudang atau pabrik.
5. Dengan fungsi pembiayaan, saluran distribusi dapat menentukan dasar pembentukan biaya beserta efisiensinya.
6. Dengan fungsi pengambilan risiko, saluran distribusi dapat memperkirakan risiko apa saja yang dihadapi sehubungan dengan pelaksanaan kegiatan distribusi.

7. Dengan fungsi fisik, saluran distribusi dapat mengatur kesinambungan penyimpanan dan pergerakan produk fisik dari bahan mentah sampai ke pelanggan akhir.
8. Dengan fungsi pembayaran, saluran distribusi dapat menyediakan layanan (media dan cara) pembayaran kepada pelanggan.
9. Dengan fungsi kepemilikan, saluran distribusi dapat ikut mengawasi perpindahan kepemilikan aktual dari organisasi/orang kepada organisasi/orang lain, dalam hal ini dari produsen ke konsumen.

## 2.2 Klasifikasi

Menurut Mike Chapple, klasifikasi adalah teknik data mining yang dilakukan untuk memprediksi kelas atau properti dari setiap *instance* data.



**Gambar 2.1** Tahapan Klasifikasi Data Mining

Tahapan dari klasifikasi dalam data mining terdiri dari :

a. Pembangunan Model

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

b. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut atau class dari sebuah data baru yang atribut atau *class*nya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

### 2.3 Teorema Bayes

Ide dasar aturan *bayes* adalah hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa *evidence* (E) yang diamati.

Teori keputusan *bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang ditimbulkan dalam keputusan-keputusan tersebut.

Hal penting dalam *bayes* adalah

- a. Sebuah probabilitas awal atau priori H atau P(H), adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
- b. Sebuah probabilitas posterior H atau P(H|E), adalah probabilitas dari suatu hipotesis setelah bukti-bukti yang diamati ada.

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

Keterangan :

$P(H|E)$  : Probabilitas posterior bersyarat (*Conditional Probability*) suatu hipotesis H terjadi jika diberikan evidence/bukti E terjadi.

$P(E|H)$  : Probabilitas sebuah evidence E terjadi akan mempengaruhi hipotesis H.

$P(H)$  : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang evidence apapun.

$P(E)$  : Probabilitas awal (priori) *evidence* E terjadi tanpa memandang hipotesis/evidence yang lain.

Contoh Teorema Bayes adalah sebagai berikut :

Peramalan cuaca untuk memperkirakan terjadinya hujan, misal ada faktor yang mempengaruhi terjadinya hujan yaitu mendung. Jika diterapkan dalam *naive bayes* maka probabilitas terjadinya hujan jika bukti mendung sudah diamati :

$$P(\text{Hujan} | \text{Mendung}) = \frac{P(\text{Mendung} | \text{Hujan}) \times P(\text{Hujan})}{P(\text{Mendung})} \dots\dots\dots(2.2)$$

Keterangan :

$P(\text{Hujan} | \text{Mendung})$  : nilai probabilitas hipotesis hujan terjadi jika bukti mendung sudah diamati.

$P(\text{Mendung} | \text{Hujan})$  : probabilitas bahwa mendung yang diamati akan mempengaruhi terjadinya hujan.

$P(\text{Hujan})$  : probabilitas awal hujan tanpa memandang bukti apapun

$P(\text{Mendung})$  : probabilitas terjadinya mendung.

Teorema Bayes juga bisa menangani beberapa evidence, misalnya ada  $E_1$ ,  $E_2$ , dan  $E_3$ , maka probabilitas posterior untuk hipotesis hujan:

$$P(H | E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 | H) \times P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.3)$$

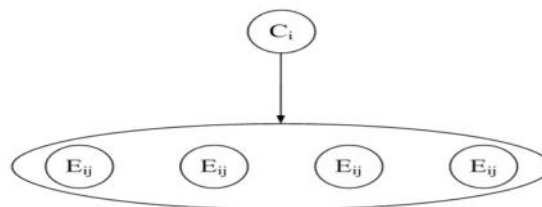
$$P(H | E_1, E_2, E_3) = \frac{P(E_1 | H) \times P(E_2 | H) \times P(E_3 | H) \times P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.4)$$

Bentuk diatas dapat diubah menjadi:

Untuk contoh diatas, jika ditambahkan *evidence* suhu udara dan angin

$$P(\text{Hujan} | \text{Mendung, Suhu, Angin}) = \frac{P(\text{Mendung} | \text{Hujan}) \times P(\text{Suhu} | \text{Hujan}) \times P(\text{Angin} | \text{Hujan}) \times P(\text{Hujan})}{P(\text{Mendung, Suhu, Angin})} \dots\dots\dots(2.5)$$

Pengklasifikasian menggunakan teorema *bayes* ini membutuhkan biaya komputasi yang mahal (waktu processor dan ukuran *memory* yang besar) karena kebutuhan untuk menghitung nilai probabilitas untuk tiap nilai dari perkalian kartesius untuk tiap nilai atribut dan tiap nilai kelas.



**Gambar 2.2** Ilustrasi Teorema Bayes

**2.4 Naive Bayes Classifier**

Klasifikasi *naive bayes* adalah metode yang berdasarkan probabilitas dan teorema *bayes* dengan asumsi bahwa setiap variabel bersifat bebas (*independence*) dan mengasumsikan bahwa keberadaan sebuah fitur tidak ada kaitannya dengan keberadaan fitur yang lain. Asumsi keindependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasikan suatu data. Formulasi *naive bayes* untuk klasifikasi adalah

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \dots\dots\dots(2.6)$$

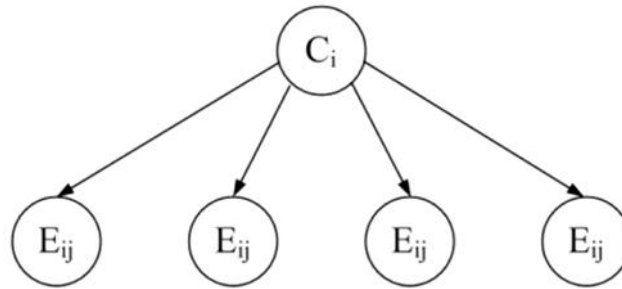
Keterangan :

- P(Y|X) : Probabilitas data dengan vektor X pada kelas Y
- P(Y) : Probabilitas awal kelas Y



$\prod_{i=1}^q P(X_i|Y)$  : Probabilitas independen kelas Y dari semua fitur dalam vektor X

Karena P (X) selalu tetap, sehingga dalam perhitungan prediksi nantinya cukup hanya dengan menghitung  $P(Y) \prod_{i=1}^q P(X_i|Y)$ .



**Gambar 2.3** Ilustrasi Naive Bayes.

Umumnya, *naive bayes* mudah dihitung untuk fitur bertipe kategoris seperti pada contoh diatas. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam Naive Bayes, yaitu :

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasi fitur kontinu ke dalam fitur ordinal.
2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi *Gaussian* biasanya dipilih untuk mempresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas  $P(X_i|Y)$ . Untuk setiap kelas  $y_j$ , probabilitas bersyarat kelas  $y_j$  untuk fitur  $X_i$  adalah

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left( -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \dots\dots\dots(2.7)$$

Keterangan :

$\mu_{ij}$  : mean sampel  $X_i$  ( $\bar{x}$ ) dari semua data latih.

$2\sigma_{ij}^2$  : varian sampel ( $s^2$ ) dari data latih.

Untuk mengukur nilai akurasi yang didapat dari hasil pengujian, menggunakan rumus 2.8. Sedangkan untuk mengukur tingkat kesalahannya menggunakan rumus 2.9.

$$Akurasi = \frac{Jumlah\ data\ yang\ diprediksi\ secara\ benar}{Jumlah\ prediksi\ yang\ dilakukan} \quad (2.8)$$

$$Laju\ error = \frac{Jumlah\ data\ yang\ diprediksi\ secara\ salah}{Jumlah\ prediksi\ yang\ dilakukan} \quad (2.9)$$

Sensitivitas akan mengukur proporsi positif asli yang dikenali (diprediksi) secara benar sebagai positif asli. Rumus perhitungannya menggunakan rumus 2.10. Sedangkan spesifisitas akan mengukur proporsi negatif asli yang dikenali (diprediksi) secara benar sebagai negatif asli. Rumus perhitungannya menggunakan rumus 2.10.

$$Sensitivitas = \frac{TP}{TP+FN} \dots\dots\dots (2.10)$$

Keterangan:

TP : Hasil data distribusi semen dengan kelas prioritas yang diklasifikasikan secara benar mempunyai kelas prioritas

FN : Hasil data distribusi semen dengan kelas prioritas yang diklasifikasikan secara salah mempunyai kelas bukan prioritas.

$$Spesifisitas = \frac{TN}{FP+TN} \dots\dots\dots (2.11)$$

Keterangan:

TN : Hasil data distribusi semen dengan kelas bukan prioritas yang diklasifikasikan secara salah mempunyai kelas prioritas.

FP : Hasil data distribusi semen dengan kelas bukan prioritas yang diklasifikasikan secara benar mempunyai kelas bukan prioritas.

### 2.4.1 Algoritma Klasifikasi *Naive Bayes*

Algoritma klasifikasi *naive bayes* dihitung sesuai dengan rumus *naive bayes*  $P(Y) \prod_{i=1}^q P(X_i|Y)$ , yang langkah-langkah perhitungannya dijelaskan sebagai berikut :

1. Menghitung nilai probabilitas kelas berdasarkan data latih  $P(Y)$
2. Menghitung nilai probabilitas tiap fitur berdasarkan data latih  $\prod_{i=1}^q P(X_i|Y)$

Untuk fitur bertipe numerik menggunakan rumus berikut :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}\right)$$

Fitur numerik berikut ini dihitung tiap data uji.

3. Menghitung nilai probabilitas akhir  
Mengalikan hasil dari  $P(Y)$  dan  $\prod_{i=1}^q P(X_i|Y)$  pada masing-masing kelas dan data uji.
4. Data uji akan diklasifikasikan pada kelas dengan nilai probabilitas akhir terbesar.

Berikut flowchart perhitungan *naive bayes* berdasarkan penjelasan diatas.



**Gambar 2.4** Flowchart *Naive Bayes*

### 2.4.2 Karakteristik *Naive Bayes*

Karakteristik *naive bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *naive bayes* sebagai berikut :

1. Metode *naive bayes* teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*). *naive bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *naive bayes* karena asumsi independensi tersebut sudah tidak ada.

*Naive Bayes* memiliki beberapa keuntungan dan kekurangan yaitu sebagai berikut :

1. Keuntungan *Naive Bayes*
  - a. Cepat dan efisiensi ruang.
  - b. Kokoh terhadap atribut yang tidak relevan.
  - c. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
  - d. Menangani kuantitatif dan data diskrit.
2. Kekurangan *Naive Bayes*
  - a. Tidak berlaku jika *probabilitas* kondisionalnya adalah nol, apabila nol maka *probabilitas* prediksi akan bernilai nol juga.
  - b. Mengasumsikan variabel bebas.

Contoh perhitungan *naive bayes* adalah sebagai berikut :

1. Jika ada sebuah data uji berupa hewan musang dengan nilai fitur: penutup kulit = rambut, melahirkan = ya, berat = 15, masuk kelas manakah untuk hewan musang tersebut ?

**Tabel 2.2** Contoh Data Latih Klasifikasi Hewan

<b>Nama hewan</b>	<b>Penutup kulit</b>	<b>Melahirkan</b>	<b>Berat</b>	<b>Kelas</b>
Ular	Sisik	Ya	10	Reptil
Tikus	Bulu	Ya	0.8	Mamalia
Kambing	Rambut	Ya	21	Mamalia
Sapi	Rambut	Ya	120	Mamalia
Kadal	Sisik	Tidak	0.4	Reptil
Kucing	Rambut	Ya	1.5	Mamalia
Bekicot	Cangkang	Tidak	0.3	Reptil
Harimau	Rambut	Ya	43	Mamalia
Rusa	Rambut	Ya	45	Mamalia
Kura-Kura	Cangkang	Tidak	7	Reptil

**Tabel 2.3** Contoh Perhitungan Nilai Probabilitas Tiap Fitur

<b>Penutup kulit</b>		<b>Melahirkan</b>	
Mamalia	Reptil	Mamalia	Reptil
Sisik = 0	Sisik = 2	Ya = 6	Ya = 1
Bulu = 1	Bulu = 0	Tidak = 0	Tidak = 3
Rambut = 5	Rambut = 0		
Cangkang = 0	Cangkang = 2		

P(Kulit = Sisik   Mamalia) = 0	P(Kulit = Sisik   Reptil) = 0.5	P(Lahir = Ya   Mamalia) = 1	P(Lahir = Ya   Reptil) = 0.25
P(Kulit = Bulu   Mamalia) = 1/6	P(Kulit = Bulu   Reptil) = 0	P(Lahir = Tidak   Mamalia) = 0	P(Lahir = Tidak   Reptil) = 0.75
P(Kulit = Rambut   Mamalia) = 5/6	P(Kulit = Rambut   Reptil) = 0		
P(Kulit = Cangkang   Mamalia) = 0	P(Kulit = Cangkang   Reptil) = 0.5		
<b>Berat</b>		<b>Kelas</b>	
Mamalia	Reptil	Mamalia	Reptil
$\bar{x}_{mamalia} = 38.55$ $s^2_{mamalia} = 1960.255$ $s_{mamalia} = 44.275$	$\bar{x}_{reptil} = 4.425$ $s^2_{reptil} = 23.6425$ $s_{reptil} = 4.8624$	Mamalia = 6 P(Mamalia) = 6/10 = 0.6	Reptil = 4 P(Reptil) = 4/10 = 0.4

Hitung nilai probabilitas untuk fitur dengan tipe numerik yaitu berat.

$$P(\text{Berat} = 15 | \text{Mamalia}) = \frac{1}{\sqrt{2f} \cdot 44.275} \exp \frac{-(15-38.55)^2}{2 \times 1960.255} = 0.0078$$

$$P(\text{Berat} = 15 | \text{Reptil}) = \frac{1}{\sqrt{2f} \cdot 4.8624} \exp \frac{-(15-4.425)^2}{2 \times 23.6425} = 0.0077$$

Hitung probabilitas akhir untuk setiap kelas:

$$\begin{aligned} P(X | \text{Mamalia}) &= P(\text{Kulit} = \text{Rambut} | \text{Mamalia}) \times P(\text{Lahir} = \text{Ya} | \text{Mamalia}) \times \\ &\quad P(\text{Berat} = 15 | \text{Mamalia}) \\ &= 5/6 \times 1 \times 0.0078 = 0.0065 \end{aligned}$$

$$\begin{aligned} P(X | \text{Reptil}) &= P(\text{Kulit} = \text{Rambut} | \text{Reptil}) \times P(\text{Lahir} = \text{Ya} | \text{Reptil}) \times P(\text{Berat} \\ &= 15 | \text{Reptil}) \\ &= 0 \times 0.25 \times 0.0077 = 0 \end{aligned}$$

Nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir:

$$\begin{aligned} P(\text{Mamalia} | X) &= \alpha \times P(\text{Mamalia}) \times P(X | \text{Mamalia}) \\ &= \alpha \times 0.6 \times 0.0065 \\ &= 0.0039\alpha \end{aligned}$$

$$P(\text{Reptil} | X) = \alpha \times P(\text{Reptil}) \times P(X | \text{Reptil}) = \alpha \times 0.4 \times 0 = 0$$

Untuk  $\alpha = 1/P(X)$  pasti nilainya konstan sehingga tidak perlu diketahui karena terbesar dari dua kelas tersebut tidak dapat dipengaruhi  $P(X)$ .

Karena nilai probabilitas akhir (posterior probability) terbesar ada di kelas **mamalia (0.0039)**, maka data uji musang diprediksi sebagai kelas **mamalia**.

## 2.5 Penelitian Sebelumnya

*Naive bayes* merupakan metode populer yang banyak digunakan untuk klasifikasi. Beberapa riset yang telah dilakukan berkaitan dengan kasus klasifikasi yang menggunakan metode *naive bayes* dan penelitian tentang bantuan langsung sementara masyarakat, antara lain :

Penelitian oleh Gresika Duita pada tahun 2016 yang berjudul “*Klasifikasi Penerimaan Bantuan Langsung Sementara Masyarakat (BLSM) Dengan Metode Naive Bayes (Studi Kasus: Kelurahan Pekauman Kecamatan Gresik)*”. Adapun permasalahan yang diambil dalam penelitian ini yaitu Bagaimana mengklasifikasikan masyarakat yang berhak menerima Bantuan Langsung Sementara Masyarakat (BLSM) di Kelurahan Pekauman Kecamatan Gresik. Penelitian ini bertujuan untuk membuat sistem yang dapat mengetahui masyarakat yang berhak menerima Bantuan Langsung Sementara Masyarakat (BLSM) di Kelurahan Pekauman Kecamatan Gresik. Atribut yang digunakan sebagai data latih sistem adalah penghasilan, kondisi rumah, makan/hari, jumlah keluarga, tanggungan anak sekolah dan kelas. Dari 458 data dari Kelurahan Pekauman, data tersebut diambil 40% yang akan dijadikan sebagai data uji dan 60% akan menjadi data training. Jadi jumlah pembagiannya adalah 275 data sebagai data training dan 183 data untuk data uji. Keakuratan ketepatan hasil klasifikasi dengan *naive bayes* untuk mengetahui tingkat kecenderungan penyelesaian studi mahasiswa sistem memiliki akurasi kebenaran sistem sebesar 99.45% dengan nilai error 0.55%.

Sedangkan penelitian lain yang digunakan sebagai rujukan dalam tugas akhir ini adalah penelitian dilakukan oleh Friday Aries L pada tahun 2015

dengan judul “*Aplikasi Klasifikasi Penentuan Penerimaan Beras Miskin (Raskin) di Desa Sidomulyo Kecamatan Deket Kabupaten Lamongan Dengan Metode Klasifikasi Naive Bayes*”. Adapun permasalahan yang diambil dalam penelitian ini yaitu bagaimana cara menentukan masyarakat yang berhak menerima beras miskin (Raskin) berdasarkan regulasi Pemerintah di Ds. Sidomulyo Kec. Deket Kab. Lamongan. Penelitian ini bertujuan untuk membuat sistem klasifikasi yang dapat menentukan pembagian beras miskin (Raskin) kepada masyarakat Ds. Sidomulyo Kec. Deket Kab. Lamongan dengan lebih tepat sasaran. Atribut yang digunakan adalah pekerjaan, penghasilan, harta benda (kendaraan), kondisi rumah, jumlah keluarga dan kelas. Penelitian ini menggunakan 60 data yang diperoleh dari desa sidomulyo kecamatan deket kabupaten lamongan dalam percobaan dengan pembagian 50 data latih dan 10 data uji. Tingkat akurasi prediksi 85 %, laju error 15%, sensitivitas 99,17% dan spesifisitas 70,59%.