### Contoh tipe Data numerik

#### **Dataset Awal**

| **Fitur 1 (Kata “Diskon”)** | **Fitur 2 (Kata “Beli”)** | **Label (Kelas)** |
| --- | --- | --- |
| 5 | 3 | Spam |
| 0 | 1 | Bukan Spam |
| 7 | 6 | Spam |
| 1 | 0 | Bukan Spam |
| 4 | 2 | Spam |

#### **Langkah-Langkah Perhitungan**

##### **Langkah 1: Hitung Probabilitas Prior ()**

* Jumlah total data:
* Jumlah email spam:
* Jumlah email bukan spam:

##### **Langkah 2: Hitung Mean dan Varians untuk Setiap Fitur**

Untuk setiap kelas (Spam dan Bukan Spam), hitung mean () dan varians () dari setiap fitur.

###### **Fitur 1 (Kata “Diskon”)**

* **Kelas Spam**: Data = {5, 7, 4}
  + Mean ():
  + Variance ():
* **Kelas Bukan Spam**: Data = {0, 1}
  + Mean ():
  + Variance ():

###### **Fitur 2 (Kata “Beli”)**

* **Kelas Spam**: Data = {3, 6, 2}
  + Mean ():
  + Variance ():
* **Kelas Bukan Spam**: Data = {1, 0}
  + Mean ():
  + Variance ():

##### **Langkah 3: Rumus Gaussian**

Rumus Gaussian untuk menghitung likelihood:

##### **Langkah 4: Prediksi Email Baru**

Misalkan email baru memiliki fitur: - Kata “diskon”: - Kata “beli”:

Hitung untuk setiap kelas ( dan ).

###### **Kelas Spam**

1. **Likelihood untuk Fitur 1 (“Diskon”)**:
   * Hitung komponen:
   * Hasil:
2. **Likelihood untuk Fitur 2 (“Beli”)**:
   * Hitung komponen:
   * Hasil:
3. **Posterior Probability untuk Spam**:

###### **Kelas Bukan Spam**

1. **Likelihood untuk Fitur 1 (“Diskon”)**:
   * Hitung komponen:
   * Hasil:
2. **Likelihood untuk Fitur 2 (“Beli”)**:
   * Hitung komponen:
   * Hasil:
3. **Posterior Probability untuk Bukan Spam**:

##### **Langkah 5: Kesimpulan**

Karena , maka email tersebut diklasifikasikan sebagai **Spam**.

#### 

#### **Contoh Untuk Tipe data Kategorical**

#### **Dataset: Klasifikasi Cuaca**

Dataset ini berisi informasi tentang kondisi cuaca (fitur kategorikal) dan apakah seseorang memutuskan untuk bermain di luar atau tidak (label kelas).

| **Fitur 1 (Cuaca)** | **Fitur 2 (Suhu)** | **Label (Kelas)** |
| --- | --- | --- |
| Cerah | Panas | Tidak Bermain |
| Mendung | Sejuk | Bermain |
| Hujan | Dingin | Tidak Bermain |
| Cerah | Sejuk | Bermain |
| Hujan | Panas | Tidak Bermain |

#### **Penjelasan Dataset**

1. **Fitur 1 (Cuaca)**: Kondisi cuaca, dengan nilai kategorikal: “Cerah”, “Mendung”, “Hujan”.
2. **Fitur 2 (Suhu)**: Kondisi suhu, dengan nilai kategorikal: “Panas”, “Sejuk”, “Dingin”.
3. **Label (Kelas)**: Apakah seseorang memutuskan untuk bermain di luar (“Bermain”) atau tidak (“Tidak Bermain”).

### Contoh untuk tipe data Kategorikal

#### **Langkah-Langkah Perhitungan Naive Bayes untuk Data Kategorikal**

##### **Langkah 1: Hitung Probabilitas Prior ()**

* Jumlah total data:
* Jumlah “Bermain”:
* Jumlah “Tidak Bermain”:

##### **Langkah 2: Hitung Likelihood ()**

Untuk setiap fitur, hitung probabilitas kemunculan setiap nilai kategorikal dalam setiap kelas.

###### **Fitur 1 (Cuaca)**

* **Kelas Bermain**: Data = {“Mendung”, “Cerah”}
* **Kelas Tidak Bermain**: Data = {“Cerah”, “Hujan”, “Hujan”}

###### **Fitur 2 (Suhu)**

* **Kelas Bermain**: Data = {“Sejuk”, “Sejuk”}
* **Kelas Tidak Bermain**: Data = {“Panas”, “Dingin”, “Panas”}

##### **Langkah 3: Prediksi Data Baru**

Misalkan kita memiliki data baru: - Cuaca: “Hujan” - Suhu: “Panas”

Hitung probabilitas posterior untuk setiap kelas menggunakan rumus Naive Bayes:

###### **Kelas Bermain**

###### **Kelas Tidak Bermain**

##### **Langkah 4: Kesimpulan**

Karena , maka prediksi untuk data baru adalah **Tidak Bermain**.

### 

### Contoh Tipe Data Campuran

#### **Dataset: Klasifikasi Pembelian Produk**

| **Fitur 1 (Jenis Kelamin)** | **Fitur 2 (Usia)** | **Label (Kelas)** |
| --- | --- | --- |
| Pria | 25 | Tidak Beli |
| Wanita | 35 | Beli |
| Pria | 45 | Beli |
| Wanita | 20 | Tidak Beli |
| Pria | 30 | Tidak Beli |

#### **Penjelasan Dataset**

1. **Fitur 1 (Jenis Kelamin)**: Data kategorikal dengan nilai “Pria” atau “Wanita”.
2. **Fitur 2 (Usia)**: Data numerik yang menunjukkan usia pelanggan.
3. **Label (Kelas)**: Apakah pelanggan membeli produk (“Beli”) atau tidak (“Tidak Beli”).

#### **Langkah-Langkah Perhitungan Naive Bayes untuk Data Campuran**

##### **Langkah 1: Hitung Probabilitas Prior ()**

* Jumlah total data:
* Jumlah “Beli”:
* Jumlah “Tidak Beli”:

##### **Langkah 2: Hitung Likelihood untuk Fitur Kategorikal**

Untuk fitur kategorikal (Jenis Kelamin), hitung probabilitas kemunculan setiap nilai dalam setiap kelas.

###### **Fitur 1 (Jenis Kelamin)**

* **Kelas Beli**: Data = {“Wanita”, “Pria”}
* **Kelas Tidak Beli**: Data = {“Pria”, “Wanita”, “Pria”}

##### **Langkah 3: Hitung Likelihood untuk Fitur Numerik**

Untuk fitur numerik (Usia), asumsikan data mengikuti distribusi Gaussian (normal). Hitung mean () dan varians () untuk setiap kelas.

###### **Fitur 2 (Usia)**

* **Kelas Beli**: Data = {35, 45}
  + Mean ():
  + Variance ():
* **Kelas Tidak Beli**: Data = {25, 20, 30}
  + Mean ():
  + Variance ():

Rumus Gaussian:

##### **Langkah 4: Prediksi Data Baru**

Misalkan kita memiliki data baru: - Jenis Kelamin: “Pria” - Usia: 32

Hitung probabilitas posterior untuk setiap kelas menggunakan rumus Naive Bayes:

###### **Kelas Beli**

1. **Likelihood untuk Jenis Kelamin**:
2. **Likelihood untuk Usia**:
   * Hitung komponen:
   * Hasil:
3. **Posterior Probability untuk Beli**:

###### **Kelas Tidak Beli**

1. **Likelihood untuk Jenis Kelamin**:
2. **Likelihood untuk Usia**:
   * Hitung komponen:
   * Hasil:
3. **Posterior Probability untuk Tidak Beli**:

##### **Langkah 5: Kesimpulan**

Karena , maka prediksi untuk data baru adalah **Tidak Beli**.