

# Big Data & Data Analytics

## Psikologi Kelas III & IV

### Pertemuan 11

# Analisis Sentimen

- ❑ **Analisis Sentimen** adalah proses pendeteksian penilaian seseorang terhadap objek tertentu.
- ❑ Analisis sentiment digunakan untuk memperoleh persentase penilaian positif dan penilaian negative.
- ❑ **Tujuan** analisis sentiment adalah untuk melakukan penilaian terhadap emosi, pendapat, evaluasi yang disampaikan oleh seseorang terhadap suatu produk, layanan atau tokoh masyarakat.

# Nilai Sentimen

**Positif** : Berupa penilaian seseorang yang cenderung ke bentuk pujian.

Misal : Barangnya bagus, pengiriman barang cepat, besok beli lagi deh, dsb.

**Negatif** : Berupa penilaian seseorang yang cenderung ke arah hinaan atau hujatan atau complain.

Misal : Packingnya tidak rapi, pengiriman terlalu lama, barang tidak sesuai gambar, dsb.

**Netral** : Berupa penilaian seseorang yang tidak ada kecenderungan untuk memuji maupun menghujat.

Misal : pesanan telah diterima, standar, biasa aja, dsb.

# Sistem Temu Kembali Informasi

Data *tweet* yang telah diambil dari sosial media twitter masih merupakan data mentah maka dari itu perlu dilakukan *preprocessing* untuk mendapatkan data yang siap untuk diproses selanjutnya. Tahapan dari *preprocessing* ada *case folding*, *tokenizing*, *stopword removal* dan *stemming*.

# Tahapan Preprocessing Data

## 1. Case Folding

Dalam penulisan tweet sering ditemukan perbedaan bentuk huruf. Tahapan *case folding* akan merubah bentuk huruf menjadi huruf kecil atau disebut juga penyeragaman bentuk huruf.

|                    |  |
|--------------------|--|
| Teks <i>input</i>  | Kapok Belanja Di Lazada Sudah Transfer Uang Barang Tidak Dikirim |
| Teks <i>output</i> | kapok belanja di lazada sudah transfer uang barang tidak dikirim |

## 2. Tokenizing

Tahap *tokenizing* yakni tahap pemotongan string inputan berdasarkan kata yang menyusunnya. Pada dasarnya proses *tokenizing* adalah pemenggalan kalimat menjadi kata.

|                   |   |
|-------------------|---|
| Teks <i>input</i> | kapok belanja di lazada sudah transfer uang barang tidak dikirim          |
| Teks <i>ouput</i> | kapok, belanja, di, lazada, sudah, transfer, uang, barang, tidak, dikirim |

### 3. Stopword Removal

Stopword removal adalah proses menghilangkan kata-kata yang tidak memiliki arti seperti kata “yang”, “di”, “itu” dan lain sebagainya.

|                    |   |
|--------------------|---|
| Teks <i>input</i>  | kapok, belanja, di, lazada, sudah, transfer, uang, barang, tidak, dikirim |
| Teks <i>output</i> | kapok, belanja, lazada, uang, barang, dikirim                             |

### 4. Stemming

|                    |   |
|--------------------|---|
| Teks <i>input</i>  | kapok, belanja, lazada, uang, barang, dikirim   |
| Teks <i>output</i> | <u>kapok</u> , <u>belanja</u> , <u>lazada</u> , <u>uang</u> , <u>barang</u> , <u> kirim</u> |

# Opinion Word

Opinion word atau kata opini merupakan kumpulan kamus kata positif, netral dan negatif dalam bahasa indonesia. Opinion word dibuat pertama kali oleh Liu dalam penelitiannya kemudian dari penelitian diterjemahkan ke dalam bahasa indonesia yang berguna untuk menghitung kata sentimen dalam kalimat. Contoh kata positif, netral dan negatif pada tabel 2.5, 2.6, 2.7 berikut.

**Tabel 2. 5** Kumpulan kata positif

|        |           |         |          |           |         |
|--------|-----------|---------|----------|-----------|---------|
| Abadi  | Adaptif   | Adidaya | Adikuasa | Adil      | Aduhai  |
| Bangun | Agresif   | Ahli    | Anyar    | Aplikatif | Arif    |
| Asih   | Aspiratif | Cepat   | Cerah    | Cerdas    | Canggih |
| Wujud  | Ganteng   | Fasih   | Favorit  | Fenomenal | Janji   |



**Tabel 2. 6** Kumpulan kata negatif

|         |           |        |       |          |       |
|---------|-----------|--------|-------|----------|-------|
| Abai    | Abal-abal | Aborsi | Gagal | Abnormal | Banci |
| Bandel  | Azab      | Babi   | Bodoh | Bolos    | Cabul |
| Tinggal | Buta      | Cebol  | Cekal | Dosa     | Duka  |
| Fatal   | Erotis    | Fitnah | Genit | Goda     | Lulus |

**Tabel 2. 7** Kumpulan kata netral

|        |          |         |                |         |            |
|--------|----------|---------|----------------|---------|------------|
| A      | Aba-aba  | Abad    | Abah           | Abstrak | Masyarakat |
| Bicara | Acung    | Bahasa  | Akta           | Deputi  | Derajah    |
| Derak  | Detektor | Digital | Diskualifikasi | Ebi     | Eceng      |

# Metode Naïve Bayes Classifier

Klasifikasi *bayes* yang ditemukan oleh Thomas Bayes pada abad ke-18. Teori klasifikasi bayes adalah pendekatan statistika yang fundamental dalam data mining. Pendekatan ini berdasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas [2].

Kaidah Bayes

# Langkah – Langkah Metode Naïve Bayes

## 1. Menyiapkan data training yang sudah diberi label

**Tabel 2. 8** Contoh data latih

| Dokumen   | Teks   | Label               |
|-----------|--|---------------------|
| <u>D1</u> | Terimakasih lazada <u>barang</u> pesanan saya sampai dengan selamat pengemasan <u>barang</u> juga bagus. Good. | <u>C2 (Positif)</u> |
| <u>D2</u> | <u>Ngeri belanja di lazada banyak barang palsu</u>   | C1 (Negatif)        |
| <u>D3</u> | Suka belanja di <u>lazada</u> lengkap kualitas bagus harganya murah  | <u>C2 (Positif)</u> |
| D4        | <u>Barang</u> di <u>lazada</u> jelek seperti di pasar  | <u>C1 (Negatif)</u> |

## 2. Menentukan Data Uji

**Tabel 2. 9** Data uji

| Dokumen | Teks   | Label |
|---------|--|-------|
| D5      | Kapok belanja di lazada sudah transfer uang barang tidak dikirim | ?     |

## 3. Menentukan Peluang Data Training

$$P(c_i) = \frac{\text{jumlah dokumen dengan label } c_i}{\text{jumlah seluruh dokumen}}$$

Dari data training di atas diperoleh

$$\underline{P(\text{positif})} = \frac{2}{4} = \frac{1}{2}$$

$$P(\text{negatif}) = \frac{2}{4} = \frac{1}{2}$$

Tabel 2. 10 Frekuensi kemunculan kata

| Kata               | D1 <sup>4</sup> | D2 <sup>-</sup> | D3 <sup>4</sup> | D4 <sup>-</sup> |
|--------------------|-----------------|-----------------|-----------------|-----------------|
| Terimakasih        | <u>1</u>        | 0               | <u>0</u>        | 0               |
| <u>Lazada</u>      | 1               | 1               | 1               | 1               |
| <u>Barang</u>      | 2               | 1               | 0               | 1               |
| <u>Pesan</u>       | 1               | 0               | 0               | 0               |
| Sampai             | 1               | 0               | 0               | 0               |
| Selamat            | 1               | 0               | 0               | 0               |
| Kemas              | 1               | 0               | 0               | 0               |
| Bagus              | 1               | 0               | 1               | 0               |
| Ngeri <sup>u</sup> | 0               | 1               | 0               | 0               |
| Belanja            | 0               | 1               | 1               | 0               |
| Palsu              | 0               | 1               | 0               | 0               |
| Suka               | 0               | 0               | 1               | 0               |
| Lengkap            | 0               | 0               | 1               | 0               |
| Kualitas           | 0               | 0               | 1               | 0               |
| Harga              | 0               | 0               | 1               | 0               |
| Murah              | 0               | 0               | 1               | 0               |
| Jelek              | 0               | 0               | 0               | 1               |
| Pasar              | 0               | 0               | 0               | 1               |

#### 4. Menentukan kemunculan kata dalam dokumen

$$P(\text{terimakasih, positif}) = \frac{1+1}{17+18} = \frac{2}{35}$$

$$P(\text{Lazada, positif}) = \frac{2+1}{17+18} = \frac{3}{35}$$

$$P(\text{terimakasih, negatif}) = \frac{0+1}{9+18} = \frac{1}{27}$$

$$P(\text{Ngeri, negatif}) = \frac{1+1}{27} = \frac{2}{27}$$

## 5. Menentukan peluang untuk setiap kelas data training

$$P(w_{kj} \mid c_i) = \frac{P_k + 1}{n + \text{term}}$$

Di mana

$P(w_{kj} \mid c_i)$  : peluang dari setiap term (kata)

$P_k$  : banyaknya kemunculan term  $w_{kj}$

$n$  : semua jumlah kata yang ada pada topik (label)  $c_i$

$\text{term}$  : jumlah seluruh kata

**a. Peluang untuk kemunculan setiap kata untuk label positif**

$$p(\text{terimakasih, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{lazada, positif}) = (2 + 1) / (17 + 18) = 3 / 35$$

$$p(\text{barang, positif}) = (2 + 1) / (17 + 18) = 3 / 35$$

$$p(\text{pesan, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{sampai, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{selamat, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{kemas, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{bagus, positif}) = (2 + 1) / (17 + 18) = 3 / 35$$

$$p(\text{ngeri, positif}) = (0 + 1) / (17 + 18) = 1 / 35$$

$$p(\text{belanja, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{palsu, positif}) = (0 + 1) / (17 + 18) = 1 / 35$$

$$p(\text{suka, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{lengkap, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{kualitas, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{harga, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{murah, positif}) = (1 + 1) / (17 + 18) = 2 / 35$$

$$p(\text{jelek, positif}) = (0 + 1) / (17 + 18) = 1 / 35$$

$$p(\text{pasar, positif}) = (0 + 1) / (17 + 18) = 1 / 35$$

## b. Peluang kemunculan setiap kata pada label negative

$$p(\text{terimakasih, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{lazada, negatif}) = (2 + 1) / (9 + 18) = 3 / 27$$

$$p(\text{barang, negatif}) = (2 + 1) / (9 + 18) = 3 / 27$$

$$p(\text{pesan, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{sampai, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{selamat, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{kemas, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{bagus, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{ngeri, negatif}) = (1 + 1) / (9 + 18) = 2 / 27$$

$$p(\text{belanja, negatif}) = (1 + 1) / (9 + 18) = 2 / 27$$

$$p(\text{palsu, negatif}) = (1 + 1) / (9 + 18) = 2 / 27$$

$$p(\text{suka, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{lengkap, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{kualitas, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{harga, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{murah, negatif}) = (0 + 1) / (9 + 18) = 1 / 27$$

$$p(\text{jelek, negatif}) = (1 + 1) / (9 + 18) = 2 / 27$$

$$p(\text{pasar, negatif}) = (1 + 1) / (9 + 18) = 2 / 27$$



## 6. Menentukan $V_{map}$ untuk data uji

$$\frac{1}{2}$$

$$\frac{1}{10}$$

$$V_{map} = P(c_i) \times P(w_{kj}, c_i)$$

a.  $V_{map}$  untuk sentiment positif

$$p(\text{positif}) \times (p(\text{kapok}, \text{positif}) \times p(\text{belanja}, \text{positif}) \times p(\text{lazada}, \text{positif}) \times$$

$$p(\text{uang}, \text{positif}) \times p(\text{barang}, \text{positif}) \times p(\text{ kirim}, \text{positif}))$$

$$= 2/4 \times (1/35 \times 2/35 \times 3/35 \times 1/35 \times 3/35 \times 1/35)$$

$$= 0.000000000489 = \frac{36}{7.353.062.500}$$

b.  $V_{map}$  untuk sentiment negative

$$p(\text{negatif}) \times (p(\text{kapok}, \text{negatif}) \times p(\text{belanja}, \text{negatif}) \times p(\text{lazada}, \text{negatif})$$

$$\times p(\text{uang}, \text{negatif}) \times p(\text{barang}, \text{negatif}) \times p(\text{ kirim}, \text{negatif}))$$

$$= 2/4 \times (1/27 \times 2/27 \times 3/27 \times 1/27 \times 3/27 \times 1/27)$$

$$= 0.00000000232 = \frac{36}{1.549.681.956}$$

$$P(\text{Kapok}, \text{positif}) = \frac{0+1}{35} = 1/35$$

$$P(\text{uang}, \text{positif}) = \frac{0+1}{35} = 1/35$$

$$P(\text{Kirim}, \text{positif}) = \frac{0+1}{35} = 1/35$$

$$P(\text{Kapok}, \text{negatif}) = \frac{0+1}{27} = 1/27$$

$$P(\text{uang}, \text{negatif}) = \frac{0+1}{27} = 1/27$$

$$P(\text{Kirim}, \text{negatif}) = \frac{0+1}{27} = 1/27$$

## 6. Klasifikasi Data Uji

$$V_{map}^- = P(\text{negatif}) \times P(\text{Bagus, negatif}) \times P(\text{Lazada, negatif}) \times P(\text{Bagus, negatif}) \\ \times P(\text{packing, negatif}) \times P(\text{rapi, negatif}) \\ = \frac{1}{2} \times \frac{3}{27} \times \frac{3}{27} \times \frac{1}{27} \times \frac{1}{27} \times \frac{1}{27}$$

Karena  $V_{map}(\text{negative}) > V_{map}(\text{positif})$  maka data uji tersebut diklasifikasikan ke dalam sentiment negative.

$D_6$ : Barang di Lazada Bagus tetapi packingnya tidak

$V_{map} \text{ positif}$  rapi

$$P(\text{packing, positif}) = \frac{0+1}{35} = \frac{1}{35}$$

$$P(\text{rapi, positif}) = \frac{0+1}{28} = \frac{1}{28}$$

$V_{map} \text{ negatif}$

$$P(\text{packing, negatif}) = \frac{0+1}{27} = \frac{1}{27}$$

$$P(\text{rapi, negatif}) = \frac{0+1}{27} = \frac{1}{27}$$

$$V_{map}^+ = P(\text{positif}) \times P(\text{Bagus, positif}) \times P(\text{Lazada, positif}) \\ \times P(\text{Bagus, positif}) \times P(\text{packing, positif}) \times P(\text{rapi, positif}) \\ = \frac{1}{2} \times \frac{3}{35} \times \frac{3}{35} \times \frac{3}{35} \times \frac{1}{35} \times \frac{1}{35} \\ = \frac{9}{105.043.750}$$

karena  $V_{map}^- > V_{map}^+$  maka  $D_6$  termasuk ke sentiment negatif