

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Analisis Sentimen**

Analisis sentimen atau *opinion mining* adalah deteksi sikap-sikap terhadap objek atau orang [beineke et al. 2014]. Analisis sentimen dapat digunakan untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, institusi, produk atau pada sebuah kondisi tertentu. Nilai dari analisis sentimen bisa dipecah menjadi 3 yakni, sentimen positif, sentimen negatif dan sentimen netral atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif [2].

Analisis sentimen bertujuan untuk melakukan penilaian terhadap emosi, sikap, pendapat, evaluasi yang disampaikan oleh seseorang pembicara atau penulis terhadap sebuah produk atau terhadap tokoh masyarakat. Alasan tersebut menyebabkan beberapa penelitian terutama pada *review* produk didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses *opinion mining*.

#### **2.2 Twitter**

Twitter merupakan salah satu media sosial yang memungkinkan penggunanya untuk mengirim dan membaca pesan *tweet* berupa teks, gambar atau sebuah video. Media sosial twitter berbeda dengan media sosial lainnya terutama dalam penulisan status atau *tweet*. Media sosial selain twitter tidak ada batasan karakter yang dapat dituliskan sedangkan twitter hanya memberikan 280 karakter yang dapat ditulis sebagai status atau cuitan.

Twitter bersifat publik, maksudnya adalah semua yang dituliskan atau dibagikan dapat dilihat oleh semua pengguna lainnya, namun pengguna twitter dapat membatasi pengiriman *tweet* hanya bisa dilihat temannya saja atau biasa disebut sebagai *follower*. Twitter memiliki fitur utama yakni dapat menuliskan status atau cuitan serta dapat melakukan pengiriman pesan kepada pengguna lain, fitur lain dari sosial media twitter sebagai berikut :

### 1. Following

Salah satu fitur andalan dari media sosial twitter yakni following. Fitur ini memungkinkan pengguna untuk saling terhubung dengan pengguna lain atau bisa disebut sebagai pertemanan. Setiap unggahan tweet dari pengguna yang telah di follow maka dapat dilihat di beranda pengguna yang telah mengikutinya.

### 2. Retweet

Fitur retweet merupakan fitur yang memudahkan pengguna untuk meneruskan atau menyebarkan tweet pengguna lain sehingga dapat muncul di beranda pribadi

### 3. Hashtag

Hashtag atau tagar merupakan fitur dari twitter yang dapat mengelompokkan sebuah tweet. Dimana setiap tweet yang ditulis bisa ditambahkan dengan hashtag berupa kata atau keyword dari tweet tersebut. Salah satu fungsi dari adanya hashtag yakni untuk mengelompokkan atau memudahkan pencarian terhadap kata kunci dari tweet.

### 4. Trending Topic

Fitur trending topik merupakan fitur yang menampilkan topik atau hashtag yang sedang populer atau banyak dibahas oleh pengguna tweet. Adanya trending topic membuat pengguna mengetahui hal apa saja yang sedang viral di kalangan masyarakat.

## **2.3 Tokoh Politik**

Tokoh politik atau *political actor* adalah seorang yang karena latar belakang sejarahnya, sepak terjangnya dalam perjuangan dan idealismenya, dikenal oleh masyarakat sehingga segala pendapatnya atau pemikirannya maupun perbuatannya diikuti. Tokoh-tokoh politik mempunyai kemampuan-kemampuan karismatis, oratorik yang bersifat mempersatukan, menegahi dan dapat memanipulasikan simbol-simbol sehingga dapat mengendalikan massa.

## 2.4 Sistem Temu Kembali Informasi

### 2.4.1 Preprocessing

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*.

#### 2.4.1.1 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.

**Tabel 2. 1** Contoh proses *case folding*

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

#### 2.4.1.2 Tokenizing

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

**Tabel 2. 2** Contoh proses *tokenizing*

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

#### 2.4.1.3 Stopword removal

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

**Tabel 2. 3** Contoh proses *stopword removal*

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

Teks <i>output</i>	kapok, belanja, lazada, uang, barang, dikirim
--------------------	---

#### 2.4.1.4 Stemming

Stemming merupakan proses pengubahan kata menjadi kata dasar yang sesuai dengan aturan bahasa Indonesia. Proses stemming menggunakan bantuan algoritma Nazief dan Adriani.

**Tabel 2. 4** Contoh proses *stemming*

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

#### 2.4.2 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

Empas	Empat	Entar	Entah	Falak	Fakultas
-------	-------	-------	-------	-------	----------

### 2.5 Naïve Bayes Classifier

Klasifikasi *bayes* yang ditemukan oleh Thomas Bayes paada 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].

Berikut contoh langkah-langkah perhitungan menggunakan *naive bayes classifier* :

**Tabel 2. 8** Contoh data latih

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

#### 1. Menentukan data uji

**Tabel 2. 9** Data uji

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

#### 2. Menentukan probabilitas kategori dari data latih

$$p(ci) = \frac{\text{Jumlah dokumen dengan label } ci}{\text{Keseluruhan dokumen}} \dots\dots\dots (2.1)$$

$$p(\text{positif}) = 2/4$$

$$p(\text{negatif}) = 2/4$$

3. Menghitung probabilitas kosakata (*term frequency*) dari data latih**Tabel 2. 10** Frekuensi kemunculan kata

Kata	D1	D2	D3	D4
Terimakasih	1	0	0	0
Lazada	1	1	1	1
Barang	2	1	0	1
Pesan	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	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. Menghitung probabilitas untuk setiap kelas data latih

$$p(wkj.ci) = \frac{Pk+1}{n+term} \dots\dots\dots (2.2)$$

Dimana,

$p(wkj.ci)$  = probabilitas term

$pk + 1$  = fitur kemunculan term  $wkj + 1$ , penambahan 1 dilakukan agar hasil perhitungan tidak 0 (nol)

$n$  = semua jumlah kata yang ada pada topik  $ci$

$term$  = jumlah kata keseluruhan

a. Probabilitas untuk sentimen 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. Probabilitas untuk sentimen negatif

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5. Menentukan Vmap untuk data uji

$$Vmap = p(ci) \times p(wkj. ci) \dots\dots\dots (2.3)$$

a. Vmap untuk sentimen positif

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

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

$$= 0.000000000489$$

b. Vmap untuk sentiment negatif

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

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

$$= 0.00000000232$$

6. Klasifikasi data uji

Jadi dari perhitungan diatas data uji tersebut termasuk dalam kategori kelas negatif (-) karena nilai Vmap Negatif lebih besar (>) dari Vmap Positif

## 2.6 Pengujian Klasifikasi

Sebuah proses klasifikasi memerlukan proses pengujian mengenai hasil dari klasifikasi. Hal itu perlu dilakukan untuk mendapatkan akurasi dari perhitungan yang dilakukan. Proses pengujian klasifikasi menggunakan *matriks confusion* dimana penjelasan *matriks confusion* seperti pada tabel berikut [5].

**Tabel 2. 11** Matriks confusion

Keterangan	Relevan	Tidak Relevan
Terambil	True Positif (TP)	False Positif (FP)

Tidak Terambil	False Negatif (FN)	True Negatif (TN)
----------------	--------------------	-------------------

Dimana,

True Positif (TP) = teridentifikasi secara benar

False Positif (FP) = teridentifikasi secara salah

False Negatif = tertolak secara benar

True Negatif = tertolak secara salah

Rumus untuk menghitung akurasi seperti pada rumus 2.4 sebagai berikut.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} \text{-----} (2.4) \\
 &= \frac{TP + TN}{TP + FP + FN + TN}
 \end{aligned}$$

Sebuah algoritma klasifikasi berusaha untuk membentuk model yang mempunyai nilai akurasi yang tinggi. Umumnya model yang dibangun dapat memprediksi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji barulah kinerja model dari sebuah algoritma klasifikasi ditentukan.

## 2.7 Normalisasi data

Nilai-nilai atribut data yang berbeda rentangnya seringkali perlu dilakukan normalisasi atau disandarisasikan agar proses data mining tidak menjadi bisa. Normalisasi data dapat dilakukan ke dalam rentang angka yang kecil seperti rentang [0, 1] maksudnya nilai hasil normalisasi berada diantara angka 0 (nol) sampai satu (1). Sehingga semua atribut akan memiliki bobot yang sama [5].

$$X_{ik} = \frac{X_{ik} - \min(X_k)}{\max(X_k) - \min(X_k)} \text{-----} (2.4)$$

Dimana

$X_{ik}$  = nilai asl

$\min(X_k)$  = nilai minimal (terkecil) dari kelompok k

$\max(X_k)$  = nilai maksimal (terbesar) dari kelompok k

## 2.8 Simple Additive Weighting

Menurut Fishburn dan MacCrimmon dalam (Munthe, 2013) mengemukakan bahwa Metode Simple Additive Weight (SAW), sering juga dikenal dengan istilah metode penjumlahan terbobot. Konsep dasar metode Simple Additive Weight

(SAW) adalah mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif pada semua atribut.

Menurut Fishburn dan MacCrimmon Ada beberapa langkah dalam penyelesaian metode Simple Additive Weight (SAW) adalah sebagai berikut :

1. Menentukan kriteria-kriteria yang dijadikan acuan dalam pendukung keputusan yaitu  $C_i$ .
2. Menentukan bobot preferensi atau tingkat kepentingan pada kriteria

$$\text{Preferensi kategori} = \frac{\text{Jumlah data pada kategori}}{\text{jumlah semua data}} \dots\dots\dots (2.5)$$

3. Menentukan rating kecocokan setiap alternatif pada setiap kriteria.
4. Membuat matriks keputusan berdasarkan kriteria ( $C_i$ ).
5. Kemudian melakukan normalisasi matriks berdasarkan persamaan yang disesuaikan dengan jenis atribut sehingga diperoleh matriks ternormalisasi  $R$ . Normalisasi menggunakan normalisasi linear jangkauan  $[0, 1]$  rumus (2.4)
6. Hasil akhir diperoleh dari proses perankingan yaitu penjumlahan dari perkalian matriks ternormalisasi  $R$  dengan vector bobot.

$$v_i = \sum_{j=0}^n w_j r_{ij} \dots\dots\dots (2.6)$$

sehingga diperoleh nilai terbesar yang dipilih sebagai alternatif terbaik ( $A_i$ ) sebagai solusi

## 2.9 Penelitian Sebelumnya

1. ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER. Pada tahun 2013 Ahmad Fathan Hidayatullah dari fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada telah melakukan penelitian sebagai tugas akhir. Penelitian tersebut difungsikan untuk mengetahui nilai sentimen pada tokoh publik dengan metode *naive bayes* dan *support vector machine* dengan akurasi untuk *naive bayes* 73.81% dan akurasi *support vector machine* sebesar 83.14% [6].