

## Analisis Sentimen pelanggan Riview Shopee di Twitter

Shopee adalah salah satu perusahaan retail berbasis online yang banyak digunakan oleh masyarakat baik di Indonesia maupun luar negeri. Menawarkan berbagai produk seperti makanan, barang perabotan rumah tangga, buku, jasa gambar, jasa penjualan tiket dan banyak lainnya. Untuk dapat terus berkembang maka diperlukannya sebuah feedback dari pelanggan tentang tingkat kepuasan berdasarkan respon positif atau negatif. Untuk mengetahui tingkat kepuasan pelanggan yaitu menggunakan sebuah analisis sentiment untuk mengevaluasi kebutuhan atau produk dari pelanggan, mengetahui kekurangan perusahaan sehingga dapat menjadi cara untuk perbaikan dan meningkatkan pendapatan suatu perusahaan. Untuk mengetahui pendapat-pendapat masyarakat tentang produk Shopee, Shopee membentuk grup di Twitter yang dapat digunakan untuk mengetahui tanggapan dari pelanggan. Dimana Twitter adalah salah satu media sosial yang cukup populer dan sangat mudah digunakan oleh semua kalangan usia. Untuk melakukan analisis, metode yang digunakan yaitu Cross Industry Process for Data Mining (CRISP-MD).

### 1. Business Understanding

Tujuan dari proses ini yaitu untuk mengetahui tingkat kepuasan pelanggan berdasarkan pendapat positif maupun negatif tentang produk yang ditawarkan Shopee di Twitter serta menggunakan Naïve Bayes untuk melihat tingkat akurasi.

### 2. Data Understanding

Data yang diperlukan yaitu diambil melalui percakapan di Twitter, mulai dari bulan Januari sampai dengan bulan Mei 2022. Percakapan dibagi menjadi data uji dan data latih yang terdiri atas komentar positif dan negatif.

### 3. Data Preparation

Data yang telah terkumpul dari bulan Januari hingga Mei 2022 dilakukan pembersihan. Data diambil dari 600 data dengan perincian 300 data uji yaitu komentar positif 150 dan komentar negatif 150 dan data uji sebesar 300 data dengan pembagian antara komentar positif dan negatif sama. Data diambil dengan aplikasi Rstudio

### 4. Modelling

Data yang sudah dibersihkan kemudian dimasukkan ke algoritma Naive Bayes untuk dilakukan perhitungan tingkat keakurasian. Tahapannya yaitu sebagai berikut.

#### a. Naïve Bayes

Algoritma Naive Bayes digunakan untuk menghitung probabilitas maupun statistik.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Figure 1 Rumus Naive Bayes

Keterangan :

X : Data yang belum diketahui

H : Hipotesis dari data class

P(H|X) : Probabilitas hipotesis nilai berdasarkan kondisi nilai X

P(H) : Probabilitas hipotesis nilai H

P(X|H) : Probabilitas nilai X berdasarkan hipotesis nilai H.

P(X) : Probabilitas nilai.

b. *Convusion Matrix*

*Couvusion Matrix* digunakan untuk mengetahui nilai berdasarkan tingkat keakurasian. *Couvusion Matrix* memiliki 4 nilai yaitu sebagai berikut :

		Actual Values	
		1 (Postive)	0 (Negative)
Predicted Values	1 (Postive)	<b>TP</b> (True Positive)	<b>FP</b> (False Positive) Type I Error
	0 (Negative)	<b>FN</b> (False Negative) Type II Error	<b>TN</b> (True Negative)

Gambar 2. Tabel *Convulsion Matriks*

Keterangan :

TP : Pernyataan positif yang benar-benar terjadi

TN : Pernyataan negatif yang benar-benar terjadi

FP : Suatu prediksi positif, tetapi kebenarannya bersifat negatif

FN : Suatu prediksi negatif, tetapi kebenarannya bersifat positif

Cara Penghitungannya :

$$Accuracy = ( TP + TN ) / ( TP + FP + FN + TN )$$

c. Tampilan Rstudio

Dimana Rstudio digunakan untuk mengambil komentar-komentar dari Twitter dengan menggunakan rumus koding serta memasukkan consumer key dan consumer sehingga rumus dapat dijalankan.

```

1 setwd("C:/Users/Vijayan/Desktop/scraps")
2
3 library(tw)
4 library(twitteR)
5 library(copys)
6 library(httr)
7 library(bas44enc)
8
9 #download file from http://url.kiss.se/CA/secret.png destfile = "cakeys.png"
10
11 # Secret dengan key with key
12 consumer_key <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
13 consumer_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
14 access_token <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
15 access_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
16
17 setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
18 tw <- searchTwitter("shopware", n = 100, retryonrate.limit = 100)
19 #tw <- searchTwitter("KRT hanggala 402", n = 200, retryonrate.limit = 100)
20 view(tw)
  
```

Gambar 3. Tampilan *RStudio*

d. Hasil *Crowling Data*

Pengambilan data masih tercampur belum dilakukan cleaning



tw	list [1000]	List of length 1000
[1]	S4 [1] (tweetEntity)	
[2]	S4 [1] (tweetEntity)	
[3]	S4 [1] (tweetEntity)	
[4]	S4 [1] (tweetEntity)	
[5]	S4 [1] (tweetEntity)	
[6]	S4 [1] (tweetEntity)	
[7]	S4 [1] (tweetEntity)	
[8]	S4 [1] (tweetEntity)	
[9]	S4 [1] (tweetEntity)	
[10]	S4 [1] (tweetEntity)	
[11]	S4 [1] (tweetEntity)	

Gambar 4. *Crawling Data*

#### e. Data Frame

Data Frame yaitu berisi percakapan Twitter yang berada dalam tahap campur antara komentar negatif maupun positif, belum dilakukan pemisahan.



	text
1	semangat yaaa kak good luck hi
2	ashia mimin syatet deh nih kak cuss ntengin terus stingan a...
3	shia kakcp
4	wihh mantul banget nih kak ah
5	ntas deh kak mantul bangettt yukkk stay disini aja yaww jan...
6	min cek dm dong
7	aamin kak semangat trus ya buat ikutan kontisinya goodlu...
8	aamin makasih min
9	hi kak waah jangan simis diungg hehe siaa buat kejutan ha...
10	jangan simis dong kak harus usaha teruss ah
11	aamin yah kak semangattttt av

Gambar 5. Hasil Data Frame

#### f. Program Naïve Bayes

Ini merupakan proses dimana data-data yang didapat di Twitter bisa digunakan untuk perhitungan tingkat keakurasiann. Semua data komentar positif dan negative kemudian dimasukkan dalam program dan dihitung dengan rumus Convusion Metrix.



Gambar 6. Tampilan Program *Naive Bayes*

$$\begin{aligned}
 \text{Accuracy} &= (TP+TN) / (TP+FP+FN+TN) \\
 &= (15+14) / (15+1+0+14) \\
 &= (29 / 30) * 100 \\
 &= 97\%
 \end{aligned}$$

#### 5. Evaluasi

Hasil penerapan menggunakan algoritma Naïve Bayes sudah sesuai dengan yang diharapkan yaitu mengecek tingkat keakurasian analisis sentiment.

## 6. Deployment

Penerapan algoritma Naïve Bayes untuk melakukan akurasi data analisis sentiment sudah dapat di terapkan karena hasilnya sama seperti yang diharapkan pada tahap business understanding.

Demikianlah yang bisa saya sampaikan mengenai Analisis Sentiment Pelanggan Review Shopee di Twitter, bahwa dalam metode CRISP-MD menggunakan algoritma Naïve Bayes dapat digunakan untuk menghitung Tingkat keakurasian berdasarkan data- data komentar positif dan negative yang di dapat di Twitter dan dapat dilakukan pemisahan antara data yang membawa komentar positif dan yang membawa komentar negative. Dengan melakukan analisis sentiment maka akan berpengaruh kepada perusahaan seperti mengetahui kekurangan pada bisnis, dapat menarik pelanggan baru dan mengetahui kepuasan yang dirasakan oleh pelanggan.

## Referensi

Ruger AH, Analisis S, Shopee P, Suyanto M, Kurniawan MP. Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma Naive Bayes. JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY. 2021;1(2).

Firmansyah F, Yulianto A. Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study : Gramedia Matraman Jakarta. JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING. 2021 Jan 18;4(2):383–92.

## Github Repository Link :

<https://github.com/Nuvisa09/analisis-sentimen.git>