

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/362218813>

# TWITTER SENTIMENT ANALYSIS TERHADAP PENGGUNA E-COMMERCE MENGGUNAKAN TEXT MINING

Conference Paper · November 2021

DOI: 10.53026/sntem.v1i2.592

CITATION

1

READS

73

4 authors:



**Arif Rahman Hakim**

Universitas Medika Suherman

7 PUBLICATIONS 5 CITATIONS

SEE PROFILE



**Dewi Marini Umi Atmaja**

Universitas Medika Suherman

9 PUBLICATIONS 15 CITATIONS

SEE PROFILE



**Deny Haryadi**

Institut Teknologi Telkom Jakarta

13 PUBLICATIONS 29 CITATIONS

SEE PROFILE



**Niko Suwaryo**

4 PUBLICATIONS 4 CITATIONS

SEE PROFILE

## **TWITTER SENTIMENT ANALYSIS TERHADAP PENGGUNA E-COMMERCE MENGGUNAKAN TEXT MINING**

**Arif Rahman Hakim<sup>1</sup>, Dewi Marini Umi Atmaja<sup>1</sup>, Deny Haryadi<sup>2</sup> Niko Suwaryo<sup>1</sup>**

<sup>1</sup>Bisnis Digital, Universitas Medika Suherman, Jl.Raya Industri Pasir Gombong, Cikarang-Bekasi 17530

<sup>2</sup>Teknologi Informasi, Institut Teknologi Telkom Jakarta, Jl. Daan Mogot, Cengkareng, Jakarta Barat 11710

\*E-mail: [arif@medikasuherman.ac.id](mailto:arif@medikasuherman.ac.id), [dewi@medikasuherman.ac.id](mailto:dewi@medikasuherman.ac.id),  
[denyharyadi@ittelkom-jkt.ac.id](mailto:denyharyadi@ittelkom-jkt.ac.id), [niko@medikasuherman.ac.id](mailto:niko@medikasuherman.ac.id).

### **ABSTRAK**

*Tokopedia dan Bukalapak merupakan perusahaan yang mengusung bisnis marketplace dan mall online. Untuk menanggapi pendapat, kritik, saran dan masalah komplain, Tokopedia dan Bukalapak mempunyai akun khusus pada Twitter yang diberi nama @tokopediacare dan @bukabantuan. Jenis komentar tersebut berbentuk unstructured text dalam jumlah besar. Kondisi ini dapat mengakibatkan perusahaan belanja online dapat melewatkan informasi yang berguna dari sekumpulan dokumen teks. Oleh karena itu, dilakukan analisis text mining dengan membandingkan metode klasifikasi menggunakan algoritma Support Vector Machine (SVM) dan algoritma Naive Bayes. Berdasarkan hasil analisis dapat diketahui bahwa metode klasifikasi menggunakan algoritma Support Vector Machine (SVM) lebih baik digunakan jika dibandingkan dengan algoritma Naive Bayes dengan hasil ketepatan klasifikasi mencapai 55,92% pada akun Tokopediacare dan 53,04% akun Bukabantuan. Selain itu untuk mengetahui kata-kata yang paling sering muncul dalam tanggapan positif, tanggapan negatif maupun tanggapan netral pada masing-masing e-commerce dilakukan pula visualisasi kata menggunakan word cloud.*

**Kata kunci:** *twitter, text mining, machine learning*

### **1. PENDAHULUAN**

Sebuah survei yang diselenggarakan Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) menunjukkan bahwa jumlah pengguna internet di Indonesia meningkat tiap tahunnya hingga akhir tahun 2018 mencapai 171,17 juta orang, dari total penduduk sebanyak 264,16 juta orang. Perkembangan tersebut berdampak pada berbagai bidang. Salah satunya ialah maraknya kegiatan berbelanja melalui media internet. Tokopedia dan Bukalapak mempunyai sebuah akun *Twitter costumer care* yaitu @tokopediacare untuk Tokopedia dan @bukabantuan untuk Bukalapak. Akun tersebut adalah salah satu bentuk layanan pelanggan khusus melalui *online* yang disediakan untuk menanggapi tanggapan, pendapat, kritik, saran dan masalah komplain. Komentar dalam *twitter* berbentuk teks, sehingga perlu dilakukan analisis *text mining*. Tujuan dari *text mining* yaitu untuk memperoleh informasi dari sekumpulan dokumen serta mendukung proses *knowledge discovery* pada koleksi dokumen yang besar, sumber data yang digunakan pada *text mining* merupakan kumpulan teks dengan format yang tidak teratur atau minimal semi teratur [1].

Konsumen dari belanja online dapat memberikan *feedback* secara terbuka, saling berkomentar dalam waktu cepat dan tidak terbatas terhadap pelayanan dari Tokopedia dan Bukalapak melalui *twitter costumer care*. Mengetahui *sentiment* dari konsumen belanja *online* secara manual dapat merugikan waktu, dan tenaga. Oleh karena itu, pada penelitian ini dilakukan analisis *text mining* dengan membandingkan metode klasifikasi *Support Vector Machine* dan *Naive Bayes*. Informasi yang didapatkan dari *twitter* tidak dapat menggambarkan struktur komunikasi dan tingkat partisipasi dari setiap pelanggan. Oleh karena itu diperlukan suatu metode yang dapat memvisualisasikan atau memeriksa pola interaksi pengguna Tokopedia dan Bukalapak menggunakan *Word*. tujuan yang akan dicapai dalam penelitian ini adalah mengetahui metode klasifikasi terbaik antara *Support Vector*

*Machine* dan *Naive Bayes*. Selain itu, juga ingin mengetahui kata-kata yang sering muncul pada masing-masing sentimen menggunakan visualisasi *word cloud*. Manfaat yang dapat diperoleh dari penelitian ini adalah mampu mengetahui proses mengekstrak dan mengolah data teks secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini dengan hasil metode klasifikasi terbaik. Selain itu juga dapat mengidentifikasi pola pengguna *twitter* konsumen Tokopedia dan Bukalapak sehingga dapat dimanfaatkan lebih maksimal untuk menyebarkan informasi secara lebih efektif. Dengan diketahuinya hasil visualisasi kata berdasarkan hasil tampilan *word cloud* maka dapat diketahui kata apa saja yang paling sering muncul dalam tanggapan positif negatif, maupun tanggapan netral [2].

## 2. METODE

### A. Data Training dan Data Testing

Data set dalam penelitian ini adalah tanggapan mengenai “Tokopedia” dan “Bukalapak” pada media sosial *twitter* dari tanggal 10 Agustus 2019 sampai 26 Agustus 2019. Data tanggapan yang diperoleh sebanyak 6748 tanggapan untuk Tokopedia, dan 4491 tanggapan untuk Bukalapak. Penentuan jumlah data *training* dan testing menggunakan perbandingan rasio dengan jumlah masing-masing data sebanyak 1.000 data tanggapan positif, 1.000 data tanggapan negatif dan 1.000 data tanggapan netral untuk Tokopedia, sehingga total data tanggapan yang digunakan pada Tokopedia sebesar 3000 data tanggapan sedangkan untuk Bukalapak sebanyak 3000 data tanggapan dan untuk kategori tanggapan positif, negatif maupun netral masing-masing sebesar 1.000 data tanggapan. Pada penelitian ini untuk seluruh data dibagi ke dalam 5 kombinasi perbandingan rasio *data training* dan *data testing* seperti pada tabel dibawah ini.

**Tabel 1. Jumlah Sampel (Data Training) untuk Tokopedia dan Bukalapak**

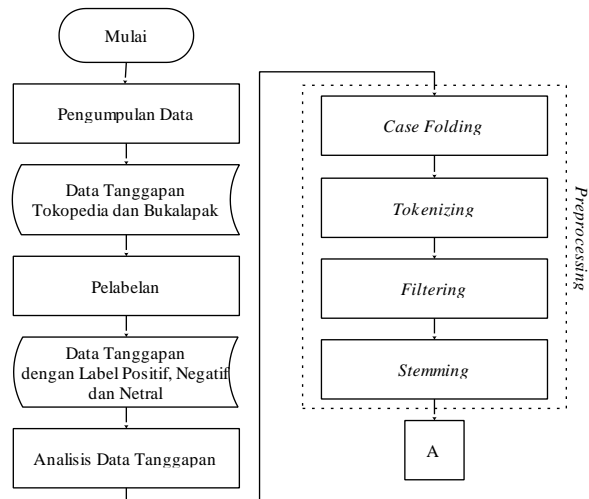
No.	Jumlah Data	Rasio (%) Training : Testing	Data Training	Data Testing
1	3000	20:80	600	2400
2		40:60	1200	1800
3		60:40	1800	1400
4		80:20	2400	600
5		90:10	2700	300

### B. Metode Analisis

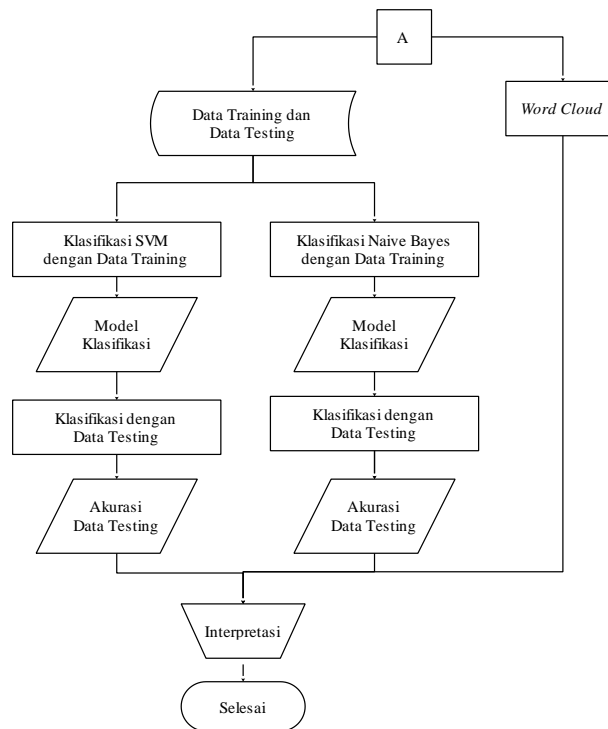
Penelitian ini menggunakan metode klasifikasi dengan algoritma *Support Vector Machine* (SVM) dan *Naive Bayes* dalam pengklasifikasian data tanggapan, serta *Word Cloud* untuk menampilkan kata yang sering muncul atau sering dibahas. *Tools* analisis yang digunakan adalah *Microsoft Microsoft Excel 2016* dan *R 3.6.0*.

### C. Tahapan Penelitian

*Flowchart* yang menunjukkan tahapan penelitian dari mulai pengumpulan data hingga diperoleh kesimpulan ditunjukkan pada gambar dibawah ini.



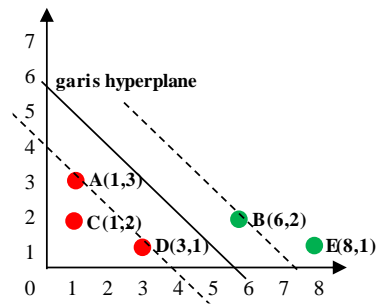
**Gambar 1 Tahapan Penelitian**



**Gambar 2 Tahapan Penelitian (lanjutan)**

#### D. Simulasi Perhitungan

1. Simulasi *Support Vector Machine* pada Data Terpisah Secara Linear  
Prinsip kerja *Support Vector Machine* yaitu mencoba menemukan *hyperplane* terbaik [3].



**Gambar 3. Grafik Hyperplane**

Berdasarkan gambar diatas dapat disajikan dalam bentuk tabel berikut.

**Tabel 2. Data Simulasi Support Vector Machine Linear**

Titik	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
A	1	3	-1
B	6	2	1
C	1	2	-1
D	3	1	-1
E	8	1	1

Formulasi titik minimal fungsi tujuan  $\frac{1}{2} \|w\|^2$  dengan memperhatikan pembatas, berdasarkan diatas diperoleh persamaan *constraint* (pembatas) sebagai berikut.

$$-w_1 - 3w_2 - b \geq 1 \quad (1)$$

$$6w_1 + 2w_2 + b \geq 1 \quad (2)$$

$$-w_1 - 2w_2 - b \geq 1 \quad (3)$$

$$-3w_1 - w_2 - b \geq 1 \quad (4)$$

$$8w_1 + w_2 + b \geq 1 \quad (5)$$

Eliminasi Persamaan 1 dan Persamaan 2 diperoleh.

$$-w_1 - 3w_2 - b = 1 \quad (1)$$

$$6w_1 + 2w_2 + b = 1 \quad (2)$$

$$\begin{array}{rcl} 5w_1 - w_2 & = & 2 \quad + \\ w_1 & = & 0,4 + 0,2w_2 \end{array} \quad (6)$$

Eliminasi Persamaan 3 dan Persamaan 4, serta substitusi Persamaan 5 diperoleh.

$$-w_1 - 2w_2 - b = 1 \quad (3)$$

$$-3w_1 - w_2 - b = 1 \quad (4)$$

$$2w_1 - w_2 = 0$$

$$2 * (0,4 + 0,2w_2) - w_2 = 0$$

$$0,8 + 0,4 - w_2 = 0$$

$$w_2 = 1,33$$

$$w_2 = 1,33, \text{ maka } w_1 = 0,4 + 0,2w_2 = 0,66$$

Substitusi nilai  $w_1$  dan  $w_2$  kedalam Persamaan 5 diperoleh.

$$8w_1 + w_2 + b = 1 \quad (5)$$

$$8 * 0,66 + 1,33 + b = 1$$

$$5,28 + 1,33 + b = 1$$

$$b = -5,61$$

Berdasarkan nilai  $w_1$ ,  $w_2$ , dan  $b$  yang diperoleh, maka persamaan fungsi pemisahannya yaitu:

$$f(x) = 0,66x_1 + 1,33x_2 - 5,61 \quad (7)$$

Klasifikasi untuk data selanjutnya dapat dilakukan dengan menggunakan fungsi pada Persamaan 7, dengan fungsi *sgn* tersebut semua nilai  $f(x) < 0$  masuk ke kelas -1 (negatif) dan nilai  $f(x) > 0$  masuk ke dalam kelas +1 (positif) [4].

## 2. Simulasi *Naive Bayes*:

Misalkan terdapat empat buah dokumen yang telah melalui tahapan *preprocessing*, dua dokumen diambil dari kelas positif dan dua dokumen diambil dari kelas negatif. Dokumen tersebut adalah sebagai berikut.

Dokumen 1 : beli voucher hemat murah situs

Dokumen 2 : beli voucher periode promo agustus

Dokumen 3 : server voucher habis

Dokumen 4 : habis beli voucher buka twitter

Selanjutnya akan dihitung frekuensi kemunculan kata pada setiap dokumen seperti pada tabel berikut.

**Tabel 3. Frekuensi Kemunculan Kata**

Query	Dokumen			
	1	2	3	4
beli	1	1	0	1
voucher	1	1	1	1
hemat	1	0	0	0
murah	1	0	0	0
situs	1	0	0	0
periode	0	1	0	0
promo	0	1	0	0
agustus	0	1	0	0
server	0	0	1	0
habis	0	0	1	1
buka	0	0	0	1
twitter	0	0	0	1
Kelas	Positif	Positif	Negatif	Negatif

Berdasarkan tabel diatas bahwa diketahui bahwa kelas positif terdiri dari 2 dokumen dengan jumlah kata sebanyak 10 kata dari 12 kosakata yang ada, sedangkan 2 dokumen kelas negatif terdiri dari 8 kata dari 12 kosakata yang ada. Berdasarkan jumlah kata tersebut, dapat dihitung nilai probabilitas untuk setiap kelasnya dengan menggunakan Persamaan berikut.

Probabilitas kata kelas positif

Contoh perhitungan probabilitas untuk kata “beli” yang terdapat dalam kelas positif.

$$P a_i v_j = \frac{n_i + 1}{n + \text{kosakata}} \quad (8)$$

$$P \text{ beli positif} = \frac{2 + 1}{10 + 12} = 0,136$$

dimana,

$n_i$ : jumlah kata “beli” dalam dokumen kelas positif

$n$ : jumlah seluruh kata pada dokumen kelas positif

$\text{kosakata}$  : lah kata dalam fase *training*

Nilai probabilitas untuk kata yang lain, dengan menggunakan cara perhitungan yang sama ditunjukkan oleh tabel berikut.

**Tabel 4. Probabilitas Kata Kelas Positif**

Query	Probabilitas	Query	Probabilitas
beli	0,136	promo	0,091
voucher	0,136	agustus	0,091
hemat	0,091	server	0,045
murah	0,091	habis	0,045
situs	0,091	buka	0,045
periode	0,091	twitter	0,045

### Probabilitas Kata Kelas Negatif

Contoh perhitungan probabilitas kata “beli” negatif.

$$P a_i v_j = \frac{n_i + 1}{n + \text{kosakata}} \quad (8)$$

$$P \text{ beli negatif} = \frac{1 + 1}{8 + 12} = 0,1$$

dimana,

$n_i$ : jumlah kata “beli” dalam dokumen kelas negatif

$n$ : jumlah seluruh kata pada dokumen kelas negatif

kosakata : lah kata dalam fase training

Nilai probabilitas untuk kata yang lain dalam kelas negatif ditunjukkan oleh tabel dibawah ini.

**Tabel 5. Probabilitas Kata Kelas Negatif**

Query	Probabilitas	Query	Probabilitas
beli	0,10	promo	0,05
voucer	0,15	agustus	0,05
hemat	0,05	server	0,10
murah	0,05	habis	0,15
situs	0,05	buka	0,10
periode	0,05	twitter	0,10

Nilai probabilitas kata pada masing-masing kelas tersebut, kemudian disimpan pada database yang nantinya akan digunakan untuk menguji data baru [5]. Misal ingin diketahui kelas data dari tanggapan baru “voucer hemat promo murah”. Langkah pertama yang dilakukan untuk melakukan klasifikasi adalah memecah kalimat dalam tanggapan tersebut menjadi kata per kata, selanjutnya menghitung nilai probabilitas dari kata pada masing-masing kelas dengan menggunakan tabel probabilitas kata yang telah diperoleh sebelumnya, sedangkan probabilitas masing-masing kelas ditentukan dengan menggunakan Persamaan 22 [6].

$$P v_j = \frac{\text{doc}_j}{\text{training}} \quad (9)$$

$$P \text{ positif} = \frac{2}{4} = 0,5$$

$$P \text{ negatif} = \frac{2}{4} = 0,5$$

Probabilitas untuk tanggapan baru yang ingin diklasifikasikan ditentukan dengan menggunakan Persamaan 21. Nilai hasil perhitungan ditunjukan oleh tabel berikut.

**Tabel 6. Nilai Probabilitas Tanggapan Baru**

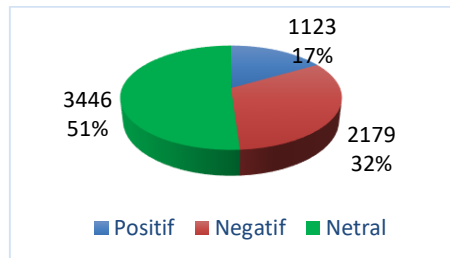
Kelas	Voucer	Hemat	Promo	Murah	Nilai Probabilitas
Positif (P = 0,5)	0,136	0,091	0,091	0,091	0,000051
Negatif (P = 0,5)	0,15	0,05	0,05	0,05	0,000009

Hasil klasifikasi dari kelas tanggapan baru tersebut adalah kelas atau kategori yang memiliki probabilitas tertinggi. Berdasarkan tabel 6 diketahui bahwa nilai probabilitas tertinggi adalah probabilitas kelas positif, maka tanggapan baru “voucer hemat promo murah” masuk ke dalam kelas positif.

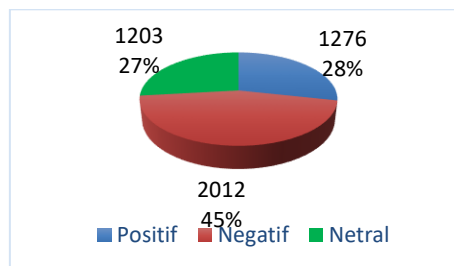
### 3. PEMBAHASAN

#### A. Analisis Data

Secara keseluruhan gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori untuk Tokopedia dan Bukalapak ditunjukkan oleh gambar berikut.



Gambar 4. Jumlah Tanggapan pada *e-commerce* Tokopedia



Gambar 5. Jumlah Tanggapan *e-commerce* Bukalapak

Berdasarkan Gambar 4 diketahui bahwa dari total tanggapan yang diperoleh dari Tokopedia yaitu 6748 tanggapan, sebanyak 1123 (17%) tanggapan merupakan tanggapan positif yaitu tanggapan yang mendukung atau menyukai Tokopedia sedangkan 2179 (32%) tanggapan merupakan tanggapan negatif yaitu tanggapan yang kurang menyukai suatu hal yang berhubungan dengan Tokopedia dan untuk tanggapan netral sebanyak 3446 (51%) tanggapan. Sedangkan Gambar 5 menunjukkan total tanggapan yang diperoleh dari Bukalapak yaitu 4491 tanggapan, sebanyak 1276 (28%) tanggapan merupakan tanggapan positif yaitu tanggapan yang mendukung atau menyukai Bukalapak sedangkan 2012 (45%) tanggapan merupakan tanggapan negatif yaitu tanggapan yang kurang menyukai suatu hal yang berhubungan dengan Bukalapak dan sisanya untuk tanggapan netral sebanyak 1203 (27%).

#### B. Perbandingan Algoritma

Berikut ini hasil *macro average recall*, *macro average precision* dan *average accuracy* dengan menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* pada 5 perbedaan rasio data yang telah ditentukan pada *e-commerce* Tokopedia dan Bukalapak. Metode nilai *macro average recall*, *macro average precision* dan *average accuracy* digunakan untuk mengetahui kinerja machine learning secara keseluruhan dari seluruh data set yang ada.

Tabel 7. *Macro Average Recall* dan *Precision* serta *Average Accuracy* Tokopedia

Nilai	SVM	Naive Bayes
<i>Macro Average Recall</i>	0.5545	0.4768
<i>Macro Average Precision</i>	0.5454	0.5009
<i>Average Accuracy</i>	0.5545	0.4768



**Tabel 8. Perbandingan Accuracy Tokopedia menggunakan 10-fold cross validation**

No	Rasio Perbandingan (%)	Accuracy	
		SVM	Naive Bayes
1	20:80	0,5079	0,5313
2	40:60	0,5467	0,5294
3	60:40	0,5625	0,5567
4	80:20	0,5767	0,5350
5	90:10	0,6020	0,5300
Rata-rata		0,5592	0,5365

**Tabel 9. Macro Average Recall dan Precision serta Average Accuracy Bukalapak**

Nilai	SVM	Naive Bayes
Macro Average Recall	0.5295	0.4973
Macro Average Precision	0.5283	0.5339
Average Accuracy	0.5295	0.4676

**Tabel 10. Perbandingan Accuracy Bukalapak menggunakan 10-fold cross validation**

No	Rasio Perbandingan (%)	Accuracy	
		SVM	Naive Bayes
1	20:80	0,4988	0,5413
2	40:60	0,5239	0,5283
3	60:40	0,5375	0,5292
4	80:20	0,5350	0,4750
5	90:10	0,5567	0,4433
Rata-rata		0,5304	0,5034

### C. Analisis Hasil Pengujian

Perbandingan kinerja *machine learning* difokuskan pada tingkat akurasi dalam menentukan algoritma yang memiliki kinerja terbaik didasarkan pada tingkat akurasi yang tinggi [7]. Perbandingan akurasi berdasarkan jumlah data uji pada algoritma *Support Vector Machine* dan *Naive Bayes* pada masing-masing *e-commerce* berdasarkan hasil evaluasi menggunakan *10-fold cross validation* ditunjukkan oleh tabel berikut.

**Tabel 11. Perbandingan Accuracy pada rasio data yang berbeda menggunakan 10-fold cross validation**

No	Rasio Perbandingan	Acc Tokopedia		Acc Bukalapak	
		SVM	NB	SVM	NB
1	20:80 (600:2400)	0.5079	0.5313	0.4988	0.5413
2	40:60 (1200:1800)	0.5467	0.5294	0.5239	0.5283
3	60:40 (1800:1200)	0.5625	0.5567	0.5375	0.5292
4	80:20 (2400:600)	0.5767	0.5350	0.5350	0.4750
5	90:10 (2700:300)	0.6020	0.5300	0.5567	0.4433
Rata-rata		0.5592	0.5365	0.5304	0.5034

**Tabel 12. Perbandingan Average Accuracy SVM dan NB pada Tokopedia dan Bukalapak menggunakan 10-fold cross validation**

Rata-rata Akurasi	Tokopedia	Bukalapak
SVM	55.92%	53.04%
Naive Bayes	53.65%	50.34%

Dari hasil pengujian, terlihat bahwa metode *Support Vector Machine* memiliki tingkat performa yang lebih besar daripada *Naive Bayes* dengan selisih akurasi sebesar 2,27% untuk Tokopedia dan 2,7% untuk Bukalapak. Hal ini dikarenakan ketidakmampuan *Naive Bayes* dalam mengatasi masalah data berdimensi tinggi, untuk Tokopedia berdimensi 3000×4049 yang artinya terdiri dari 3000 dokumen dalam bentuk baris dan 4049 variabel berupa *term* (kata) dalam bentuk kolom sedangkan untuk Bukalapak berdimensi 3000×3019. Data berdimensi tinggi sangat berpengaruh dalam perhitungan parameter yang akhirnya menyebabkan rendahnya performansi *Naive Bayes*. Akan tetapi perbedaan akurasi tidak terlalu besar yang menunjukkan bahwa *Naive Bayes* masih merupakan metode yang baik dalam pengklasifikasian teks. *Naive Bayes* dapat menjadi pilihan mengingat sifat komputasinya yang lebih ringan.

Algoritma dengan akurasi terendah adalah *Naive Bayes* dengan rata-rata akurasi sebesar 53.65% untuk Tokopedia dan 50.34% untuk Bukalapak. Namun perlu diketahui bahwa model-model yang digunakan pada klasifikasi teks dengan menggunakan *Machine Learning* tidak akan memiliki hasil yang sama pada setiap data set, maksudnya adalah tidak ada model yang lebih superior dan memiliki akurasi yang paling tinggi pada setiap data set yang dilakukan proses training dan testing [8]. Misalnya pada perbandingan rasio 80:20 pada Bukalapak, dimana nilai akurasi *Naive Bayes* Bukalapak lebih tinggi daripada SVM. Teorema ini dinamakan sebagai teorema *No Free Lunch*.

Faktor lain yang menyebabkan rendahnya akurasi dari masing-masing algoritma adalah adanya dokumen uji yang semua *term*-nya tidak terdapat di dokumen latih, dengan kata lain *machine learning* tidak pernah mengenal *term* ini sebelumnya. Apabila ada *term* baru dari dokumen uji, perhitungan tidak dapat dilanjutkan karena *term* tersebut tidak memiliki nilai untuk dimasukkan ke persamaan dalam *machine learning* [9]. Alasan inilah yang membuat dokumen uji bersentimen netral sulit terdeteksi sebagai kelas sentimen netral, karena pada umumnya kalimat bersentimen netral memiliki variasi *term* yang lebih beragam daripada kelas sentimen yang lain. Tidak seperti pada kelas sentimen positif dan negatif, kelas sentimen netral cenderung tidak memiliki *term* yang khas, sehingga sulit dikenali, contohnya kata "info" yang mempunyai frekuensi tertinggi di sentimen positif dan netral yang terlihat pada visualisasi *word cloud* Tokopedia.

#### 4. SIMPULAN

- a. Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, penulis dapat menarik beberapa kesimpulan sebagai berikut: Dengan menggunakan perbandingan data latih dan data uji sebesar 20% : 80%, 40% : 60%, 60% : 40%, 80% : 20% dan 90%:10% diperoleh hasil klasifikasi sentimen dengan metode *Support Vector Machine* (SVM) dan menggunakan *10-fold cross validation* diperoleh nilai rata-rata akurasi sebesar 55,92% untuk Tokopedia dan 53,04% untuk Bukalapak. Sedangkan dengan metode *Naive Bayes* dan menggunakan *10-fold cross validation* diperoleh tingkat akurasi yang lebih kecil daripada metode *Support Vector Machine* (SVM) yaitu nilai rata-rata akurasi sebesar 53,65% untuk Tokopedia dan 50,34% untuk Bukalapak.
- b. Berdasarkan tingkat akurasi yang dihasilkan, *machine learning* dengan menggunakan algoritma *Support Vector Machine* (SVM) memiliki kinerja yang lebih baik dibanding algoritma *Naive Bayes* dalam melakukan klasifikasi tanggapan mengenai *e-commerce* Tokopedia maupun Bukalapak yang diperoleh melalui media sosial *Twitter*.
- c. Tidak terdapat perbedaan kinerja *machine learning* yang signifikan, baik ketika diaplikasikan pada data tanggapan *e-commerce* Tokopedia maupun Bukalapak, kinerja *machine learning* dengan menggunakan algoritma *Support Vector Machine* (SVM) memiliki nilai rata-rata akurasi lebih tinggi dibandingkan algoritma *Naive Bayes*.

- d. Tanggapan positif pada Tokopedia sering dikaitkan dengan kata “info”, “kasih”, “terima” dan sebagainya, sedangkan pada Bukalapak sering dikaitkan dengan kata “bantu”, “transaksi”, “cek”, “kendala” dan sebagainya. Untuk tanggapan negatif pada Tokopedia sering dikaitkan dengan kata “saldo”, “hasil”, “salah”, “nasabah”, “curi”, “cyber”, “crime” dan sebagainya, sedangkan pada Bukalapak sering dikaitkan dengan kata “mohon”, “transaksi”, “admin”, “bantu”, “cek”, “kendala” dan sebagainya. Sedangkan untuk tanggapan netral pada Tokopedia sering dikaitkan dengan kata “info”, “kendala”, “maaf”, “cek”, “terima”, “kasih” dan seterusnya, dan untuk Bukalapak sering dikaitkan dengan kata “cek”, “mohon”, “transaksi”, “bantu”, “admin”, “tolong”, “resi”, “nomor”, “kirin”, “bayar” dan seterusnya.

## Saran

Berdasarkan hasil analisis dan kesimpulan, dapat diberikan saran sebagai berikut:

- a. Data yang digunakan pada penelitian ini hanya satu periode, yaitu pada bulan Agustus 2019, sehingga perlu diperbanyak periode waktu pengumpulan data agar informasi yang diperoleh lebih mendalam.
- b. Pada penelitian ini, proses sentiment analysis dilakukan hanya berdasarkan pembobotan yang diukur dari angka kemunculan dan mengabaikan makna kata. Akan lebih baik bila pada penelitian selanjutnya diterapkan juga kedekatan suatu kata dengan kata yang lain (semantic). Misalnya mengkolaborasi tabel sinonim dengan tabel wordnet.
- c. Perlu adanya sebuah optimasi dalam model tersebut sehingga hasil yang didapatkan lebih sempurna. Maka saran untuk penelitian selanjutnya adalah melakukan setting parameter yang lebih tepat sehingga didapatkan tingkat akurasi yang lebih baik dan melakukan eksperimen dengan model-model lain.
- d. Bagi pihak e-commerce, hasil ekstraksi informasi dari tanggapan yang telah diberikan oleh pengguna khususnya tanggapan yang berbentuk negatif dapat dijadikan bahan evaluasi dalam peningkatan kepuasan pengguna dan memberikan pelayanan semaksimal mungkin, serta untuk pengembangan pembaharuan aplikasi serta bisnis selanjutnya.

## 5. DAFTAR PUSTAKA

- [1] E. B. Santoso and A. Nugroho, “Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook,” *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [2] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm,” *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [3] H. Tuhuteru, “Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine,” *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.
- [4] J. LING, I. P. E. N. KENCANA, and T. B. OKA, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square,” *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014, doi: 10.24843/mtk.2014.v03.i03.p070.
- [5] D. M. U. Atmaja and R. Mandala, “Analisa Judul Skripsi untuk Menentukan Peminatan Mahasiswa Menggunakan Vector Space Model dan Metode K-Nearest Neighbor,” *IT Soc.*, vol. 4, no. 2, pp. 1–6, 2020, doi: 10.33021/itfs.v4i2.1182.
- [6] W. Setyobudi, A. Alwi, and I. P. Astuti, “Sentimen Analisis Twitter Terhadap Penyelenggaraan Gojek Traveloka Liga 1 Indonesia,” *Komputek*, vol. 2, no. 1, p. 56, 2018, doi: 10.24269/jkt.v2i1.68.
- [7] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and

- random forest algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [8] R. Risnantoyo, A. Nugroho, and K. Mandara, “Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm,” *J. Informatics Telecommun. Eng.*, vol. 4, no. 1, pp. 86–96, 2020, doi: 10.31289/jite.v4i1.3798.
- [9] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.