

## ANALISIS SENTIMEN TERHADAP DAMPAK COVID-19 PADA PERFORMA TOKOPEDIA MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Dinda Tri Wisudawati<sup>1</sup>, Tiani Wahyu Utami<sup>2</sup>, Prizka Rismawati Arum<sup>3</sup>

<sup>1,2,3</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Muhammadiyah Semarang

dindatriwisudawati@gmail.com

**Abstrak.** Tokopedia merupakan *e-commerce* populer di Indonesia. Hal tersebut didukung dengan *rating* Tokopedia yang tinggi pada *Google Play*. Diperlukan sebuah metode yang mampu mengkategorikan *reviews* pengguna secara otomatis, apakah tergolong ke dalam klasifikasi positif atau negatif. Analisis Sentimen menggunakan *Support Vector Machine* (SVM) merupakan metode yang digunakan. Konsep SVM merupakan usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space* dengan memaksimalkan jarak antar kelas. Sehingga SVM dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang. Klasifikasi menggunakan SVM pada periode sebelum munculnya Covid-19 di Indonesia (Februari 2020) menghasilkan akurasi sebesar 87% dan 84% pada periode sesudah munculnya Covid-19 (April 2020). Hasil menunjukkan bahwa walaupun Covid-19 muncul di Indonesia, performa Tokopedia masih tetap terjaga dan pengguna masih tetap memberikan penilaian suka sekali. Hal ini dibuktikan dengan penurunan jumlah *review* negatif dari 43% pada Februari 2020 menjadi 27% pada April 2020.

**Kata Kunci:** *Review, Google Play, Tokopedia, Support Vector Machine, Analisis Sentimen*

### 1. Pendahuluan

Berdasarkan Digital 2020 diketahui bahwa pengguna internet di seluruh dunia mencapai angka 4,5 milyar orang dari seluruh penduduk di dunia yaitu sebesar 7,7 milyar. Indonesia berada pada peringkat tiga dunia dengan pertumbuhan populasi yang mengakses internet sebesar 17% dalam satu tahun terakhir. Angka ini sama dengan 25,3 juta pengakses internet baru dalam setahun (Wearesocial, 2020).

Berdasarkan hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2018, pengguna internet di Indonesia sebesar 64,8%. Atau sebanyak 171,17 juta jiwa pengguna internet di Indonesia dari total keseluruhan penduduk Indonesia sebesar 264,16 juta jiwa. Pertumbuhan pengguna internet di Indonesia dari tahun 1998 s.d. 2018 berdasarkan hasil survey yang dilakukan oleh APJII dapat dilihat pada Gambar 1 berikut:

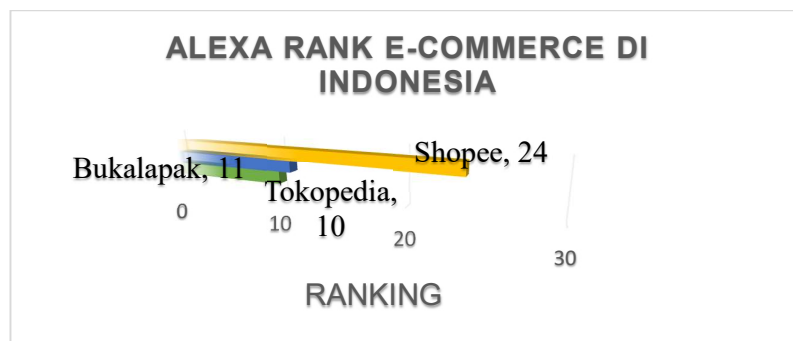


**Gambar 1.** Pertumbuhan Pengguna Internet

Salah satu perkembangan teknologi saat ini yang banyak dilirik oleh penduduk Indonesia adalah *e-commerce*. Menurut (Turban, dkk, 2012). *E-commerce (electronic commerce)* adalah pembelian, penjualan dan pemasaran barang serta jasa melalui sistem elektronik seperti radio, televisi dan jaringan komputer atau internet.

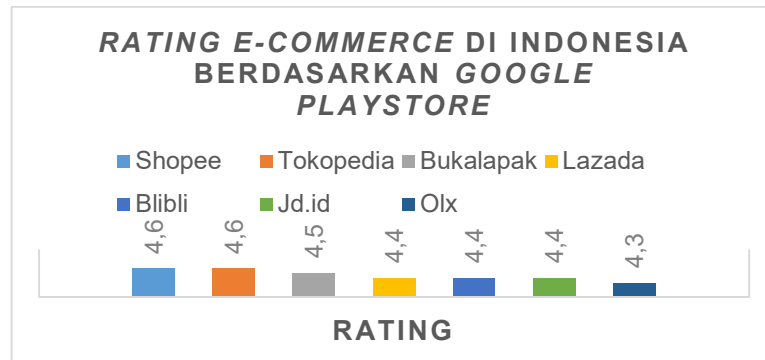
Hasil survei yang dilakukan oleh Jakpat, diketahui bahwa 60,5% penduduk Indonesia lebih memilih untuk belanja secara *online* karena lebih cepat dan efisien (65,7%), terdapat banyak promo dan diskon (62,9%), harga cenderung lebih murah (59,3%), dan fleksibilitas waktu berbelanja (59%) (Binus, 2019).

Berdasarkan data Alexa per tanggal 10 Juni 2020 *e-commerce* Indonesia yang berada pada peringkat 50 besar dapat dilihat pada Gambar 2 berikut:



**Gambar 2.** Alexa Rank E-Commerce di Indonesia (10 Juni 2020)

Berdasarkan Gambar di atas, Tokopedia memiliki ranking yang paling tinggi diantara *e-commerce* lainnya. Selain itu, berdasarkan *Google Play* per 7 Juni 2020, beberapa *rating* untuk aplikasi *e-commerce* dapat dilihat pada Gambar berikut:



**Gambar 3.** Rating *e-commerce* di Indonesia Berdasarkan *Google Play* (7 Juni 2020)

Berdasarkan gambar di atas, semakin membuktikan bahwa Tokopedia memang merupakan *e-commerce* yang paling populer di Indonesia, terbukti dengan *rating* 4.6 (*Google Play*, 2020). Tokopedia berhasil menjadi salah satu perusahaan internet Indonesia dengan pertumbuhan yang pesat. Tokopedia sendiri dapat diunduh melalui *Google Play*.

*Google Play* adalah layanan konten digital milik *Google* yang terdiri dari toko produk-produk *online* seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media. *Google Play* terdapat fitur berisi *reviews* dari para pengguna.

Di sisi lain, saat ini tengah terjadi pandemi *Coronavirus Disease 2019* (Covid-19) yang disebabkan oleh SARS-CoV-2 pada awal tahun 2020 yang berasal dari Wuhan, Tiongkok pada Desember 2019. Berdasarkan data WHO sampai 10 Juni 2020 negara yang sudah terjangkit sebanyak 213 negara, jumlah total pasien terinfeksi di seluruh dunia yaitu 7.342.779. Salah satu negara adalah Indonesia. Sampai 10 Juni 2020, Covid-19 sudah menyebar ke 424 kabupaten/kota di 34 provinsi di Indonesia dengan jumlah pasien positif sebanyak 34.316, pasien sembuh sebanyak 12.129 dan angka kematian di Indonesia sebesar 5,71% (Kompas, 2020).

Covid-19 sangat berpengaruh terhadap perekonomian di Indonesia yang dapat dilihat dari berbagai sisi. Salah satunya memberikan dampak pada *e-commerce*. *Review* pengguna pun semakin bervariasi setelah adanya Covid-19 ini. Dengan menggunakan *text mining* dapat dilihat pembicaraan atau *review* apa saja

yang sering dibahas oleh pengguna akibat dampak dari Covid-19 terhadap *e-commerce*. Menurut (Kristiyanti, 2015), analisis sentimen adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas.

Penelitian yang dilakukan oleh (Faadilah, 2020) terkait *review* pengguna *Tokopedia* menggunakan *Long Short Term Memory*, mendapatkan akurasi sebesar 93,3%. Penelitian lainnya dilakukan oleh (Darsyah, 2014) mengenai Klasifikasi Tuberkulosis dengan Pendekatan Metode *Supports Vector Machine* (SVM) menggunakan fungsi kernel RBF, dihasilkan akurasi sebesar 98%.

Dari penjelasan latar belakang di atas, akan dilakukan penelitian mengenai *review* *Tokopedia* pada *Google Play* sebelum (Februari 2020) dan sesudah (April 2020) munculnya Covid-19 untuk melihat performa *Tokopedia* setelah munculnya Covid-19 di Indonesia menggunakan metode *Support Vector Machine* (SVM) karena SVM memiliki tingkat akurasi paling tinggi dalam klasifikasi teks (Naradhipa & Purwarianti, 2012). Maka judul penelitian ini yaitu “Analisis Sentimen Terhadap Dampak Covid-19 Pada Performa *Tokopedia* Menggunakan *Support Vector Machine*”.

## 2. Metode Penelitian

### 2.1 Sumber Data

Data dalam penelitian ini adalah data sekunder yang diambil menggunakan proses *scraping* data pada kolom *review* *Tokopedia* berbahasa Indonesia melalui *Google Play* dengan bantuan *Jupyter Notebook*. Data yang diambil yakni dua periode waktu yaitu sebelum munculnya *Covid-19* di Indonesia pada Februari 2020 dan sesudah munculnya *Covid-19* pada April 2020.

### 2.2 Variabel Penelitian

Variabel yang digunakan dapat dilihat pada Tabel berikut:

**Tabel 1.** Variabel Penelitian

Variabel	Tipe Data	Keterangan
<i>Rating</i>	<i>Ordinal</i>	Tingkat kepuasan pengguna
<i>Date</i>	<i>Date</i>	Tanggal dibuatnya <i>review</i>
<i>Review</i>	<i>String</i>	Isi <i>review</i> pengguna

### 2.3 Analisis Data

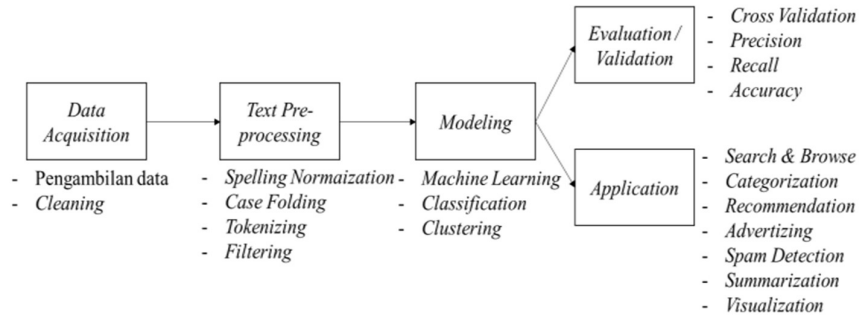
Dalam penelitian ini digunakan beberapa *software* di antaranya *R Studio* dan *Microsoft Excel 2016*. Adapun metode analisis yang digunakan yaitu:

- a. Analisis Deskriptif, digunakan untuk memberikan gambaran umum *review* Tokopedia pada *Google Play* serta tingkat kepuasannya.
- b. Analisis Sentimen, digunakan untuk melakukan pelabelan data ke dalam kelas sentimen positif dan negatif.
- c. *Support Vector Machine* (SVM), digunakan untuk melakukan klasifikasi *review* pengguna Tokopedia yang berbentuk positif dan negatif. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. SVM berusaha menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak (*margin*) antar kelas. Kelebihan *Support Vector Machine* yaitu metode yang paling akurat untuk teks klasifikasi (Moraes, dkk, 2013).

Langkah-langkah SVM sebagai berikut:

- a. Membagi data ke dalam data *training* dan *testing*.
- b. Membangun model SVM menggunakan fungsi kernel *Radial Basis Function* (RBF). Menurut hasil penelitian yang dilakukan oleh (Fauzi, dkk), penggunaan kernel RBF mampu menghasilkan akurasi 100%.
- c. Menentukan hasil prediksi model SVM melalui tabel *confussion matrix*.
- d. Menghitung akurasi prediksi untuk model SVM.

Menurut (Purbo, 2017), proses *text mining* seperti berikut:



**Gambar 4.** Proses Dasar *Text Mining*

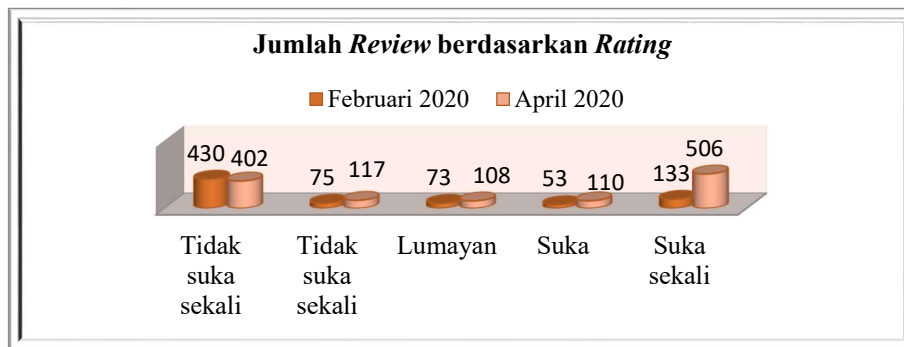
### 3. Hasil dan Pembahasan

#### 3.1 Hasil *Scraping Data*

*Scraping data reviews* Tokopedia melalui situs *Google Play* didapatkan pada Februari 2020 sebanyak 764 *reviews* dan April 2020 sebanyak 1.243 *reviews*.

#### 3.2 Analisis Data

Didapat jumlah *review* berdasarkan *rating* pada Gambar berikut:



**Gambar 5.** Perbandingan *Rating* Aplikasi

Terjadi peningkatan jumlah pengguna yang memberikan *rating* “suka sekali”, awalnya sebanyak 133 *review* dan April 2020 sebanyak 506 *review*. Sedangkan terjadi penurunan jumlah pengguna yang memberikan *rating* “tidak suka sekali” pada Februari sebanyak 430 dan pada April 2020 menjadi 402 *review*.

### 3.3 Pelabelan Kelas Sentimen

Proses pelabelan dilakukan secara otomatis dengan cara menghitung skor sentimen menggunakan kamus *lexicon* dan manual. Untuk skor  $< 0$  masuk ke dalam kelas negatif,  $= 0$  kelas netral dan  $> 0$  masuk dalam kelas positif. Pada penelitian ini digunakan dua pelabelan kelas sentimen yaitu sentimen positif dan negatif. Karena sebagai bahan pertimbangan bahwa kelas sentimen netral dianggap kurang memberikan manfaat bagi pihak Tokopedia.

### 3.4 Klasifikasi dengan *Support Vector Machine*

Hasil algoritma *Support Vector Machine* (SVM) kemudian dihitung nilai akurasi untuk memprediksi data baru. Pembagian data latih dan uji yaitu 80%:20%.

Penelitian ini menggunakan *kernel Radial Basis Function* (RBF) untuk proses klasifikasi sebelum dan sesudah munculnya Covid-19 di Indonesia. Perhitungan akurasi, mengikuti tabel *confusion matrix* berikut:

**Tabel 2.** *Confusion Matrix*

Prediksi	Februari 2020		April 2020	
	Positif	Negatif	Positif	Negatif
Positif	83	16	173	33
Negatif	4	50	8	36

Kemudian didapatkan akurasi menggunakan *kernel* RBF untuk Februari 2020 sebesar 87% sedangkan April 2020 sebesar 84%. Diketahui bahwa hasil klasifikasi SVM untuk kedua periode masuk ke dalam kategori klasifikasi “Baik”, karena masing-masing nilai akurasi berada pada rentang 80-90%, untuk tabel dapat dilihat seperti berikut (Sasongko, 2016):

**Tabel 3.** Kategori Klasifikasi

Nilai Akurasi	Kategori Klasifikasi
91-100%	Excellent (Sangat baik)
81-90%	Good (Baik)
71-80%	Cukup Baik
61-70%	Poor (Buruk)
51-60%	Fail (Sangat buruk)

### 3.5 Visualisasi



Visualisasi *Reviews Positif* Februari 2020    Visualisasi *Reviews Positif* April 2020

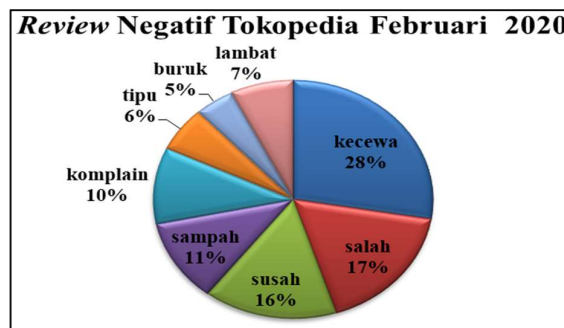
Semakin besar ukuran kata menggambarkan semakin banyak frekuensi kata tersebut digunakan oleh pengguna pada *review* Tokopedia. Berdasarkan visualisasi di atas, terlihat beberapa kata yang sering muncul yaitu “belanja”, “bagus”, “gratis”, “bantu”. Dimana keduanya berfokus kepada barang maupun fasilitas yang Tokopedia berikan.

Untuk *reviews* negatif terlihat pada gambar di bawah ini:



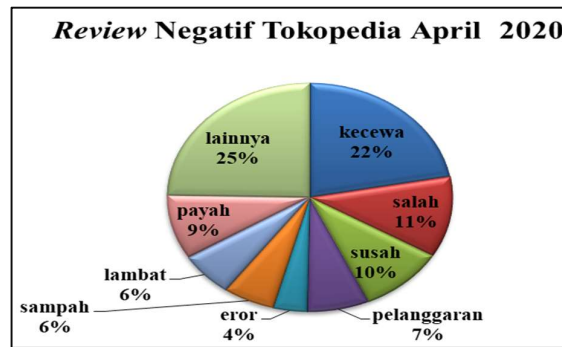
Visualisasi *Reviews Negatif* Februari 2020    Visualisasi *Reviews Negatif* April 2020.

Terlihat kata yang sering muncul pada Februari 2020 dan April 2020 yaitu “kecewa”, “susah”, “salah”.



Gambar 6. *Reviews* Negatif Februari 2020





**Gambar 7.** *Reviews* Negatif April 2020

Berdasarkan dua gambar di atas, terlihat bahwa persentase *reviews* negatif untuk kata “kecewa” mengalami penurunan, awalnya sebesar 28% pada Februari 2020 menjadi 22% pada April 2020. Dan kemunculan kata-kata lainnya seperti salah, sampah, susah pun mengalami penurunan. Ini bukti bahwa Covid-19 tidak mempengaruhi performa Tokopedia di mata pelanggan.

#### 4. Kesimpulan

Pada Februari 2020 sebanyak 56% pengguna sangat tidak menyukai dan 17% sangat menyukai Tokopedia. Sedangkan, April 2020 sebanyak 32% pengguna sangat tidak menyukai dan 41% sangat menyukai Tokopedia.

Hasil klasifikasi sentimen menggunakan (SVM) pada Tokopedia selama Februari 2020 terdapat 133 *review* terklasifikasi dengan tepat pada kelas sentimennya, serta menghasilkan akurasi sebesar 87%, sehingga tergolong ke dalam klasifikasi baik. Sedangkan pada April 2020, 209 *review* yang terklasifikasi dengan tepat pada kelas sentimennya, serta menghasilkan akurasi sebesar 84%, sehingga tergolong ke dalam klasifikasi baik.

Pada Februari 2020, sebanyak 43% pengguna menilai negatif Tokopedia. Sedangkan pada April 2020, sebanyak 27%. Dapat dilihat bahwa setelah munculnya Covid-19, *review* negatif Tokopedia semakin menurun. Disisi lain dapat dilihat dari jumlah pengguna yang tetap memberikan skor “sangat menyukai” pada April 2020 mengalami peningkatan menjadi 904 pengguna dan sebesar 73% pengguna Tokopedia memberikan penilaian positif. Sehingga dapat disimpulkan bahwa setelah munculnya Covid-19 di Indonesia, tidak

berpengaruh berarti terhadap performa Tokopedia dan Tokopedia mampu menjaga performa menjadi lebih baik.

### Daftar Pustaka

- Afifah, Faadilah. 2020. *Analisis Sentimen pada Reviews Aplikasi Tokopedia di Google Play Store Menggunakan Metode Long Term Short Memory*. Skripsi: Program Studi Matematika UIN Syarif Hidayatullah Jakarta.
- APJII. 2018. *Penetrasi & Perilaku Pengguna Internet Indonesia*. <https://apjii.or.id/survei2018>. Akses, 10 Juni 2020.
- Cortes, C., & V. Vapnik. 1995. *Support Vector Networks Dalam Machine Learning Vol. 20*, 273-297.
- Darsyah, Moh. Yamin. 2014. *Klasifikasi Tuberkulosis dengan Pendekatan Metode Supports Vector Machine (SVM)*. *Statistika*, 2(2), 37-41.
- Fatkurokhman, Fauzi, Moh. Yamin D, Tiani Wahyu U. *Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota se-Indonesia dengan Pendekatan Smooth Super Vector Machine (SSVM) Kernel Radial Basis Function (RBF)*. Semnas Pendidikan, Sains & Teknologi, UNIMUS. ISBN.978-602-61599-6-0.
- Google Play. 2020. Tokopedia. [https://play.google.com/store/apps/details?id=com.tokopedia.tkpd&hl=in\\_US&showAllReviews=true](https://play.google.com/store/apps/details?id=com.tokopedia.tkpd&hl=in_US&showAllReviews=true). Diakses, 7 Juni 2020.
- Kompas. 2020. Update Virus Corona di Dunia. [www.kompas.com](http://www.kompas.com). Diakses, 10 Juni 2020.
- Kristiyanti, D.A. 2015. *Analisis Sentimen Review Produk Kosmetik Melalui Komparasi Feature Selection*. Bekasi. ISBN 978-602-72850-0-2.74.
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. 2013. *Document-level sentiment classification: An empirical comparison between SVM and ANN*. *Expert Systems with Applications*, 40(2), 621-633. <https://doi.org/10.1016/j.eswa.2012.07.05>.
- Onno, W. Purbo. 2017. *Text Mining Analisis Medsos, Kekuatan Brand & Intelegen di Internet*. Yogyakarta. Andi Yogyakarta.
- Turban, E., King, D., Lee, J.K., Liang, T.-P., Turban, D.C. 2012. *Electronic Commerce: A Managerial and Social Networks. Perspective 8 th ed. Revised Edition*. USA: Sringer. Wong, CK.
- Wearesocial. 2020. Digital in 2020. <https://wearesocial.com/digital-2020>. Diakses 10 Juni 2020