



**ANALISIS SENTIMEN TERHADAP DAMPAK COVID-19 PADA
PERFORMA *E-COMMERCE* DI INDONESIA MENGGUNAKAN
SUPPORT VECTOR MACHINE
(*REVIEW APLIKASI TOKOPEDIA PADA
GOOGLE PLAY*)**

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

**DINDA TRI WISUDAWATI
B2A219016**

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN
ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2020**

PERSETUJUAN PEMBIMBING

Nama : Dinda Tri Wisudawati
NIM : B2A219016
Program Studi : SI STATISTIKA
Judul Skripsi : Analisis Sentimen Terhadap Dampak Covid-19 Pada Performa *E-Commerce* Di Indonesia Menggunakan *Support Vector Machine* (Review Aplikasi Tokopedia Pada *Google Play*)

Skripsi ini telah disetujui oleh pembimbing pada tanggal 24 September 2020 untuk diajukan ke sidang ujian lulus Skripsi Program Sarjana.



Pembimbing Utama,

Tiani Wahyu Utami, M.Si

NIK.28.6.1026.341

Pembimbing Pendamping

Prizka Rismawati Arum, M.Stat

NIK.CP.1026.071

PENGESAHAN KELULUSAN

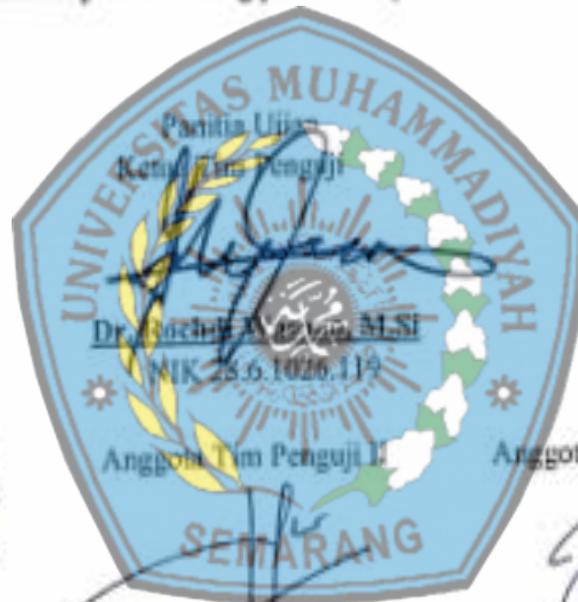
Skripsi dengan Judul "Analisis Sentimen Terhadap Dampak Covid-19 Pada Performa *E-Commerce* Di Indonesia Menggunakan *Support Vector Machine* (*Review* Aplikasi Tokopedia Pada *Google Play*)" yang disusun oleh:

Nama : Dinda Tri Wisudawati

NIM : B2A219016

Program Studi : Statistika

Telah dipertahankan dalam Sidang Panitia Ujian Skripsi Program Sarjana, Universitas Muhammadiyah Semarang pada 24 September 2020.



Anggota Tim Penguji I

Indah Manfaat Nur, M.Si

NIK. 28.6.1026.221

Anggota Tim Penguji II

Tiani Wahyu Utami, M.Si

NIK.28.6.1026.341

Anggota Tim Penguji III

Prizka Rismawati Arum, M.Stat

NIK.CP.1026.071



Mengetahui
Ketua Program Studi

Indah Manfaat Nur, M.Si

NIK. 28.6.1026.221

PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa :

1. Karya tulis saya, skripsi ini, adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana baik di Universitas Muhammadiyah Semarang maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing dan masukan Tim Penguji.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai kutipan dalam paragraf dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi ini.



Semarang, September 2020
Yang membuat pernyataan,



Dinda Tri Wisadawati

B2A219016

ANALISIS SENTIMEN TERHADAP DAMPAK COVID-19 PADA PERFORMA E-COMMERCE DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE (REVIEW APLIKASI TOKOPEDIA PADA GOOGLE PLAY)

Dinda Tri Wisudawati¹, Tiani Wahyu Utami², Prizka Rismawati Arum³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

e-mail : 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: *Google Play, Tokopedia, Support Vector Machine, Analisis Sentimen*

ABSTRACT

Tokopedia is a popular e-commerce in Indonesia. This is supported by Tokopedia's high rating on Google Play. A method is needed that can automatically categorize user reviews, whether they are classified as positive or negative. Sentiment analysis using a Support Vector Machine (SVM) is the method used. The SVM concept is an effort to find the best hyperplane that functions as a separator of two classes in the input space by maximizing the distance between classes. So that SVM can guarantee high generalizability for future data. Classification using SVM in the period before the Covid-19 factor in Indonesia (2020) resulted in an accuracy of 87% and 84% in the period after calculating Covid-19 (April 2020). The results show that even though Covid-19 appeared in Indonesia, Tokopedia's performance was still maintained and still gave very good results. This is evidenced by the decrease in the number of negative reviews from 43% in February 2020 to 27% in April 2020.

Keywords: *Google Play, Tokopedia, Support Vector Machine, Sentiment Analysis*

PENDAHULUAN

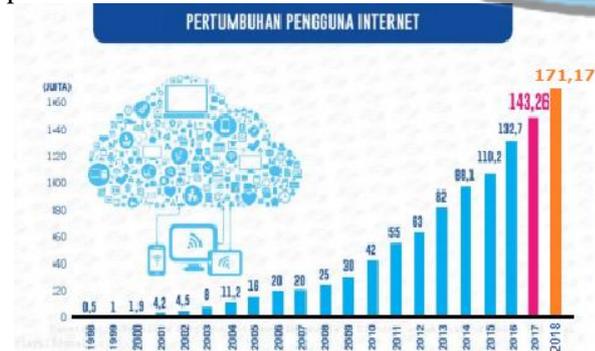
Di era modern saat ini, perkembangan teknologi di Indonesia maupun dunia sudah berkembang dengan begitu cepat dan pesat sehingga menjadikan kondisi saat ini sebagai era digital. Era digital ini telah membuat manusia tidak dapat lepas dari elektronik. Karena semua data dan informasi bahkan jasa sudah sangat banyak tersebar melalui internet.

Berdasarkan Digital 2020 diketahui bahwa pengguna internet di seluruh dunia telah mencapai angka 4,5 milyar orang dari seluruh penduduk di dunia yaitu sebesar 7,7 milyar. Angka ini menunjukkan bahwa pengguna internet telah mencapai lebih dari 60 persen penduduk dunia atau lebih dari separuh populasi

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, dan tentunya mengalami kenaikan dari tahun 2017 sebesar 10,12% atau sebanyak 143,26 juta jiwa yang sudah menyebar merata di seluruh Indonesia.

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

Untuk melihat ranking masing-masing *e-commerce* di Indonesia dapat dilihat melalui Alexa. 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, terdapat tiga *e-commerce* yang dinilai paling populer menurut Alexa diantaranya Shopee, Bukalapak dan Tokopedia. Tokopedia memiliki ranking yang paling tinggi diantara *e-commerce* lainnya yaitu ranking 10 di Indonesia.

Selain itu, banyaknya aplikasi *e-commerce* pun sudah terdaftar di *Google Play*. 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. Dimana angka *rating* tersebut dapat dikatakan cukup tinggi (*Google Play*, 2020).

Tokopedia adalah salah satu perusahaan jual beli berbasis digital terbesar di Indonesia dengan misi mencapai pemerataan ekonomi secara digital. Dimana Tokopedia dapat menjangkau sekitar 97% kecamatan di seluruh Indonesia dengan jumlah pengguna aktif sekitar 90 juta setiap bulannya. Tokopedia berhasil menjadi salah satu perusahaan internet Indonesia dengan pertumbuhan yang pesat. Dengan mengusung model bisnis C2C (*Customer to Customer*). C2C merupakan aktivitas bisnis (penjualan) yang dilakukan oleh konsumen kepada konsumen lainnya (Sandhausen, 2008).

Google Play adalah layanan konten digital milik *Google* yang terdiri dari toko produk-produk *online* seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media. Dalam *Google Play* dilengkapi dengan adanya fitur berisi *reviews* dari para pengguna yang dapat digunakan untuk melihat *reviews* dari pengguna aplikasi serta melihat performa suatu perusahaan.

Jumlah data *review* pengguna aplikasi Tokopedia yang masuk ke situs *Google Play* terus bertambah seiring berjalannya waktu. Berdasarkan data *Google Play* per 7 Juni 2020, jumlah *review* pengguna Tokopedia sebanyak 2.601.555.

Di satu 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 yang terjangkit 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% sebanyak 1.959 (Kompas, 2020).

Covid-19 sangat berpengaruh terhadap perekonomian di Indonesia yang dapat dilihat dari berbagai sisi. Mulai dari semakin berkurangnya pengunjung di pusat perbelanjaan seperti pasar tradisional, *supermarket*, *mall* bahkan untuk

sementara waktu beberapa tempat umum yang harus ditutup. Untuk mencegah penularan sesuai dengan kebijakan pemerintah “*stay at home*”, dan menjadikan sebagian perilaku masyarakat Indonesia beralih pada belanja *online*. Selain itu, memberikan dampak pada berbagai hal termasuk *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 atau 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 hasil tingkat akurasi untuk klasifikasi sentimen 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 dalam mengukur ketepatan klasifikasi sebesar 98%. Demikian juga penelitian yang dilakukan oleh Irfani, dkk (2020) yang meneliti tentang *review* aplikasi Ruangguru menggunakan *Support Vector Machine*. Diperoleh hasil pengujian untuk keakurasian berada di kisaran 90%.

Dari penjelasan latar belakang di atas, akan dilakukan penelitian mengenai *review* Tokopedia sebelum dan sesudah Covid-19 di *Google Play* untuk mengetahui bagaimana opini pengguna terhadap Tokopedia. Akan dilakukan klasifikasi *review* pengguna Tokopedia apakah termasuk *review* positif atau negatif. Dalam penelitian ini akan menyajikan analisis sentimen menggunakan metode *Support Vector Machine* (SVM) karena SVM memiliki tingkat akurasi paling tinggi dalam klasifikasi teks (Naradhipa & Purwarianti, 2012). Objek penelitian ini adalah *review* pengguna Tokopedia di *Google Play* sebelum adanya Covid-19 pada Februari 2020 dan sesudah adanya Covid-19 pada April 2020. Maka pada penelitian ini penulis akan melakukan Analisis Sentimen Terhadap Dampak Covid-19 Pada Performa *E-Commerce* Di Indonesia Menggunakan *Support Vector Machine* (*Review* Aplikasi Tokopedia Pada *Google Play*).

TINJAUAN PUSTAKA

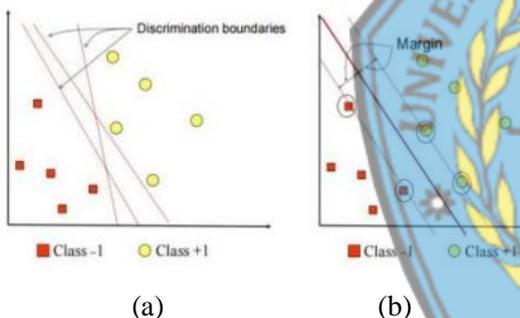
1. Support Vector Machine

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 sebagai suatu teknik klasifikasi yang efisien untuk masalah nonlinear. SVM juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network* (NN).

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).

2. SVM pada Data Terpisah secara Linear

Ilustrasi SVM pada data terpisah secara linear dapat dilihat pada Gambar di bawah ini:



Gambar 4. Ilustrasi SVM pada data terpisah secara linear

(Sumber: Nugroho, Witarto & Handoko, 2003)

Gambar di atas memperlihatkan konsep dasar *Support Vector Machine* (SVM) dimana penyebaran data memiliki dua kelas yang ditunjukkan oleh kotak warna merah (kelas negatif yang dinotasikan dengan -1) dan lingkaran warna kuning (kelas positif yang dinotasikan dengan +1). Garis solid warna merah pada Gambar 2b sebelah kanan menunjukkan *hyperplane* pemisah terbaik, yaitu yang terletak tepat di tengah-tengah kedua kelas, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam pada bidang pembatas adalah *support vector*. Upaya mencari lokasi *hyperplane* optimal ini merupakan inti dari proses pembelajaran pada SVM. Bidang pembatas pertama membatasi kelas pertama sedangkan pada

pembatas bidang kedua membatasi kelas kedua sehingga diperoleh persamaan :

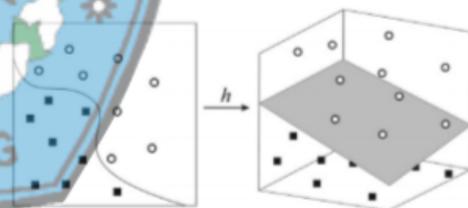
$$\begin{aligned}x_i \cdot w + b &= 0, \text{ untuk } y_i = 0 \text{ (hyperplane)} \\x_i \cdot w + b &\geq +1, \text{ untuk } y_i = +1 \text{ (kelas positif)} \\x_i \cdot w + b &\leq -1, \text{ untuk } y_i = -1 \text{ (kelas negatif)}\end{aligned}$$

Keterangan :

w = bobot
x = data (*input*)
b = bias

3. SVM pada Data Tidak Terpisah secara Linear

Dalam beberapa kasus, dapat ditemukan bahwa himpunan data tidak dapat dipisahkan secara linear. SVM mampu menyelesaikan permasalahan tidak linear dengan menggunakan teknik kernel (Cortes dan Vapnik, 1995). Penggunaan kernel ini memetakan vektor masukan pada ruang berdimensi rendah ke ruang berdimensi lebih tinggi. Gambar 5 menunjukkan bahwa data masukan yang tidak dapat dipisahkan secara linear kemudian ditransformasikan ke dalam ruang berdimensi lebih tinggi (*feature space*). Jika pada data linear, *hyperplane* berbentuk sebuah garis yang memisahkan antar kelas, maka pada data non linear, *hyperplane* akan berbentuk sebuah bidang yang memisahkan antar kelas.



Gambar 5 Transformasi dari *input space* ke *feature space*

(Sumber: Moraes dkk, 2013)

Kasus data yang tidak terpisah secara linear diasumsikan bahwa kelas pada *input space* tidak dapat terpisah secara sempurna. Selain itu, *feature space* pada kenyataannya memiliki dimensi yang lebih tinggi dari pada vektor input (*input space*). Hal ini menyebabkan komputasi pada *feature space* sangat besar sehingga terdapat kemungkinan *feature space* memiliki jumlah *feature* tak terhingga. Maka untuk mengatasi masalah ini SVM menggunakan “*kernel trick*”.

4. Kernel Trick pada SVM

Efektivitas *Support Vector Machine* sangat dipengaruhi oleh jenis fungsi kernel yang dipilih

dan diterapkan berdasarkan karakteristik data (Haddi, dkk, 2013). Inilah yang disebut *Kernel Trick*, dimana formulasi fungsi tersebut sebagai berikut:

$$K(x_i, x_j) = \phi_i(x_i) \cdot \phi_j(x_j)$$

Beberapa *kernel* yang umum digunakan pada SVM di antaranya:

a. Polynomial

Kernel trick polynomial diformulasikan untuk digunakan dalam menyelesaikan masalah klasifikasi, dimana dataset pelatihan yang digunakan sudah normal. Berikut persamaan:

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d$$

b. Radial Basis Function (RBF)

Kernel RBF ini merupakan *kernel* yang paling banyak digunakan dalam penyelesaian masalah klasifikasi untuk dataset yang tidak terpisah secara *linear*, dikarenakan pada *kernel* ini memiliki akurasi prediksi yang sangat baik. Persamaan yang dimiliki sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \exp(-\|\vec{x}_i - \vec{x}_j\|^2) \gamma$$

c. Sigmoid Kernel

Sigmoid merupakan *kernel trick* SVM yang merupakan pengembangan dari jaringan saraf tiruan, dimana *kernel* ini dinyatakan dengan persamaan berikut:

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i \cdot \vec{x}_j + \beta)$$

Dari keseluruhan *kernel trick* tersebut, *kernel trick radial basisfunction* merupakan *kernel trick* yang memberikan hasil terbaik pada proses klasifikasi khususnya untuk data yang tidak dapat dipisahkan secara *linear*. Untuk klasifikasi sebuah objek data x dapat diformulasikan sebagai berikut :

$$f(x) = \sum_{i=1, \vec{x}_i \in SV}^n \alpha_i y_i K(\vec{x}_i, \vec{x}_j) + b$$

5. Text Mining

Text mining (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007). *Text*

mining biasanya termasuk kategorisasi teks, teks *clustering*, ekstraksi konsep, analisis sentiment, merangkum dokumen dan pemodelan hubungan entitas (misalnya mempelajari hubungan antar entitas) (Purbo, 2017).

Untuk mengubah teks menjadi data untuk dianalisis, pada umumnya menggunakan *Natural Language Processing* (NLP) (Purbo, 2017). Tahapan-tahapan dalam *text mining* secara umum adalah *text preprocessing* dan *feature selection* (Feldman & Sanger, 2007).

6. Text Preprocessing

Text preprocessing berfungsi untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur. Secara umum proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut:

a. Spelling Normalization

Merupakan proses perbaikan atau substitusi kata-kata yang salah eja atau disingkat dalam bentuk tertentu.

b. Case Folding

Merupakan tahapan untuk merubah semua huruf dalam dokumen menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk mempermudah pencarian. Karena tidak semua dokumen teks atau *database* konsisten dalam penggunaan huruf kapital.

c. Tokenizing

Merupakan tahapan untuk memotong dokumen menjadi bagian-bagian kecil berdasarkan tiap kata yang menyusunnya atau disebut dengan *token*, disertai tahapan untuk membuang tanda baca seperti @, \$, &, tanda titik (.), koma (,) tanda tanya (?), tanda seru (!) (Manning, Raghavan, dan Schutze, 2009). Berfungsi untuk menjadikan sebuah kalimat menjadi lebih bermakna.

d. Filtering

Merupakan tahapan untuk mengambil kata-kata penting dari hasil *tokenizing* menggunakan *stopword removal* (membuang kata yang kurang penting). Sehingga kata-kata tersebut dapat dibuang dan hanya menyisakan kata-kata penting untuk dapat memiliki arti yang akan diproses ke tahap berikutnya. Contohnya adalah “yang”, “dan”, “di”, “dari” dan sebagainya (Putri, 2016).

7. Pembobotan Kata

Tahap ini bertujuan untuk memberi nilai frekuensi suatu kata sebagai bobot. *Term* dapat berupa kata, frasa atau unit hasil *indexing* lainnya

dalam suatu dokumen yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut. Karena setiap kata dalam dokumen memiliki tingkat kepentingan masing-masing, maka untuk setiap kata tersebut diberikan sebuah indikator, yaitu *term weight* (Zafikri, 2008). Menurut Zafikri (2008) *term weighting* atau pembobotan katadipengaruhi oleh hal-hal di antaranya:

a. Term Frequency (TF)

Merupakan frekuensi kemunculan sebuah kata (*term*) dalam sebuah dokumen. Semakin besar jumlah *term* yang muncul (TF tinggi) maka semakin besar bobot dokumen atau memberikan nilai kesesuaian yang semakin besar (Informatikalogi, 2016).

b. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) merupakan metode statistik numerik yang menghitung seberapa pentingnya kata dalam sebuah dokumen dimana dalam konteks ini yaitu pengurangan dominansi *term* yang sering muncul di berbagai dokumen. Metode ini digunakan sebagai bobot dalam pencarian informasi dalam *text mining* (Fanani, 2017).

Formula IDF dapat dituliskan sebagai berikut:

$$idf_j = \log\left(\frac{D}{df_j}\right)$$

Keterangan:

D : jumlah semua dokumen dalam koleksi

df_j : jumlah dokumen mengandung *term* t_j

Rumus umum untuk TF-IDF (bobot term terhadap dokumen) adalah penggabungan dari perhitungan *raw TF* dan formula IDF dengan cara mengalikan nilai keduanya, seperti berikut:

$$w_{ij} = tf_{ij} \times idf_j$$

$$w_{ij} = tf_{jj} \times \log\left(\frac{D}{df_j}\right)$$

8. Analisis Sentimen

Analisis sentimen adalah bidang ilmu data mining yang bertujuan untuk menganalisis, memahami, mengolah dan mengekstrak data teks yang berupa opini, sentimen, evaluasi, sikap, dan emosi terhadap suatu entitas seperti produk, servis, organisasi, individu, dan topik-topik tertentu. Klasifikasi sentimen digunakan untuk menyelesaikan masalah klasifikasi dua kelas, positif dan negatif. Data pengujian yang digunakan biasanya adalah *reviews/review* produk secara *online*.

9. Evaluasi Hasil Klasifikasi

Terdapat beberapa metode perhitungan yang digunakan untuk menilai performa sebuah klasifikasi misalnya Akurasi. Untuk memudahkan perhitungan, menggunakan table *confusion matrix*. Dapat dilihat pada Tabel berikut:

Tabel 1. Confusion Matrix

Prediction	Actual	
Class	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Accuracy dapat dihitung dengan persamaan berikut (Lim, dkk, 2006) :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

10. Akurasi Terbaik

Menurut (Sasongko, 2016), standar tabel kategori pengklasifikasian berdasarkan nilai akurasi dapat dilihat pada Tabel berikut:

Tabel 2. 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)

11. E-Commerce

E-commerce atau elektronik *commerce* (perdagangan secara elektronik) menurut (Turban, dkk, 2012) merupakan transaksi bisnis yang terjadi dalam jaringan elektronik seperti internet. Siapapun yang dapat mengakses komputer, memiliki sambungan ke internet dan memiliki cara untuk membayar produk-produk atau jasa yang dibeli, dapat berpartisipasi dalam *e-commerce*.

12. Tokopedia

Tokopedia merupakan salah satu perusahaan jual beli berbasis digital (*e-commerce*) terbesar di Indonesia dengan pertumbuhan pesat yang mengusung model bisnis *marketplace* dan *mall online*, serta memungkinkan setiap individu, toko kecil dan *brand* untuk membuka dan mengelola toko *online* masing-masing. Visi Tokopedia adalah "Membangun Indonesia lebih baik, lewat internet". Sedangkan misinya adalah selalu positif, memecahkan masalah, menjadi yang terbaik, generasi Indonesia yang lebih baik lagi dan fokus pada pelanggan.

13. Google Play

Google Play merupakan layanan konten digital toko aplikasi *online* milik *Google* yang menawarkan berbagai macam produk seperti music, buku, dan aplikasi yang dapat diakses melalui web, aplikasi android (*Play Store*) dan *Google TV*. *Google play* terdapat fitur *reviews* yang berisi *review* dari para pengguna.

14. Reviews

Pardiyono (2007) menyatakan bahwa teks *review* adalah teks yang berisi pemberian kritik, evaluasi terhadap karya cipta intelektual yang bertujuan untuk hasil evaluasi, atas suatu karya ilmiah, buku, atau karya seni. Fitur *review* yang diberikan oleh aplikasi-aplikasi dapat diberikan dan dilihat secara *online* pada suatu sumber yang terkoneksi pada sebuah jaringan komputer yang bersifat *online*.

METODE PENELITIAN

1. Sumber Data

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

2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel berikut:

Tabel 3. 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

3. Analisis Data

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

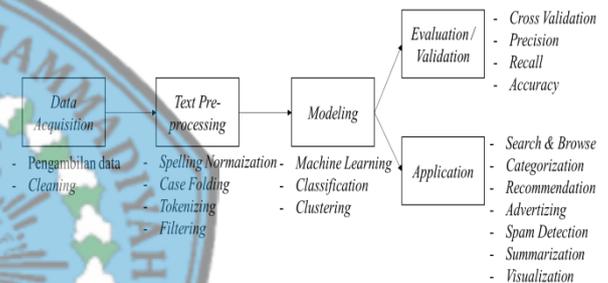
a. Analisis Deskriptif, digunakan untuk memberikan gambaran umum *review* Tokopedia pada *Google Play* serta tingkat kepuasannya.

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. Langkah-langkah SVM sebagai berikut:

- 1) Membagi data ke dalam data *training* dan *testing*.
- 2) Membangun model SVM menggunakan fungsi kernel RBF.
- 3) Menentukan hasil prediksi untuk model SVM melalui tabel *confussion matrix*.
- 4) Menghitung akurasi prediksi untuk model SVM.

Menurut (Purbo, 2017), gambaran umum proses *text mining* dapat dilihat pada gambar di bawah ini:



Gambar 6 Proses dasar text mining

HASIL DAN PEMBAHASAN

1. Hasil Scraping Data

Scraping data *reviews* Tokopedia melalui situs *Google Play* didapatkan data *review* pada Februari 2020 sebanyak 764 *record* dan April 2020 sebanyak 1.243 *record*. Proses *scraping* data menggunakan bantuan *Jupyter Notebook*.

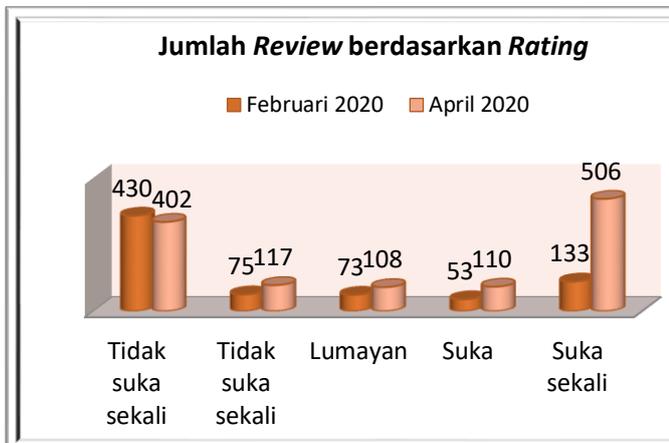
2. Analisis Deskriptif

Didapat jumlah *review* untuk Februari 2020 dan April 2020 pada Gambar berikut:



Gambar 7. Perbandingan Jumlah Review

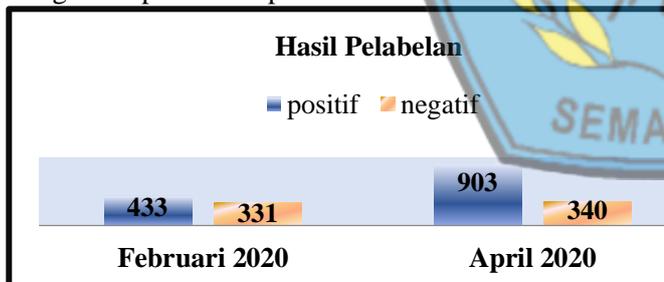
Pada Gambar di atas jumlah *review* Tokopedia pada April 2020 (sesudah munculnya Covid-19) mengalami kenaikan sebesar 63% atau terdiri dari 1.243 *review*.



Gambar 8 Perbandingan Rating Aplikasi

Gambar di atas merupakan perbandingan *rating* Tokopedia pada Februari 2020 dan April 2020. Terjadi peningkatan jumlah pengguna yang memberikan *rating* “suka sekali”, awalnya pada Februari 2020 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*.

Untuk hasil pelabelan kelas sentimen positif dan negatif dapat dilihat pada Gambar berikut :



Gambar 9. Hasil Pelabelan Reviews

Jumlah *review* positif pada kedua bulan April 2020 terlihat lebih besar dibanding *review* negatif. Pemberian nilai negatif pada Tokopedia oleh pengguna selama Februari 2020 sebesar 43%, sedangkan pada April 2020 sebesar 27%. Setelah munculnya Covid-19 ini, *review* negatif Tokopedia semakin menurun.

3. Pelabelan Kelas Sentimen

Proses pelabelan dilakukan secara otomatis dengan cara menghitung skor sentimen menggunakan kamus *lexicon* dan manual.

Pengelompokkan Kelas Sentimen dapat dilihat pada Tabel berikut:

Tabel 4. Pengelompokkan Pelabelan Berdasar Skor Sentimen

< 0	= 0	> 0
Negatif	Netral	Positif

Pada penelitian ini digunakan dua pelabelan kelas sentiment yaitu sentiment positif dan negatif. Karena sebagai bahan pertimbangan bahwa kelas sentimen netral dianggap kurang memberikan manfaat bagi pihak Tokopedia. Proses pelabelan ini menggunakan *RStudio*.

Hasil pelabelan kelas positif terjadi peningkatan, dimana pada Februari 2020 sebanyak 433 *reviews* sedangkan pada April 2020 sebanyak 903 *reviews*. Berikut hasil pelabelan kelas yang didapat:

Tabel 5. Hasil Pelabelan Kelas Sentimen

Kelas Sentimen	Sebelum Covid-19 (Feb 2020)	Setelah Covid-19 (April 2020)
Positif	433	903
Negatif	331	340
Total	764	1243

4. Klasifikasi dengan Support Vector Machine

Algoritma *Support Vector Machine* (SVM) untuk mendeteksi ciri-ciri data pada tiap kelas yang telah dibentuk pada data latih. Hasil pelatihan pada algoritma SVM kemudian diuji dengan menggunakan data uji untuk menghitung 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. Untuk menghitung akurasi, menggunakan tabel *confusion matrix*. Dapat dilihat pada tabel berikut:

Tabel 6. 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* sesuai Tabel berikut:

Tabel 7. Hasil Akurasi SVM dengan Kernel RBF

Periode	Akurasi
Februari 2020	87%
April 2020	84%

Diketahui bahwa hasil klasifikasi SVM berbasis *kernel RBF* pada Februari 2020 dan April 2020 sama-sama masuk ke dalam kategori klasifikasi yang “Baik”, karena masing-masing nilai akurasi berada pada rentang 80-90%.

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 *reviews* positif di atas, terlihat beberapa kata yang sering muncul yaitu “belanja”, “bagus”, “gratis”, “bantu”. Dimana keduanya berfokus kepada barang maupun fasilitas yang Tokopedia berikan.

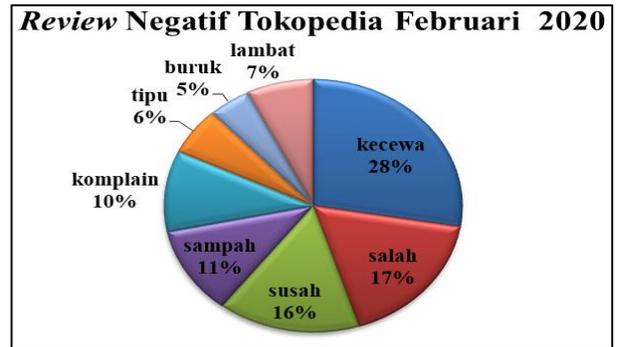
Untuk *reviews* negatif terlihat pada gambar di bawah ini:



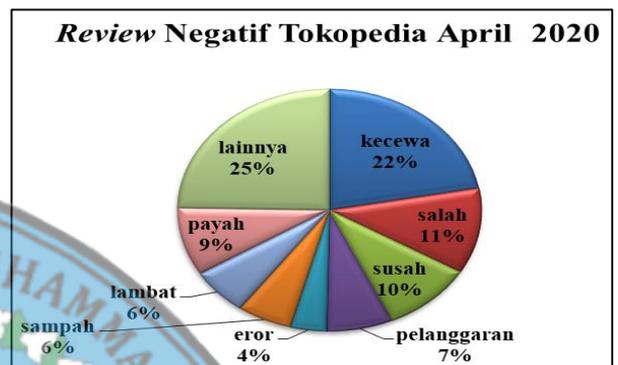
Visualisasi *Reviews Negatif* Februari 2020

Visualisasi *Reviews Negatif* April 2020

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



Gambar 10. Reviews Negatif Februari 2020



Gambar 11. 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.

KESIMPULAN

1. *Rating* aplikasi Tokopedia pada Februari 2020 sebanyak 56% pengguna sangat tidak menyukai dan 17% sangat menyukai aplikasi Tokopedia dari total 764 *review*. Sedangkan pada April 2020, sebanyak 32% pengguna sangat tidak menyukai dan 41% sangat menyukai aplikasi Tokopedia dari total 1243 *review*. Terlihat bahwa performa Tokopedia pada April 2020 lebih baik.
2. Hasil klasifikasi sentimen menggunakan *Support Vector Machine* (SVM) pada aplikasi Tokopedia selama Februari 2020 menggunakan 611 data latih dan 153 data uji. Dari data uji tersebut 87 data tergolong kelas positif dan sisanya 66 data tergolong kelas negatif serta menghasilkan tingkat akurasi sebesar 87%, sehingga tergolong ke dalam klasifikasi baik. Artinya bahwa dari 153 *review* terdapat 133

review terklasifikasi dengan tepat pada kelas sentimennya. Sedangkan pada April 2020 menggunakan 993 data latih dan 250 data uji. Dari data uji tersebut sebanyak 181 data tergolong kelas positif dan 69 data tergolong kelas negatif serta menghasilkan tingkat akurasi sebesar 84%, sehingga tergolong ke dalam klasifikasi baik. Artinya dari 250 review terdapat 209 review yang terklasifikasi dengan tepat pada kelas sentimennya.

3. Jumlah review positif pada kedua bulan tersebut pun lebih besar dibanding review negatif. Pada Februari 2020, sebanyak 43% pengguna menilai negatif Tokopedia. Sedangkan pada April 2020, sebanyak 27% pengguna memberikan nilai negatif pada Tokopedia. Dapat dilihat bahwa setelah munculnya Covid-19, review negatif Tokopedia semakin menurun. Disisi lain dapat dilihat dari jumlah pengguna yang tetap memberikan skor 5 (lima) atau sangat menyukai Tokopedia 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.
- Fanani, F. 2017. *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*. Skripsi: Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
- Feldman, R & Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Google Play. 2020. Tokopedia. https://play.google.com/store/apps/details?id=com.tokopedia.tkpd&hl=in_US&showAllReviews=true. Diakses, 7 Juni 2020.
- Haddi, E., Liu, X., & Shi, Y. 2013. *The Role of Text Pre-processing in Sentiment Analysis*. First International Conference on Information Technology and Quantitative Management, 17, 26–32. <https://doi.org/10.1016/j.procs.2013.05.05>.
- Han, J., & Kamber, M. 2012. *Data Mining: Concepts and Techniques Third Edition*. Waltham, MA: Morgan Kaufmann.
- Informatikalogi. 2016. *Pembobotan Kata atau Term Weighting TF-IDF*. Retrieved April 12, 2018, from Informatikalogi.com: <https://informatikalogi.com/term-weighting-tf-idf/>.
- Irfani, dkk, 2020. *Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine*. Yogyakarta. JBMI Vol. 16 No. 3.
- Kompas. 2020. *Update Virus Corona di Dunia*. 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.
- Lim, S. Y., Song, M.H., & Lee, S.J. 2006. *Ontology-based automatic classification of web documents*. Springer-Verlag, 690-700.
- Manning, C. D., Raghavan, P., & Schütze, H. 2009. *An Introduction to Information Retrieval – Online Edition*. Cambridge: Cambridge University Press.
- 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>.
- Naradhipa, R.A., & Purwarianti, A. 2012. *Sentiment Classification for Indonesian Message in Social Media*. *International Conference on Electrical Engineering and Informatics*: Bandung, 17-19 July 2011.
- Nugroho, Witarto & Handoko. 2003. *Application of Support Vector Machine in Bioinformatics*.

Proceeding of Indonesian Scientific Meeting in Central, 19-27.

- Onno, W. Purbo. 2017. *Text Mining Analisis Medsos, Kekuatan Brand & Intelejen di Internet*. Yogyakarta. Andi Yogyakarta.
- Pardiyono. 2007. *Pasti Bisa! Teaching Genre-Based Writing Metode Mengajar Writing Berbasis Genre Secara Efektif*. Yogyakarta: Andi Yogyakarta.
- Sandahusen, Richard. 2008. *Marketing. Hauppauge, N.Y: Barron's Educational Series*. p. 520. ISBN 0-7641-3932-0.
- Theopilus, Bayu Sasongko. 2016. *Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)*. *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 2, no.2. e-ISSN:2443-2229.
- 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.
- Zafikri, A. 2008. *Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) Pada Sistem Temu Kembali Informasi*. Skripsi: Program Studi S-1 Ilmu Komputer FMIPA USU.

