

ANALISIS SENTIMEN MEDIA SOSIAL *TWITTER* TERHADAP MASKAPAI PENERBANGAN PT GARUDA INDONESIA (PERSERO) TBK MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM)

Tantri Ayu Prasetiari¹, Iin Ernawati², Nurul Chamidah³
Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
tantriayup@gmail.com¹, iinerti@gmail.com², nurul.chamidah@upnvj.ac.id³

Abstrak. Media sosial *twitter* merupakan wadah untuk bertukar informasi. Informasi tersebut dapat berupa opini yang ditujukan oleh suatu perusahaan tertentu. Opini yang dibagikan dapat dianalisis menjadi analisis sentimen yang nantinya akan membantu dalam memahami sentimen opini masyarakat berbentuk teks yang awalnya tidak terstruktur menjadi terstruktur. Data opini tersebut dilabelkan dan diklasifikasikan menjadi kelas positif, dan kelas negatif, dengan menggunakan metode *support vector machine* (SVM), dimana data yang digunakan berasal dari sosial media *twitter* @indonesiagaruda. Dari hasil metode klasifikasi menggunakan *Support Vector Machine* (SVM) dilakukan evaluasi terhadap model dengan menggunakan *confusion matrix* terhadap perhitungan, sehingga mendapatkan nilai akurasi sebesar 88,75% dengan menggunakan kernel linear, sedangkan pada kernel *polynomial* mendapatkan nilai akurasi sebesar 75,625%. Model dengan akurasi tertinggi digunakan untuk prediksi data baru, sehingga mendapatkan hasil sentimen positif sebesar 291, dan sentimen negatif sebesar 273 dari total 564 *tweet* pada bulan Februari 2020.

Kata kunci: *twitter*, analisis sentimen, *support vector machine*

1 PENDAHULUAN

Twitter merupakan salah satu bentuk media sosial yang populer dalam bertukar informasi. Informasi yang dibagikan dapat berupa opini positif dan opini negatif berupa penilaian terhadap pelayanan yang diberikan oleh perusahaan tertentu. Penggunaan *twitter* kini banyak digunakan oleh perusahaan- perusahaan besar seperti maskapai penerbangan Garuda Indonesia yang menjadi objek penelitian ini. Media sosial *twitter* oleh Garuda Indonesia dimanfaatkan dalam berkomunikasi dan bertukar informasi kepada para pengguna media sosial khususnya kepada para pelanggannya. Informasi tersebut dapat dijadikan analisis sentimen ke dalam kelas positif dan kelas negatif.

Analisis sentimen membantu dalam memahami opini masyarakat berbentuk teks yang awalnya tidak terstruktur menjadi terstruktur dengan menggunakan *text mining* dalam pengolahan teks tersebut. Dalam proses pengklasifikasian sentimen positif dan negatif, penulis menggunakan metode klasifikasi yaitu *Support Vector Machine* (SVM).

Dengan menggunakan metode SVM diharapkan dapat digunakan dan dimanfaatkan untuk analisis sentimen dalam mengetahui kepuasan pelanggan berdasarkan kategori pelayanan seperti pelayanan terhadap fasilitas, ketepatan waktu, penilaian terhadap tarif tiket, *customer service*, serta fasilitas layanan Garudamiles dan *website* Garuda Indonesia yang diberikan oleh maskapai penerbangan Garuda Indonesia.

2 METODOLOGI

PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen adalah penelitian yang menganalisis opini dan sentimen masyarakat terhadap entitas seperti produk, layanan, dll dalam teks [1]. analisis otomatis dari konten *online* untuk mengekstraksi opini membutuhkan pemahaman mendalam tentang teks alami oleh mesin. Penelitian analisis sentimen dapat dikategorikan dalam tingkat dokumen, seperti mengklasifikasikan dokumen terhadap *review* yang mengandung polaritas positif atau negative [9].

2.2 *Text Mining*

Text mining sebagai sarana untuk dapat menganalisis data pelanggan yang tidak terstruktur menjadi terstruktur sehingga wawasan pelanggan yang berharga dapat diperoleh dari hasil *text mining* [5]. Metode analitik ini mengidentifikasi frasa yang tersembunyi dalam data yang disimpan dalam format teks bebas, membantu memahami setiap bagian dari komunikasi.

2.3 *Text analysis*

Text analysis adalah mengubah teks menjadi data untuk analisis [4]. Istilah *text analysis* menggambarkan sebuah set dari teknik *machine learning* yang memodelkan dan menstrukturkan isi informasi dari sumber teks untuk analisis data [8].

2.4 *Text Preprocessing*

Text preprocessing merupakan tahap sebuah proses dalam menyeleksi data yang awalnya tidak terstruktur menjadi terstruktur, dengan menghilangkan *noise* pada teks. Tahapan umum pada *text preprocessing* adalah sebagai berikut :

a) *Cleaning*

Cleaning adalah proses membersihkan *noise* seperti menghapus karakter- karakter yang ada pada data *tweet*, diantaranya sebagai berikut *html tags*, tanda baca (*punctuations*), sebagian besar angka (*most numbers*), ruang kosong (*white spaces*) [3].

b) *Case Folding*

Case folding adalah proses dalam mengubah huruf besar menjadikan semuanya huruf kecil [6].

c) *Tokenizing*

Tokenizing adalah sebuah proses dalam melakukan pemotongan sebuah kata dalam suatu kalimat yang menyusunnya [2].

d) *Filtering*

Filtering adalah proses menentukan kata- kata apa saja yang dianggap penting dalam sebuah dokumen..

e) *Stemming*

Stemming adalah proses mengubah kata ke bentuk dasarnya.

2.5 *Term Frequency – Inverse Document Frequency (TF IDF)*

Term Frequency – Inverse Document Frequency adalah *statistic numeric*, dimana pada proses ini melakukan analisis terhadap pentingnya sebuah kata pada suatu dokumen.

$$TF\ IDF = TF \times \log \frac{|D|}{DF_i} \quad (1)$$

Keterangan dari rumus $TF - IDF$ di atas adalah sebagai berikut

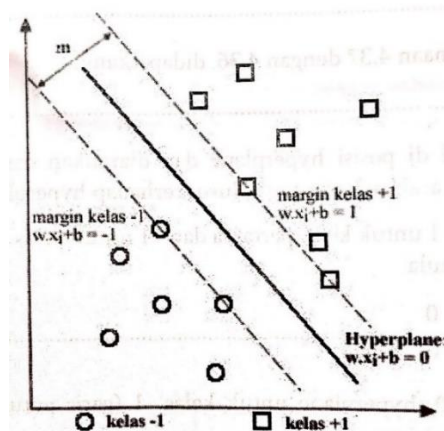
TF = Jumlah frekuensi kata terpilih / jumlah kata
 IDF = \log (Jumlah dokumen / jumlah frekuensi kata terpilih)
 D = Jumlah semua dokumen yang ada

2.6 Support Vector Machine (SVM)

Support vector machine (SVM) *Support vector machine (SVM)* adalah suatu teknik untuk melakukan prediksi, salah satunya dalam kasus klasifikasi. Dengan mencari *hyperplane* terbaik yang berfungsi sebagai pemisah 2 buah kelas.

2.6.1 Konsep Support Vector Machine (SVM)

Konsep dasar dari SVM adalah mencari *hyperplane* terbaik. Dimana fungsi dari *hyperplane* itu sendiri yaitu sebagai pemisah dua buah kelas pada *input space* [7].



Gambar. 1. *Hyperplane* pada klasifikasi linear *support vector machine (SVM)* pada kedua kelas -1 dan +1 adalah sebagai berikut

$$f(x) = \overline{w} \cdot \overline{x}_i + b = 0 \quad (2)$$

Sehingga diperoleh persamaan:

kelas -1 (sampel negatif), memenuhi pertidaksamaan

$$\overline{w} \cdot \overline{x}_i + b \leq -1 \quad (3)$$

kelas $+1$ (sampel positif), memenuhi pertidaksamaan

$$\overline{w} \cdot \overline{x}_i + b \geq +1 \quad (4)$$

Dimana pada rumus persamaan tersebut:

w : Parameter *hyperplane* yang dicari

b : Parameter *hyperplane* yang dicari

x_i : Data input SVM

Dalam *non linear SVM*, data \vec{x} dipetakan dalam fungsi $\phi(x)$ ke ruang vektor yang berdimensi lebih tinggi. Karena umumnya transformasi ϕ sulit untuk dipahami secara mudah, maka perhitungan *dot product* tersebut dapat digantikan dengan fungsi kernel.

$$K(\vec{x}_i, \vec{x}_j) \tag{5}$$

Penggunaan kernel pada penelitian ini adalah

- Kernel linear SVM, dengan persamaan

$$K(x_i, x_j) = x_i^T x_j \tag{6}$$

- Kernel *Polynomial*, dengan persamaan

$$K(x_i, x_j) = (x_i x_j + c)^d \tag{7}$$

2.7 Evaluasi

Dilakukan tahap evaluasi dengan menggunakan *confusion matrix* untuk mengukur kinerja metode SVM dalam memberikan hasil klasifikasi [10]. Hasil akhir yang diperoleh berupa akurasi dan *error rate* yang dihasilkan dari model yang dibangun, sehingga dapat dilakukan prediksi untuk mengetahui kepuasan pelanggan terhadap sentimen yang diberikan oleh pengguna *twitter* terhadap perusahaan yang bersangkutan.

Nilai prediksi	TRUE	(
	FALSE	()

Tabel 1 Contoh Tabel Confusion Matrix

- *True Positive (TP)*: Jumlah data positif yang terdeteksi benar.
- *True Negative (TN)*: Jumlah data negatif yang terdeteksi benar.
- *False Positive (FP)*: Data negatif tetapi terdeteksi sebagai data positif.
- *False Negative (FN)*: Data positif tetapi terdeteksi sebagai data negatif.

- Akurasi : Merupakan perhitungan dari semua nilai yang diprediksi benar dibagi dengan nilai keseluruhan data.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

- Error Rate : Merupakan perhitungan dari semua nilai yang diprediksi salah dibagi dengan nilai keseluruhan data.

$$Error Rate = \frac{FP + FN}{TP + FN + FP + TN} \tag{9}$$

3 HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini berasal dari media sosial *twitter* dengan kata kunci pencarian @IndonesiaGaruda yang diambil pada bulan Januari sampai dengan Februari 2020, dimana data pada bulan januari sebesar 800 data digunakan untuk membangun model yang dilabeli secara manual, sedangkan data pada bulan Februari digunakan untuk prediksi data baru dari model.

Tabel 2 Sampel Data

Tweet (D)	Sampel data sebelum preprocessing	Sampel data sesudah preprocessing	Kategori
D1	Terima kasih @IndonesiaGaruda, kualitas makanan bagus.	terima kasih kualitas makan bagus	Positif
D2	Terima kasih kepada @IndonesiaGaruda yang sudah memberikan pelayanan terbaiknya dlm memberi Makanan khusus muslim	terima kasih beri layan baik dlm beri makan khusus muslim	Positif
D3	@IndonesiaGaruda \nHallo min,saya sangat kecewa dengan ketidakpastian beroprasinya pesawat garuda, saya membeli tike\Xe2\x80\xa6 https://t.co/KMuVNYG4U5"	min sangat kecewa ketidakpastian beroprasinya pesawat garuda beli tiket	Negatif

Setelah dilakukannya *preprocessing* pada data *tweet*, kemudian dilakukan pembobotan kata dengan melihat frekuensi kata yang ada pada data *tweet*. Pada penelitian ini terdapat 1880 variabel kata dari 800 data *tweet* hasil pembobotan kata dengan menggunakan TF IDF.

Tabel 3 Perhitungan TF IDF

Term	Dokumen			DF	IDF	TF- IDF		
	D1	D2	D3			D1	D2	D3
bagus	1	0	0	1	0,477	0,477		
baik	0	1	0	1	0,477		0,477	
beli	0	0	1	1	0,477			0,477
beri	0	1	0	1	0,477		0,477	
beroprasinya	0	0	1	1	0,477			0,477
dlm	0	1	0	1	0,477		0,477	
garuda	0	0	1	1	0,477			0,477
kasih	1	1	0	2	0,176	0,176	0,176	

Term	Dokumen			DF	IDF	TF- IDF		
	D1	D2	D3			D1	D2	D3
kecewa	0	0	1	1	0,477			0,477
ketidakpastian	0	0	1	1	0,477			0,477
khusus	0	1	0	1	0,477		0,477	
kualitas	1	0	0	1	0,477	0,477		
layan	0	1	0	1	0,477		0,477	
makan	1	1	0	2	0,176	0,176	0,176	
min	0	0	1	1	0,477			0,477
muslim	0	1	0	1	0,477		0,477	
pesawat	0	0	1	1	0,477			0,477
sangat	0	0	1	1	0,477			0,477
terima	1	1	0	2	0,176	0,176	0,176	
tiket	0	0	1	1	0,477			0,477

Pada proses klasifikasi menggunakan metode *Support Vector Machine* (SVM) dengan pembagian *training dataset* dilakukan dengan menggunakan 80 % yaitu sebesar 640 data *tweet* data dari total keseluruhan jumlah data yang sudah dilabeli yaitu 800 data sampel, dan untuk *testing dataset* menggunakan 20 % sebesar 160 data *tweet* dari sisa data yang sudah digunakan pada *training dataset*. Pada penelitian ini menggunakan kernel linear dan kernel *polynomial* SVM dengan pemilihan parameter C dari kedua kernel tersebut yaitu bernilai 1,0. Dari penggunaan fungsi kernel tersebut didapatkan pengevaluasian dengan menggunakan *confusion matrix* dalam mengukur kinerja pengklasifikasian yang dilakukan sebelumnya, sebagai berikut.

Untuk hasil dari kernel linear didapatkan

$$Akurasi = \frac{67 + 75}{67 + 75 + 2 + 16} = 88,75\% \quad (10)$$

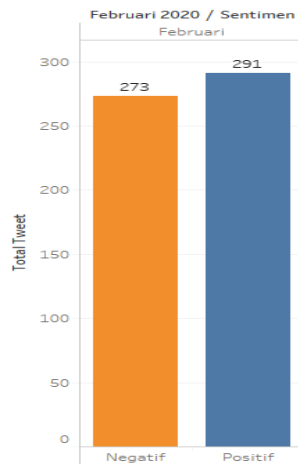
$$Error Rate = \frac{2 + 16}{67 + 16 + 2 + 75} = 11,25\% \quad (11)$$

Sedangkan untuk hasil dari kernel *polynomial*, didapatkan

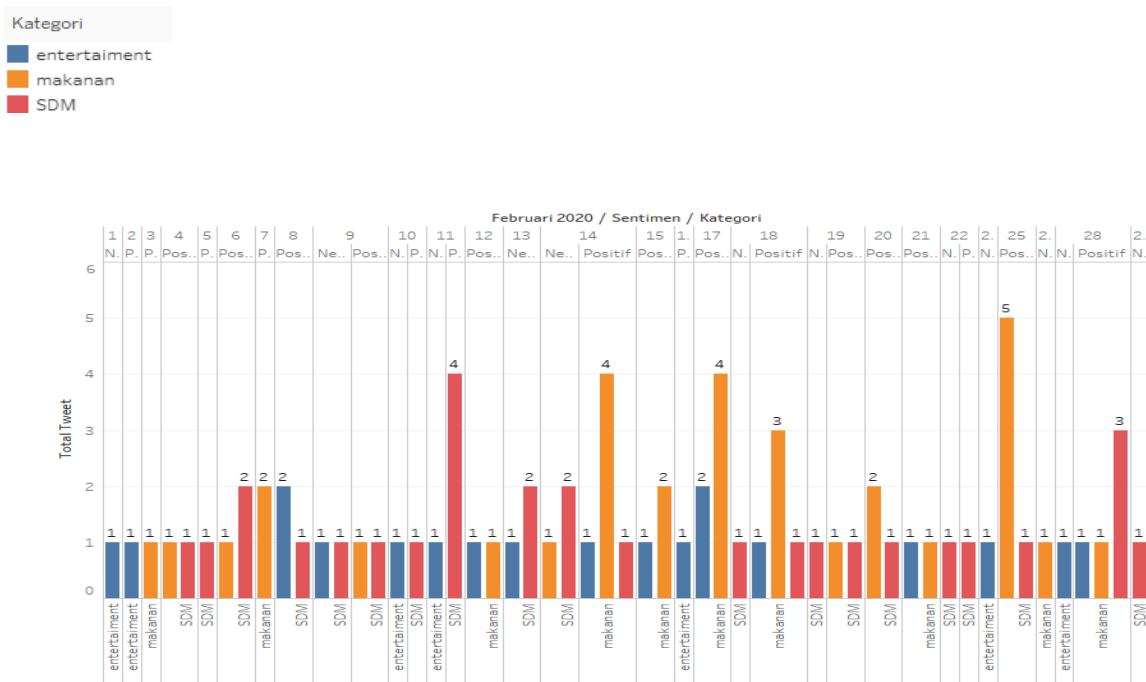
$$Akurasi = \frac{69 + 52}{69 + 52 + 0 + 39} = 75,625\% \quad (12)$$

$$Error\ Rate = \frac{0 + 39}{69 + 39 + 0 + 52} = 24,375\% \quad (13)$$

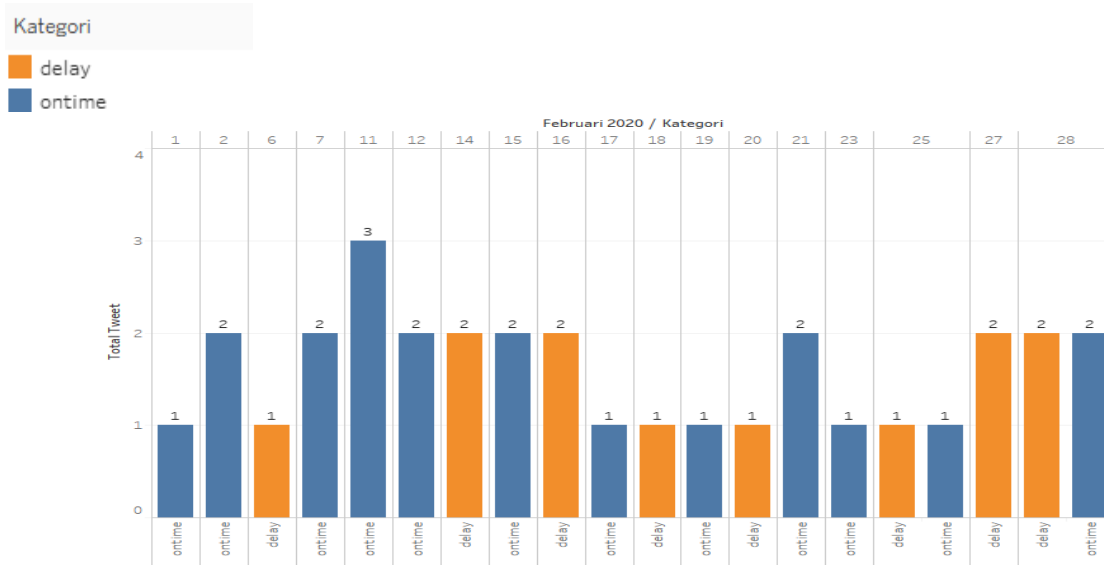
Dari model yang mendapatkan hasil akurasi yang baik yaitu model dari kernel linear, selanjutnya dilakukan prediksi data baru dengan menggunakan data *tweet* pada bulan Februari 2020 sebesar 564 *tweet* untuk dilakukan analisis sentimen berdasarkan kategori pelayanan seperti pelayanan terhadap fasilitas, ketepatan waktu, penilaian terhadap tarif tiket, *customer service*, serta fasilitas layanan GarudaMiles dan *website* Garuda Indonesia yang diberikan oleh maskapai penerbangan Garuda Indonesia, sebagai berikut.



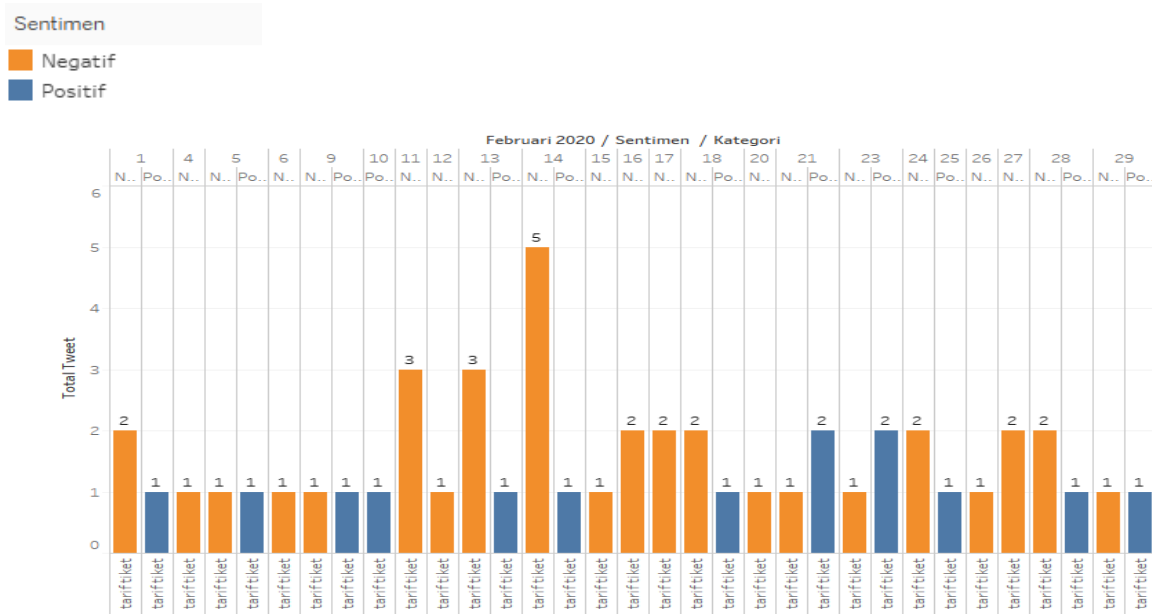
Gambar. 2. Dari gambar di atas menunjukkan jumlah sentimen positif sebesar 291 , sedangkan untuk sentimen negatif sebesar 273 atau dari total 564 data *tweet*.



Gambar. 3. Pada grafik tingkat kepuasan berdasarkan fasilitas di total sentimen sebanyak 80 *tweet*, dengan sentimen positif berjumlah 62 *tweet*, sedangkan negatif berjumlah 18 *tweet*, dimana sentimen positif makanan 30 *tweet*, negatif 2 *tweet*, sentimen positif entertainment 12 *tweet*, negatif 7 *tweet*, sentimen positif SDM 20 *tweet*, negatif 9.



Gambar. 4. Grafik tingkat ketepatan waktu menunjukkan total tweet terkait ketepatan waktu berjumlah 32 *tweet*. Sentimen positif *ontime* adalah 20 *tweet* sedangkan sentimen negatif *delay* sebanyak 12 *tweet*.



4 KESIMPULAN

Pemodelan klasifikasi yang dilakukan ke dalam kelas positif dan kelas negatif menggunakan kernel linear pada metode *support vector machine* (SVM) dengan menggunakan pembagian data latih (*training*) dan data uji (*testing*) sebesar 80% dan 20% memperoleh hasil akurasi sebesar 88,75 % dan *error rate* sebesar 11,25%, sedangkan untuk kernel polynomial hasil akurasi sebesar 75,625% dan *error rate* sebesar 24,375%. Dari penggunaan model prediksi menghasilkan sentimen positif sebanyak 291 *tweet* dan sentimen negatif sebanyak 273 *tweet* dari total keseluruhan *tweet* sebanyak 564 dengan penilaian terhadap layanan Garuda Indonesia, sentimen positif tinggi yaitu pada kategori fasilitas, ketepatan waktu, *customer service*, sedangkan untuk sentimen negatif tinggi yaitu pada kategori tarif tiket, GarudaMiles & *website*.

REFERENSI

- [1] Agarwal, Basant dan Namita Mittal. 2016. *Prominent Feature Extraction for Sentiment Analysis*. Switzerland: Springer.
- [2] Cahyono, Hary. 2019. Membangun Aplikasi Web Data Analysis dengan Framework Django: Mengenal berbagai macam pustaka di Python. Hary Cahyono.
- [3] CFA Institute. 2020. CFA Program Curriculum 2020 Level II Volumes 1-6 Box Set. United State of America: John Wiley & Sons.
- [4] Feldman, Rones dan James Sanger. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. United States of America: Library Of Congress Cataloging.
- [5] Forte Consultancy Group. 2014. *Text Mining - Going Way Beyond Just Listening to the Voice of the Customer*. Dubai: Forte Consultancy.
- [6] Hendrawan, Aria dan Rifi Wijayanti. 2020. Engineering Information and Agricultural Technology in the Global Digital Revolution. Netherlands: CRC Press.
- [7] Prasetyo, Eko. 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: ANDI.
- [8] Purbo, Onno.w, 2019. *Text Mining: Analisis Medsos, Kekuatan Brand & Intelijen di Internet*. Yogyakarta: ANDI.
- [9] Rajput, Dharmendra Singh, Thakur, Ramjeevan Singh, Basha, S. Muzamil. 2019. *Sentiment Analysis and Knowledge Discovery in Contemporary Business*. United State of America: IGI Global.
- [10] Srinivasa. K. G., Siddesh G. M., Srinidhi H. 2018. *Network Data Analytics: A Hands-On Approach for Application Development*. Switzerland: Springer.