

## Analisis Sentimen Terhadap Vaksin Covid-19 di Jejaring Sosial *Twitter* Menggunakan Algoritma *Naïve Bayes*

Rizal Al Habsi<sup>1</sup>, Realdy Agsar Dwi Anggoro<sup>2</sup>, Muhammad Arlanda Valio<sup>3</sup>, Yuni Widiastiwi, S.Kom., M.Si.<sup>4</sup>,  
Nurul Chamidah, S.Kom., M.Kom.<sup>5</sup>

<sup>1,2,3</sup>Program Studi Informatika, Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

rizalhabsi@upnvj.ac.id<sup>1</sup>, realdyada@upnvj.ac.id<sup>2</sup>, muhammadav@upnvj.ac.id<sup>3</sup>, widiastiwi@yahoo.com<sup>4</sup>,  
nurul.chamidah@upnvj.ac.id<sup>5</sup>

**Abstrak.** Semenjak adanya wabah virus corona atau COVID-19 terdapat banyak opini masyarakat pada *twitter* terkait dengan vaksin COVID-19. Penelitian ini bertujuan untuk mengetahui bagaimana opini publik terhadap vaksin COVID-19 di Indonesia pada *twitter* dengan tagar #vaksinCovid19 dan #vaksinCorona. Dalam penelitian ini, opini publik akan dikategorikan menjadi sentimen positif dan negatif serta menggunakan algoritma *Naïve Bayes* untuk proses klasifikasi *tweet*. Proses pengambilan data dimulai pada tanggal 13 Januari hingga 20 Januari 2021 menggunakan teknik *crawling* data dengan memanfaatkan fasilitas API yang disediakan oleh *twitter*. Dari hasil pelabelan data menggunakan 488 *tweet* diperoleh hasil sebanyak 251 *tweet* sentimen positif yang pro terhadap vaksin COVID-19, dan sebanyak 237 *tweet* sentimen negatif yang kontra terhadap vaksin COVID-19. Hasil klasifikasi akan masuk ke tahap pengujian dan evaluasi menggunakan *confusion matrix* untuk melihat performa dari model klasifikasi. Dari hasil pengujian model klasifikasi *naïve bayes* didapatkan hasil akurasi sebesar 82,65%, *recall* sebesar 98% dan *specificity* sebesar 66,67%.

**Kata Kunci:** Analisis Sentimen, Vaksin COVID-19, *Twitter*, *Naïve Bayes*.

### 1 Pendahuluan

Salah satu media sosial *twitter* adalah *platform* yang sangat populer dikalangan pengguna internet. Menurut data *We are Social* pada bulan Januari 2020, *twitter* merupakan salah satu *platform* yang paling banyak dipakai masyarakat Indonesia dengan total pengguna mencapai 10,65 juta pengguna [1].

Sejak merebaknya virus corona, banyak bermunculan tagar viral di *twitter* yang membuktikan bahwa tanggapan masyarakat melalui *tweet* menjadi penting karena keberadaan pandangan publik yang terwujud di masyarakat patut mendapat respons serius dari pemerintah agar dapat membuat tindakan cepat sebelum timbul masalah sosial atau krisis sosial di masyarakat. Saat ini, pemerintah harus mendorong perhatian eksklusif terhadap wabah COVID-19 sehubungan dengan jumlah case positif meningkat secara signifikan. Untuk menangkal penyebaran penyakit COVID-19, negara-negara di seluruh dunia telah bersama-sama berkomitmen untuk mengembangkan vaksin COVID-19. Selama bergulirnya waktu ke waktu sebagian masyarakat banyak sekali pro dan kontra terhadap adanya vaksin COVID-19. Dikutip melalui *website Populicenter.org*, Survei yang dilakukan oleh Populi Center menunjukkan bahwa sekitar 40% masyarakat masih tidak mau menggunakan vaksin yang dikeluarkan oleh pemerintah [2]. Sedangkan survei yang dikeluarkan oleh Kementerian Kesehatan (Kemenkes) menunjukkan bahwa sejumlah 64,8% narasumber menegaskan bahwa telah siap memakai vaksin COVID-19, sebanyak 27,8% narasumber mengungkapkan ragu-ragu, dan terdapat 7,6% yang menolak untuk diberikan vaksin [3]. Hal ini menjadi suatu yang menarik, karena sebuah penelitian yang menunjukkan tingginya sentimen negatif pada sosial media *twitter* terkait dengan COVID-19, viralnya tagar-tagar yang menampilkan reaksi kecewa dari masyarakat, serta banyak sekali pro-kontra terhadap adanya vaksin COVID-19, justru berbanding terbalik dengan reaksi survei yang dikerjakan oleh kementerian kesehatan yang menunjukkan bahwa mayoritas masyarakat tidak takut untuk diberi vaksin COVID-19.

Berdasarkan latarbelakang kasus tersebut, maka perlu suatu pembuktian dengan melakukan penelitian terkait dengan opini masyarakat terhadap vaksin COVID-19. Salah satunya dengan mendeteksi *tweet-tweet* yang mengandung unsur opini terhadap vaksin COVID-19. Penelitian ini menerapkan metode klasifikasi *Naïve Bayes*

untuk mengetahui *tweet* di jejaring sosial *twitter* tentang opini terhadap vaksin COVID-19. Tingkat keakuratan pengklasifikasi Naive Bayes lebih baik daripada model klasifikasi lainnya terlihat pada penelitian analisis sentimen COVID-19 yang membandingkan algoritma, hasil evaluasi algoritma klasifikasi Naive Bayes mendapatkan nilai akurasi yang lebih tinggi sebesar 63,21% jika dibandingkan dengan metode KNN yang hanya sebesar 58,10% [4]. Penggunaan algoritma ini disebabkan karena algoritma klasifikasi *naive bayes* mampu mengklasifikasikan dokumen dalam jumlah yang besar serta menyajikan nilai akurasi yang baik.

## 2 Landasan Teori

### 2.1 Analisis Sentimen

Analisis sentimen yaitu ilmu perhitungan suatu pandangan sentimen dan ungkapan perasaan yang diekspresikan dalam wujud teks [5]. Analisis sentimen yaitu sebuah studi tentang pendapat dan sentimen orang tentang entitas layaknya produk dan layanan dengan wujud teks. Analisis otomatis untuk mengekstrak sebuah pendapat memerlukan pengetahuan mendasar tentang teks alami pada sebuah mesin. Penelitian analisis sentimen mampu dikelaskan pada level dokumen, seperti mengkategorikan data berdasarkan komentar yang menyimpan kontradiksi positif atau negatif [6].

### 2.2 Text Mining

*Text mining* merupakan ilmu yang menggabungkan data mining dan teks analisis dengan menggunakan data tidak terstruktur atau tekstual bersama dengan data terstruktur yang bertujuan untuk eksplorasi, penemuan, dan pemodelan prediktif atau klasifikasi [7].

### 2.3 Praproses Data

Praproses data dilakukan bertujuan untuk menyiapkan data yang belum terstruktur, karena data yang mentah masih memiliki banyak *noise* hingga membentuk data yang terstruktur dan siap diproses untuk tahap selanjutnya dengan harapan untuk menghasilkan model klasifikasi yang baik. Tahapan pada praproses terdiri dari *case folding*, pembersihan data, *stemming*, tokenisasi, *stopword removal* dan normalisasi bahasa.

#### 2.3.1 Case Folding

Case Folding atau transform case merupakan langkah yang dilakukan untuk merubah semua karakter pada data sesuai kebutuhan, seperti merubah huruf besar (uppercase) menjadi huruf kecil (lowercase) atau sebaliknya [8].

#### 2.3.2 Pembersihan Data

Pembersihan data (Cleaning), yaitu proses pembersihan kata-kata atau karakter yang tidak perlu dalam tweet untuk membersihkan noise. Karakter yang dihapus diantaranya karakter URL (<https://>, dsb), emoticon, tagar (#), username (@username), ruang kosong (white space), email dan karakter lainnya [9].

#### 2.3.3 Stemming

Stemming adalah proses menghilangkan kata-kata yang mengandung awalan, akhiran dan imbuhan lainnya dalam dokumen untuk menjadikannya kata-kata dasar [10].

#### 2.3.4 Tokenisasi

Tokenisasi dapat didefinisikan sebagai proses membagi teks pada spasi dan tanda baca menjadi bagian kata serta mengembalikan daftar karakter alfabet dan nonalfabetis [11].

### 2.3.5 Stopword Removal

Stopword removal dapat di definisikan sebagai proses menghilangkan kata pada dokumen yang memiliki makna tidak ada [10].

### 2.3.6 Normalisasi Bahasa

Normalisasi bahasa didasarkan pada proses pengubahan kata tidak baku menjadi kata baku dengan mencocokkan setiap kata dalam kamus bahasa tidak baku dengan semua data dalam Kamus Besar Bahasa Indonesia (KBBI) [12]. Tahap ini dilakukan dengan menggunakan Kamus Alay – *Colloquial Indonesian Lexicon* dengan teknis mencocokkan setiap kata pada *tweet* yang telah di praproses.

## 2.4 Pembobotan Kata

Tujuan tahap ini adalah untuk memberi bobot frekuensi pada setiap kemunculan kata dalam dokumen [10]. Pembobotan kata dilakukan dengan menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF), yaitu sebuah *statistic numeric*, menganalisis terhadap sebuah kata pada dokumen [13]. Berikut rumus persamaan dari perhitungan TF-IDF.

$$TF - IDF = TF \times IDF = TF_{t,d} \times \frac{|D|}{DF_t} \quad (1)$$

Keterangan:

TF<sub>t,d</sub> : Total frekuensi kemunculan kata t pada suatu dokumen d  
 DF<sub>t</sub> : Total dokumen yang menyimpan kata t  
 D : Total seluruh dokumen yang tersedia

## 2.5 Klasifikasi *Naïve Bayes*

*Naïve Bayes* akan menghitung probabilitas sebelumnya dari setiap label, yang ditentukan dengan memeriksa frekuensi setiap label dalam set pelatihan. Setiap fitur pada data dapat menentukan label mana yang harus ditetapkan ke nilai input tertentu. Kontribusi dari setiap fitur kemudian digabungkan dengan probabilitas sebelumnya ini, untuk sampai pada perkiraan kemungkinan untuk setiap label. Label yang perkiraan kemungkinannya paling tinggi kemudian ditetapkan ke nilai input [14]. Berikut adalah rumus persamaan umum dari algoritma *naïve bayes*.

$$P(d) = \frac{P(c) \times P(d|c)}{P(d)} \quad (2)$$

Keterangan:

c : Hipotesis dokumen yang merupakan suatu kelas spesifik  
 d : Dokumen dengan kelas yang belum diketahui  
 P(c|d) : *Posterior*, probabilitas hipotesis c berdasarkan kondisi d (*posterior probability*)  
 P(c) : *Prior*, probabilitas hipotesis c (*prior probability*)  
 P(d|c) : *Likelihood*, probabilitas d berdasarkan kondisi pada hipotesis c  
 P(d) : *Evidence*, probabilitas d

Persamaan rumus yang mengukur nilai probabilitas pada masing masing kelas, seperti persamaan dibawah ini.

$$P(c) = \frac{|doc\ c|}{|document|} \quad (3)$$

Keterangan:

$P(c)$  : Probabilitas kemunculan suatu data yang memiliki kelas  $c$ .  
 $doc\ c$  : Total dari data untuk tiap kelas  $c$ .  
 $|document|$  : Total data dari setiap kelas

Rumus untuk melakukan pengukuran peluang setiap kata dari dokumen yang ada berdasarkan kategori menggunakan persamaan sebagai berikut.

$$P(c) = \frac{Count(W_i, c) + 1}{|c| + |V|} \quad (4)$$

Keterangan:

$P(c)$  : Peluang kata  $W_i$  pada *class*  $c$   
 $Count(W_i, c)$  : Total kemunculan kata  $W_i$  pada *class*  $c$ .  
 $|c|$  : Total keseluruhan kata pada *class*  $c$   
 $|V|$  : Total keseluruhan *term* (kata).

Proses klasifikasi data uji menggunakan persamaan sebagai berikut.

$$c_{MAP} = \underset{c \in V}{\operatorname{arg\,max}} P(c) \prod_i P(W_i|c) \quad (5)$$

Keterangan:

$P(c)$  : Peluang kemunculan suatu dokumen yang memiliki *class*  $c$ .  
 $P(W_i|c)$  : Peluang kata  $W_i$  pada *class*  $c$ .

Dalam pengklasifikasian menggunakan algoritma *Naïve Bayes* terdapat proses pembagian data *tweet* secara *stratified random sampling* menjadi data *training* dan data *testing* yang adil dan seimbang. Langkah pertama adalah melakukan pelatihan sistem model klasifikasi dengan menggunakan data latih sebesar 80%, dan tahap selanjutnya yaitu mengevaluasi model klasifikasi menggunakan data uji sebesar 20% dengan mengacu pada model *training*.

## 2.5 Evaluasi

Nilai target yang benar dari set data pengujian dapat membandingkan jawaban yang diprediksi dengan jawaban yang benar sehingga memerlukan persentase waktu model tersebut tepat, yang secara efektif untuk menilai setiap model dengan akurasi keseluruhannya [11]. Evaluasi sebuah model dimanfaatkan untuk mengetahui nilai performa model yang dilakukan mengikuti hasil klasifikasi menggunakan parameter berdasarkan dari tabel *confusion matrix*. Tabel 1 berikut adalah *confusion matrix* untuk klasifikasi dua *class* [15].

Tabel 1. *Confusion Matrix*

		AKTUAL	
		Positif	Negatif
PREDIKSI	Positif	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Keterangan:

1. *True Positive (TP)*, *tweet* dengan aktual positif dan diprediksi positif
2. *True Negative (TN)*, *tweet* dengan aktual negatif dan diprediksi negatif
3. *False Positive (FP)*, *tweet* dengan aktual negatif namun diprediksi positif.
4. *False Negative (FN)*, *tweet* dengan aktual positif namun diprediksi negatif.

Persamaan rumus menghitung nilai akurasi:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (6)$$

Persamaan rumus menghitung nilai *recall*:

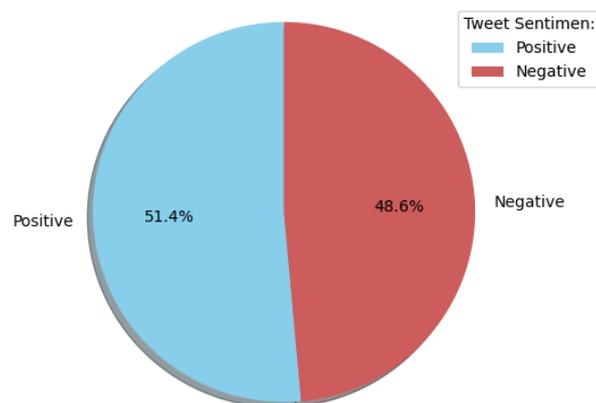
$$recall = \frac{TP}{TP + FN} \quad (7)$$

Persamaan rumus menghitung nilai *specificity*:

$$specificity = \frac{TN}{TN + FP} \quad (8)$$

### 3 Hasil dan Pembahasan

Dataset yang digunakan total sebanyak 488 *tweet* dengan menggunakan kata kunci ‘#VaksinCorona’ dan ‘#VaksinCovid19’. Data diambil pada tanggal 13 Januari sampai 20 Januari 2021 karena pada tanggal tersebut bertepatan dengan pelaksanaan pemberian vaksin pertama kali oleh Presiden RI Joko Widodo. Data *tweet* akan diberikan label dengan 2 kategori, yaitu positif yang merupakan pro terhadap vaksin COVID-19 dan negatif yang merupakan kontra terhadap vaksin COVID-19. Proses dilakukan dengan cara manual menggunakan 2 orang penilai berdasarkan sudut pandang seseorang dalam menilai suatu bacaan. Teknik pelabelan data menggunakan nilai kesetaraan *kappa value* untuk mengukur tingkat kesepakatan antar kelas label. Hasil dari total pelabelan data sebanyak 488 *tweet* diperoleh sebanyak 251 *tweet* positif dan 237 *tweet* negatif dengan proporsi data *tweet* dapat dilihat pada Gambar 1 berikut.



**Gambar 1.** Hasil Pelabelan Data

Berdasarkan hasil pelabelan data *tweet* tersebut diketahui bahwa sentimen positif mendapatkan hasil sebesar 51.4% sedangkan sentimen negatif mendapatkan hasil sebesar 48.6%. Sentimen masyarakat terhadap vaksin COVID-19 didominasi oleh sentimen positif, namun masih banyak juga masyarakat yang meragukan vaksin COVID-19 dengan banyaknya sentiment negatif sebesar 48.6%.

Setelah data diberi label positif dan negatif, data akan masuk ke tahap praproses data untuk menghilangkan *noise* sehingga membentuk data yang terstruktur dan siap diproses. Selanjutnya data akan diberikan bobot dengan menggunakan metode TF-IDF yang bertujuan untuk memberikan bobot atau nilai pada setiap kata dalam *tweet*

yang telah melewati tahap praproses data. Berikut ini Tabel 2 adalah contoh penerapan TF-IDF dengan menggunakan 4 sampel dokumen *tweet* data latih.

**Tabel 2.** Sampel Data Latih

Sampel Data Latih		
Data Latih		Label
D1	['uang', 'bayar', 'denda', 'kebal', 'virus', 'vaksin']	Positif
D2	['ayo', 'dukung', 'sukses', 'vaksin', 'lawan', 'pandemi']	Positif
D3	['awas', 'vaksin', 'cina', 'tolak', 'kecuali']	Negatif
D4	['tolak', 'vaksin', 'sinovac', 'cina', 'tidak', 'pidana']	Negatif

Kemudian data sampel diatas akan dihitung nilai bobot dengan menggunakan persamaan rumus TF-IDF poin (1) seperti pada Tabel 3 berikut ini.

**Tabel 3.** Hasil Perhitungan TF-IDF

Term	Dokumen				DF	IDF	TF-IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
uang	1	0	0	0	1	0,602	0,602	0	0	0
bayar	1	0	0	0	1	0,602	0,602	0	0	0
denda	1	0	0	0	1	0,602	0,602	0	0	0
kebal	1	0	0	0	1	0,602	0,602	0	0	0
virus	1	0	0	0	1	0,602	0,602	0	0	0
vaksin	1	1	1	1	4	0	0	0	0	0
ayo	0	1	0	0	1	0,602	0	0,602	0	0
dukung	0	1	0	0	1	0,602	0	0,602	0	0
sukses	0	1	0	0	1	0,602	0	0,602	0	0
lawan	0	1	0	0	1	0,602	0	0,602	0	0
pandemi	0	1	0	0	1	0,602	0	0,602	0	0
awas	0	0	1	0	1	0,602	0	0	0,602	0
cina	0	0	1	1	2	0,301	0	0	0,301	0,301
tolak	0	0	1	1	2	0,301	0	0	0,301	0,301
kecuali	0	0	1	0	1	0,602	0	0	0,602	0
sinovac	0	0	0	1	1	0,602	0	0	0	0,602
tidak	0	0	0	1	1	0,602	0	0	0	0,602
pidana	0	0	0	1	1	0,602	0	0	0	0,602

Dalam proses klasifikasi, data akan melewati 2 tahap yaitu tahap latih (*training*) dan tahap uji (*testing*). Data akan terlebih dahulu melewati proses pembagian data menggunakan metode *hold out* dimana data akan dibagi secara acak yang adil dan seimbang menjadi data latih (*training*) dan data uji (*testing*) menggunakan 1591 kata dari total 488 data *tweet*. Perbandingan data latih dengan data uji yaitu sebesar 80:20 secara acak dan seimbang seperti pada Tabel 4 berikut.

**Tabel 4.** Pembagian Data

	Label Positif	Label Negatif	Total Jumlah
Data Latih ( <i>Training</i> )	201	189	390
Data Uji ( <i>Testing</i> )	50	48	98
Total Jumlah	251	237	488

Data yang diperoleh dari proses pembobotan seperti contoh pada Tabel 3 digunakan sebagai data latih dan acuan untuk membentuk model klasifikasi menggunakan algoritma *naïve bayes*, kemudian akan dicari nilai probabilitas kategori dan probabilitas setiap kata dari data pelatihan. Setelah melakukan proses *training* tahap selanjutnya adalah melakukan tahap uji (*testing*). Pada tahap ini dilakukan pengujian dengan menggunakan data uji yang mengacu model *naïve bayes* yang telah dibentuk pada tahap latih (*training*). Tahap uji (*testing*) yang mengacu pada model *training* dengan menggunakan algoritma *Naïve Bayes* akan menghitung nilai probabilitas untuk setiap





## 4 Kesimpulan dan Saran

Kesimpulan yang dapat diambil dari penelitian analisis sentiment terhadap vaksin COVID-19 di jejaring sosial *twitter* adalah data yang digunakan merupakan *tweet* yang diperoleh dari media sosial *twitter* pada tanggal 13 Januari 2021 sampai dengan 20 Januari 2021 sebanyak 488 dokumen *tweet* yang terdiri dari 2 kategori kelas masing-masing sebanyak 251 positif dan 237 negatif. Hasil dari pelabelan opini masyarakat terhadap vaksin COVID-19 memiliki presentase 51,4% sentimen positif yang pro terhadap vaksin dan 48,6% sentimen negatif yang kontra terhadap vaksin. Berdasarkan hasil presentase tersebut didapat kesimpulan bahwa opini masyarakat terhadap vaksin COVID-19 cenderung seimbang antara opini positif maupun negatif. Banyak masyarakat yang sudah mau untuk divaksin namun tidak sedikit pula masyarakat yang masih ragu bahkan menolak untuk divaksin. Dalam proses visualisasi pun terlihat bahwa kata “vaksin” justru sering muncul dalam sentimen negatif sebanyak 294 dibandingkan sentimen positif yang hanya muncul sebanyak 282 kali.

Klasifikasi opini masyarakat terhadap vaksin COVID-19 dengan menggunakan data *tweet* dilakukan dengan memberikan label sentimen positif dan negatif. Setelah diberikan label, data *tweet* masuk ke tahap praproses dengan tahapan *case folding*, *cleaning*, *stemming*, tokenisasi, *stopword removal*, dan normalisasi bahasa. Data hasil praproses akan diberi bobot dengan menggunakan metode TF-IDF untuk selanjutnya data tersebut digunakan untuk membuat model klasifikasi dengan perbandingan data 80% untuk *training* dan 20% untuk *testing*. Performa yang dihasilkan oleh model klasifikasi dengan menggunakan 488 *tweet* mendapatkan hasil nilai akurasi sebesar 82,65%, *recall* sebesar 98%, dan *specificity* sebesar 66,67%.

Adapun saran yang diberikan untuk pengembangan penelitian ini dimasa mendatang agar menjadi lebih baik yaitu dalam melakukan proses normalisasi kata, diharapkan untuk menambah kosa kata dalam kamus normalisasi sehingga semakin banyak *slang word* yang akan dinormalisasi. Untuk penelitian berikutnya untuk memperhatikan tahapan pra proses dengan baik, karena pada penelitian ini masih terdapat *noise* yang dapat mengurangi performa dari model klasifikasi. Penelitian selanjutnya diharapkan untuk mencoba algoritma klasifikasi yang lain seperti *Support Vector Machine*.

## Referensi

- [1] S. Kemp, “Digital 2020: Indonesia,” *We Are Social*, 2020. <https://datareportal.com/reports/digital-2020-indonesia> (accessed Dec. 28, 2020).
- [2] N. F. Afifah, “Survei Nasional: Evaluasi Umum Pemerintahan Joko Widodo - KH. Ma’ruf Amin Dan Penerimaan Terhadap Vaksin COVID-19,” *Populi Center*, 2020. <https://populicenter.org/evaluasi-umum-pemerintahan-joko-widodo-kh-maruf-amin-dan-penerimaan-terhadap-vaksin-covid-19/> (accessed Jan. 05, 2020).
- [3] Kementerian Kesehatan Republik Indonesia, ITAGI, WHO, and UNICEF, “Survei Penerimaan Vaksin COVID-19 di Indonesia,” 2020.
- [4] M. Syarifuddin, “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn,” *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [5] 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.
- [6] B. Agarwal and N. Mittal, *Prominent Feature Extraction for Sentiment Analysis*. Berlin: Springer International Publishing, 2016.
- [7] D. Jared, “Big Data, Data Mining, and Machine Learning Value Creation for Business Leaders and Practitioners,” 2014.
- [8] I. Y. Anggraini, S. Sucipto, and R. Indriati, “Cyberbullying Detection Modelling at Twitter Social Networking,” *JUITA J. Inform.*, vol. 6, no. 2, p. 113, 2018.
- [9] C. Institute, *CFA Program Curriculum 2020 Level II Volumes 1-6 Box Set*. United State of America: John Wiley & Sons, 2020.
- [10] D. Sarkar, *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data*, vol. 32, no. 1. Apress, 2016.
- [11] B. Bengfort, R. Bilbro, and T. Ojeda, *Applied Text analysis with Python*, vol. 53, no. 9. 2018.
- [12] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [13] O. W. Purbo, *Text Mining: Analisis Medsos, Kekuatan Brand & Intelijen di Internet*. Yogyakarta: ANDI, 2019.

- [14] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.
- [15] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.