

Analisis Sentimen Terkait Layanan Gofood dan Grabfood pada Media Sosial Twitter Menggunakan Algoritma *Support Vector Machine* (SVM)

Realdy Agsar Dwi Anggoro¹, Rizal Al Habsi², Muhammad Arlanda Valio³, Yuni Widiastiwi, S.Kom., M.Si.⁴,
Nurul Chamidah, S.Kom., M.Kom.⁵

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta

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

realdyada@upnvj.ac.id¹, rizalhabsi@upnvj.ac.id², muhammadav@upnvj.ac.id³, widiastiwi@yahoo.com⁴,
nurul.chamidah@upnvj.ac.id⁵

Abstrak. Kepopuleran media sosial *twitter* di Indonesia menyebabkan cepatnya penyebaran informasi melalui dunia maya ini. Informasi tersebut berupa opini terhadap suatu layanan yang dapat dianalisis sentimennya agar dapat memahami makna dan emosi ini yang terkandung dalam kalimat opini tersebut. Penelitian bertujuan untuk mengetahui bagaimana opini publik mengenai layanan gofood dan grabfood pada media sosial *twitter* dengan kata kunci gofood dan grabfood yang diambil selama 6 hari pada tanggal 23 Februari sampai dengan 27 Februari 2021 dan tanggal 5 Maret 2021. Data tersebut kemudian akan dilabeli menjadi kelas positif atau kelas negatif serta diklasifikasikan dengan algoritma *Support Vector Machine*. Hasil evaluasi model pertama memiliki akurasi sebesar 80,18%, *recall* sebesar 100%, dan *specificity* sebesar 14%. Karena data pada penelitian ini tidak seimbang, maka pada model kedua diterapkan metode *undersampling* pada data latih untuk mengatasi hal tersebut. Hasil evaluasi model kedua memiliki akurasi sebesar 79,26%, *recall* sebesar 86,23%, dan *specificity* sebesar 56%.

Kata Kunci: twitter, analisis sentimen, *Support Vector Machine*, *undersampling*.

1 Pendahuluan

Perkembangan ilmu pengetahuan dan teknologi yang pesat seperti sekarang ini banyak memunculkan teknologi-teknologi baru pada perangkat terkini yang mengakibatkan meningkatnya pengguna internet di dunia termasuk di Indonesia. Berdasarkan laporan dari *WeareSocial Hootsuite* menunjukkan bahwa 4,5 miliar orang di seluruh dunia telah menggunakan internet yang berarti bahwa lebih dari 60% penduduk di dunia adalah pengguna internet. Dari 4,5 miliar atau lebih dari 60% pengguna internet di dunia, ternyata 3,8 miliar orang diantaranya sudah menggunakan sosial media. Di Indonesia, pengguna sosial media aktif sebanyak 160 juta dari jumlah penduduk di Indonesia yaitu 272,1 juta penduduk [1].

Salah satu media sosial yang populer di Indonesia adalah Twitter. Data laporan dari *Wearesocial Hootsuite* pengguna media sosial twitter di Indonesia mencapai 56% dari jumlah penduduk di Indonesia [1]. Kepopuleran twitter menyebabkan layanan ini sering digunakan oleh pengguna untuk berbagai keperluan, misalnya seperti sarana pembelajaran, sarana untuk kampanye, dan sering juga digunakan sebagai sarana untuk menyampaikan opini terhadap suatu layanan atau produk.

Selain perkembangan di bidang komunikasi, layanan dibidang transportasi pun mengalami perkembangan contohnya seperti inovasi yang dilakukan oleh perusahaan gojek dan grab dengan meluncurkan layanan pesan antar gofood dan grabfood. Layanan tersebut sangat populer di Indonesia dan mempengaruhi gaya hidup masyarakat sekarang ini. Layanan pesan antar gofood maupun grabfood pastinya memiliki kelebihan dan kekurangannya. Setiap pelanggan memiliki pengalaman yang berbeda terhadap pelayanan yang diberikan. Sebagian pengguna layanan tersebut menyampaikan opininya melalui media sosial seperti twitter.

Berlatarkan hal tersebut, akan dilakukan analisis sentimen pengguna layanan gofood dan grabfood pada media sosial twitter. *Tweet* pengguna layanan gofood dan grabfood yang diambil dari media sosial twitter akan diklasifikasikan menjadi kelas positif atau kelas negatif. Pada penelitian Nurrun Muchammad Shiddieqy Hadna

dan kawan-kawan mengenai perbandingan metode klasifikasi untuk analisis sentimen menunjukkan bahwa hasil analisis sentimen yang menggunakan algoritma *Support Vector Machine* memiliki hasil performa yang lebih baik dibandingkan dengan metode klasifikasi yang lain [2]. Oleh karena itu, pada penelitian ini metode klasifikasi yang digunakan adalah *Support Vector Machine* (SVM).

2 Landasan Teori

2.1 Analisis Sentimen

Analisis sentimen merupakan proses untuk menganalisis data teks secara otomatis untuk memperoleh informasi emosional pada suatu kalimat opini [3]. Tujuan dari analisis sentimen adalah untuk mengelompokkan teks pada suatu kalimat ataupun dokumen dan selanjutnya menentukan apakah pendapat yang diungkapkan pada kalimat atau dokumen itu positif atau negatif [4].

2.2 Text Mining

Text mining adalah suatu proses untuk memperoleh data teks untuk menemukan suatu informasi yang sebelumnya tidak diketahui dari sumber data teks yang berbeda [4]. Berdasarkan kutipan tersebut, *text mining* bertujuan untuk menemukan suatu informasi yang dapat mewakili isi dari sumber data teks.

2.3 Praproses Data

Praproses data adalah sebuah proses pembersihan data mentah dengan cara membuang data yang tidak diperlukan agar sistem lebih mudah untuk memprosesnya [5]. Dalam analisis sentimen, tahap praproses ini sangat penting karena data yang diperoleh dari media sosial seperti *twitter* banyak terdapat kata-kata tidak baku dan tidak terstruktur serta memiliki banyak *noise*. Praproses data pada penelitian ini meliputi *case folding*, pembersihan data, *stemming*, tokenisasi, normalisasi bahasa, dan *stopword removal*.

2.3.1 Case Folding

Case folding adalah sebuah proses untuk mengubah semua huruf pada setiap dokumen *tweet* menjadi huruf kecil [6]. Proses ini bertujuan agar tidak ada kata yang tidak terdeteksi karena terdapat perbedaan format huruf.

2.3.2 Pembersihan Data

Pada proses pembersihan data ini, setiap dokumen *tweet* akan dibersihkan dengan cara menghapus beberapa *value* pada data *tweet* seperti menghapus RT, menghapus angka, menghapus hashtag, dan menghapus URL.

2.3.3 Stemming

Stemming merupakan suatu proses pemotongan imbuhan pada suatu kata untuk memperoleh kata dasar dari kata berimbuhan [7].

2.3.4 Tokenisasi

Tokenisasi merupakan proses pemotongan suatu kalimat berdasarkan karakter spasi menjadi pecahan kecil yang pada umumnya berupa kata (*token*) [6].

2.3.5 Normalisasi Bahasa

Normalisasi bahasa merupakan proses untuk mengkonversi kata tidak baku menjadi kata baku berdasarkan Kamus Besar Bahasa Indonesia (KBBI) dengan cara mencocokkan setiap kata pada kamus Bahasa tidak baku dengan seluruh data [8]. Pada proses ini setiap dokumen *tweet* yang telah di tokenisasi akan dicocokkan dengan setiap kata pada Kamus Alay - *Colloquial Indonesian Lexicon* yang telah di *custom*.

2.3.6 Stopword Removal

Stopword removal merupakan proses untuk membuang kata yang sering muncul namun tidak memiliki arti penting dan tidak memiliki pengaruh terhadap kandungan pada suatu kalimat [9].

2.4 Pembobotan Kata

Term Frequency-Inverse Document Frequency merupakan pengukuran statistik yang berfungsi untuk mengevaluasi pentingnya suatu kata dalam sebuah dokumen. Semakin kecil bobot pada suatu kata apabila kata tersebut muncul pada banyak dokumen dan akan semakin besar apabila kata tersebut sering muncul dalam suatu dokumen [10]. Berikut ini ada persamaan untuk menghitung bobot *term* dengan menggunakan metode TF-IDF:

$$w_{t,d} = tf_{t,d} \times IDF_t \quad (1)$$

Dimana nilai IDF_t diperoleh dengan menggunakan persamaan sebagai berikut:

$$IDF_t = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Keterangan:

- $w_{t,d}$: Bobot kata t pada dokumen d
- $tf_{t,d}$: frekuensi kata t pada dokumen d
- N : Jumlah seluruh dokumen
- df_t : Jumlah dokumen yang mengandung kata t

2.5 Klasifikasi Data

Klasifikasi adalah proses untuk mengidentifikasi suatu objek untuk masuk ke suatu kelas berdasarkan kriteria yang ada pada objek tersebut [11]. Pada tahap pengklasifikasian ini terdapat proses untuk membagi data *tweet* menjadi data latih dan data uji. Pada penelitian ini sebanyak 70% data *tweet* digunakan untuk pembuatan model klasifikasi (data latih). Sedangkan 30% data *tweet* digunakan untuk menguji model klasifikasi (data uji). Untuk menangani data yang *imbalance* pada tahap pembuatan model, pada penelitian ini menggunakan metode untuk menangani *imbalance* data yaitu *undersampling*. Pada penelitian ini, algoritma klasifikasi yang digunakan adalah *Support Vector Machine*.

2.5.1 Support Vector Machine

Support Vector Machine adalah suatu algoritma yang mampu menyelesaikan permasalahan secara *linear separable* maupun permasalahan *non-linear separable* [8]. Dalam melakukan klasifikasi algoritma *Support Vector Machine* menggunakan *hyperplane* untuk memisahkan 2 kelas. Konsep dasar klasifikasi pada algoritma *Support Vector Machine* adalah mencari *hyperplane* terbaik yang memiliki jarak antar kelas semaksimal mungkin [3]. Pada umumnya rumus untuk mencari *hyperplane* untuk semua dimensi adalah seperti persamaan di bawah ini:

$$y = x_1 w_1 + x_2 w_2 + \dots + x_n w_n + b = 0 \quad (3)$$

Keterangan:

$y \in \{-1, +1\}$: nilai target dari himpunan data x
 x_n : $[x_1, x_2, \dots, x_n]$ adalah vektor baris berdimensi k (banyaknya fitur).
 w_n : $[w_1, w_2, \dots, w_n]$ adalah vektor baris yang menjadi parameter bobot.
 b : bias atau *error*.

Dua pemisah sejajar dengan *hyperplane* yang didapatkan dari *support vector* pada masing-masing kelas yang dirumuskan sebagai berikut:

$$(x_n w_n) + b = 1, \text{ untuk kelas positif} \quad (4)$$

$$(x_n w_n) + b = -1, \text{ untuk kelas negatif} \quad (5)$$

Pada *Support Vector Machine*, nilai margin antara bidang pembatas dimaksimalkan sehingga:

$$\frac{2}{\|w\|} \quad (6)$$

Untuk mendapatkan *hyperplane* terbaik didapat dengan cara memaksimalkan $\frac{2}{\|w\|}$ atau sama dengan meminimumkan $\|w\|^2$, maka pencarian *hyperplane* terbaik dengan memaksimalkan nilai margin dapat diformulasikan menjadi:

$$\frac{1}{2} \|w\|^2 \quad (7)$$

dengan syarat:

$$y_i(x_i w + b) \geq 1 \quad (8)$$

Dalam menyelesaikan permasalahan *non-linear separable*, algoritma *Support Vector Machine* memanfaatkan konsep kernel dengan mencari *hyperplane* yang memiliki margin antar kelas semaksimal mungkin. Berdasarkan penelitian Nur Fitriyah dan kawan-kawannya menunjukkan bahwa model yang menggunakan kernel Gaussian RBF memiliki performa yang lebih bagus dibandingkan dengan model yang menggunakan kernel linear [3]. Oleh karena itu, pada penelitian ini fungsi kernel yang digunakan adalah kernel Gaussian RBF yang ditunjukkan pada persamaan sebagai berikut:

$$K(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right) \quad (9)$$

2.6 Evaluasi

Evaluasi model klasifikasi pada umumnya menggunakan data uji dengan ukuran tertentu, dimana data uji tersebut tidak digunakan dalam proses untuk melatih model klasifikasi. Suatu model klasifikasi memiliki output sebuah prediksi kelas dari suatu data. Klasifikasi biner merupakan klasifikasi dengan output dua kelas. Kedua kelas tersebut pada umumnya direpresentasikan dengan $\{0,1\}$, $\{+1,-1\}$ atau $\{\text{positif}, \text{negatif}\}$ [12]. Salah satu metode dalam mengevaluasi model adalah dengan menggunakan *confusion matrix* seperti pada Tabel 1 berikut ini:

Tabel 1. *Confusion Matrix*

		Hasil Prediksi	
		Positif	Negatif
Hasil yang diharapkan	Positif	True Positif (TP)	False Negatif (FN)
	Negatif	False Positif (FP)	True Negatif (TN)

Berdasarkan Tabel 1 tersebut, dalam melakukan klasifikasi data terdapat empat kemungkinan yang terjadi, diantaranya:

1. True Positif (TP), data yang diprediksi positif dan kenyataannya positif.

2. True Negatif (TN), data yang diprediksi negatif dan kenyataannya negatif.
3. False Positif (FP), data yang diprediksi positif dan kenyataannya negatif.
4. False Negatif (FN), data yang diprediksi negatif dan kenyataannya positif.

Persamaan untuk menghitung nilai akurasi:

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (10)$$

Persamaan untuk menghitung nilai *Sensitivity (recall)*:

$$Sensitivity (recall) = \frac{TP}{TP+FN} \quad (11)$$

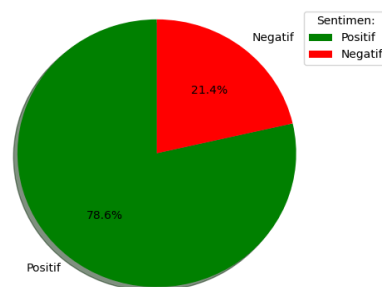
Persamaan untuk menghitung nilai *Specificity*:

$$Specificity = \frac{TN}{TN+FP} \quad (12)$$

3 Hasil dan Pembahasan

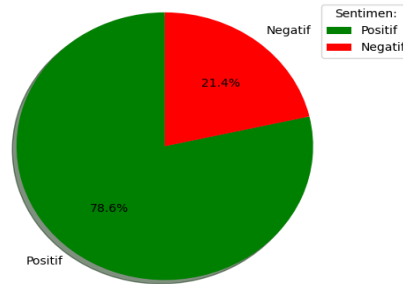
Pada penelitian ini data diperoleh dengan cara *crawling* dari media sosial *twitter* dengan kata kunci pencarian yang digunakan yaitu *gofood* dan *grabfood* yang diambil selama 6 hari pada tanggal 23 Februari sampai dengan 27 Februari 2021 dan tanggal 5 Maret 2021. Dari proses *crawling* total data yang berhasil diambil pada rentang waktu 6 hari tersebut sebanyak 1950 dokumen *tweet* yang terdiri dari 1100 dokumen *tweet* tentang *gofood* dan 850 dokumen *tweet* tentang *grabfood*. Data-data yang telah diperoleh tersebut di *filter* terlebih dahulu dengan cara menyeleksi dokumen *tweet* yang dapat digunakan pada penelitian ini serta mengecek duplikasi pada dokumen *tweet* tersebut. Total data yang diperoleh setelah melewati proses *filter* sebanyak 722 dokumen *tweet* dengan rincian 386 dokumen *tweet* tentang *gofood* dan 336 dokumen *tweet* tentang *grabfood*.

Setelah itu, dokumen *tweet* dilabeli secara manual dengan menggunakan 2 orang penilai. Pada proses ini, tiap dokumen *tweet* akan dilabelkan menjadi kelas positif atau kelas negatif berdasarkan perspektif masing-masing penilai mengenai makna kalimat *tweet* yang terkandung pada data *gofood* maupun data *grabfood*. Hasil dari tahap pelabelan data ini untuk data tentang *gofood* dengan jumlah 386 dokumen *tweet* diperoleh sebanyak 292 dokumen *tweet* positif dan 94 dokumen *tweet* negatif dengan proporsi dokumen *tweet* positif dan dokumen *tweet* negatif dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1. Visualisasi Hasil Pelabelan Data Gofood

Sedangkan untuk data tentang *grabfood* dengan jumlah 336 dokumen *tweet* diperoleh sebanyak 264 dokumen *tweet* positif dan 72 dokumen *tweet* negatif dengan proporsi dokumen *tweet* positif dan dokumen *tweet* negatif dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2. Visualisasi Hasil Pelabelan Data Gofood

Berdasarkan hasil pelabelan data tersebut dapat diketahui bahwa baik sentimen pengguna layanan gofood maupun sentimen pengguna layanan grabfood pada media sosial twitter didominasi oleh sentimen positif, sehingga dapat disimpulkan berdasarkan hasil tersebut baik layanan gofood maupun grabfood sama-sama memiliki pelayanan yang bagus di mata penggunanya berdasarkan data yang diperoleh pada penelitian ini yang bersumber dari media sosial twitter.

Setelah diberi label, dokumen *tweet* akan dibersihkan dengan cara membuang data yang tidak sesuai dengan tujuan untuk meminimalisir *noise* sehingga diharapkan dapat menghasilkan model klasifikasi dengan performa yang bagus. Selanjutnya akan masuk ke tahap pembobotan kata, pada tahap ini data yang akan dibobotkan berjumlah 722 dokumen *tweet* dari hasil penggabungan 2 data, yaitu data gofood yang berjumlah 386 dokumen *tweet* dan data grabfood yang berjumlah 336 dokumen *tweet*. Hasil dari tahap ini diperoleh fitur sebanyak 1616 kata dari 722 dokumen *tweet*. Tabel 2 di bawah ini adalah contoh dari penerapan pembobotan kata dengan menggunakan metode TF-IDF dari 3 dokumen sampel yang telah dipraproses dengan jumlah fitur sebanyak 14 kata.

Tabel 2. Dokumen Sampel

Dokumen <i>Tweet</i>	
D1	['hujan', 'pesan', 'gofood']
D2	['suka', 'benci', 'kadang', 'pesan', 'grabfood', 'makan', 'hujan']
D3	['allah', 'lupa', 'makan', 'enak', 'pikir', 'malam', 'pesan', 'gofood', 'allah', 'curhat']

Kemudian dari dokumen di atas akan dihitung nilai TF, IDF, N, dan DF nya seperti Tabel 3 berikut ini.

Tabel 3. Perhitungan TF-IDF

Kata	TF			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
Hujan	1	1	0	2	0,176	0,176	0,176	0
Pesan	1	1	1	3	0	0	0	0
Gofood	1	0	1	2	0,176	0,176	0	0,176
Suka	0	1	0	1	0,477	0	0,477	0
Benci	0	1	0	1	0,477	0	0,477	0

Kadang	0	1	0	1	0,477	0	0,477	0
Grabfood	0	1	0	1	0,477	0	0,477	0
Makan	0	1	1	2	0,176	0	0,176	0,176
Allah	0	0	2	1	0,477	0	0	0,954
Lupa	0	0	1	1	0,477	0	0	0,477
Enak	0	0	1	1	0,477	0	0	0,477
Pikir	0	0	1	1	0,477	0	0	0,477
Malam	0	0	1	1	0,477	0	0	0,477
Curhat	0	0	1	1	0,477	0	0	0,477

Setelah seluruh dokumen *tweet* dibobotkan maka akan masuk ke tahap klasifikasi data. Pada tahap ini terlebih dahulu data dibagi dengan metode *hold out estimation* dengan proporsi perbandingan data latih dan data uji sebesar 70:30 secara acak dan seimbang seperti pada Tabel 4 berikut ini.

Tabel 4. Pembagian Data

Klasifikasi	Positif	Negatif	Jumlah
Data Latih	389	116	505
Data Uji	167	50	217
Jumlah	556	166	722

Setelah proses pembagian data, data latih akan digunakan untuk membuat model klasifikasi dengan menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel Gaussian RBF. Sedangkan data uji akan digunakan untuk mengevaluasi performa dari model yang telah dibuat. Berdasarkan model yang telah dibuat didapatkan hasil evaluasi terhadap model tersebut dalam bentuk *confusion matrix* untuk mengukur performa dari model tersebut seperti Tabel 5 berikut ini.

Tabel 5. Hasil Evaluasi Model

		Hasil Prediksi	
		Positif	Negatif
Hasil yang diharapkan	Positif	TP (167)	FN (0)
	Negatif	FP (43)	TN (7)

Berdasarkan Tabel 5 di atas, maka performa model *Support Vector Machine* (SVM) dengan indikator penilaiannya yaitu akurasi, *sensitivity*, dan *specificity* dapat diperoleh dengan perhitungan seperti Tabel 6 berikut ini.

Tabel 6. Hasil Perhitungan Evaluasi Model Pertama

<i>Measure</i>	Perhitungan	Hasil
<i>Akurasi</i>	$\frac{167 + 7}{167 + 0 + 43 + 7}$	80,18 %
<i>Sensitivity</i>	$\frac{167}{167 + 0}$	100 %
<i>Specificity</i>	$\frac{7}{7 + 43}$	14 %

Hasil dari model pertama dapat dikatakan cukup bagus dikarenakan memiliki akurasi sebesar 80,18% dan *sensitivity* sebesar 100% akan tetapi *specificity* nya kecil yaitu sebesar 14%. Hal itu dikarenakan pada saat membuat model SVM dengan data latih, distribusi kelas pada data tersebut tidak merata atau *imbalance* data seperti yang dapat dilihat pada Tabel 4, terdapat *imbalance* data yang dimana data dengan kelas positif lebih banyak dibandingkan kelas negatif. Oleh karena itu, untuk menangani hal tersebut digunakan metode *imbalance* data yaitu *undersampling* agar distribusi kelas menjadi seimbang. Berikut ini adalah jumlah pembagian data setelah diterapkan metode *undersampling* pada data latih.

Tabel 7. Pembagian Data *Undersampling*

Klasifikasi	Positif	Negatif	Jumlah
Data Latih	116	116	232

Berdasarkan model SVM kedua yang telah dibuat dengan data latih yang telah diterapkan *undersampling* didapatkan hasil evaluasi terhadap model kedua tersebut dalam bentuk *confusion matrix* untuk mengukur performa dari model kedua tersebut seperti pada Tabel 8 berikut ini.

Tabel 8. Hasil Evaluasi Model *Undersampling*

		Hasil Prediksi	
		Positif	Negatif
Hasil yang diharapkan	Positif	TP (144)	FN (23)
	Negatif	FP (22)	TN (28)

Berdasarkan Tabel 8 di atas, maka performa model SVM kedua dapat dilihat seperti pada Tabel 9 berikut ini.

Tabel 9. Perhitungan Evaluasi Model Kedua

<i>Measure</i>	Perhitungan	Hasil
<i>Akurasi</i>	$\frac{144 + 28}{144 + 22 + 23 + 28}$	79,26 %
<i>Sensitivity</i>	$\frac{144}{144 + 23}$	86,23 %

- b) Dalam mengklasifikasikan data hasil analisis sentimen layanan gofood dan grabfood, performa awal dari model *Support Vector Machine* (SVM) yang dibuat dengan data latih yang belum di *undersampling* memiliki akurasi sebesar 80,18%, *recall* sebesar 100%, dan *specificity* sebesar 14%. Sedangkan model SVM dengan data latih yang sudah diterapkan *undersampling* memiliki akurasi sebesar 79,26%, *recall* sebesar 86,23%, dan *specificity* sebesar 56%.
- c) Metode *undersampling* terbukti dapat menangani masalah data *imbalance* dalam membuat model dikarenakan dapat meningkatkan akurasi kelas minoritas yang dalam penelitian ini adalah kelas negatif yang sebelumnya hanya 14% menjadi 56%.

Referensi

- [1] Hootsuite, "Indonesian Digital Report 2020," 2020. <https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2020/>.
- [2] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, 2016.
- [3] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020.
- [4] D. Rustiana and N. Rahayu, "Analisis sentimen pasar otomotif mobil:," *J. SIMETRIS*, vol. 8, no. 1, pp. 113–120, 2017.
- [5] S. Mujilawati, "Pre-Processing Text Mining Pada Data Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, 2016.
- [6] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, 2018.
- [7] D. Wahyudi, T. Susyanto, and D. Nugroho, "Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia," *J. Ilm. SINUS*, vol. 15, no. 2, pp. 49–56, 2017.
- [8] 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.
- [9] A. F. Hidayatullah, "Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016.
- [10] R. Melita, V. Amrizal, H. B. Suseno, and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018.
- [11] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 180–184, 2017.
- [12] D. T. Lukmana, S. Subanti, and Y. Susanti, "Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter," *Semin. Nas. Penelit. Pendidik. Mat. 2019 UMT*, 2019.