

COMPARISON OF RANDOM FOREST AND SUPPORT VECTOR MACHINE METHODS ON TWITTER SENTIMENT ANALYSIS (CASE STUDY: INTERNET SELEBGRAM RACHEL VENNYA ESCAPE FROM QUARANTINE)

Sudianto^{*1}, Puspa Wahyuningtias², Hapsari Warih Utami³, Uli Ahda Raihan⁴, Hasna Nur Hanifah⁵, Yehezkiel Nicholas Adanson⁶

^{1,2,3,4,5,6}Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto
Email: ¹sudianto@ittelkom-pwt.ac.id, ²18102137@ittelkom-pwt.ac.id, ³18102015@ittelkom-pwt.ac.id,
⁴18102034@ittelkom-pwt.ac.id, ⁵18102124@ittelkom-pwt.ac.id, ⁶18102144@ittelkom-pwt.ac.id

(Naskah masuk: 14 Februari 2022, Revisi: 18 Februari 2022, diterbitkan: 25 Februari 2022)

Abstract

Coronavirus (Covid-19) is an infectious disease spreading widely throughout the world. Covid-19 has been declared a pandemic. Transmission of Covid-19 spreads through the Droplet. The Indonesian government has made efforts to prevent the spread of Covid-19, one of which is the implementation of quarantine regulated through Circular Letter Number 8 of 2021 concerning International Travel Health Protocols. The case of Selebgram Rachel Vennya's escape from quarantine had become a trending topic on Twitter. Many Twitter users in Indonesia gave their opinions and comments on this case. Therefore, it is necessary to research public sentiment on the case of the escape of Selebgram Rachel Vennya from quarantine. The data used is taken from netizen comments from social media, namely Twitter, in the form of positive and negative comments; the algorithms used are Random Forest (RF) and Support Vector Machine (SVM). This study aims to compare the classification method to public sentiment regarding the case of the escape of Selebgram Rachel Vennya from quarantine using the Random Forest and SVM methods. The classification results show that the Random Forest algorithm has an accuracy value of 94%. In comparison, the SVM algorithm classification results get an accuracy value of 93%. So it can be concluded, Twitter sentiment analysis in the case study of Rachel Venya's escape from quarantine that the Random Forest algorithm got the best results.

Keywords: Covid-19, NLP, Random Forest, SVM, sentiment analysis, quarantine.

PERBANDINGAN METODE RANDOM FOREST DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TWITTER (STUDI KASUS: KABURNYA SELEBGRAM RACHEL VENNYA DARI KARANTINA)

Abstrak

Coronavirus (Covid-19) merupakan penyakit menular yang menyebar luas di seluruh dunia. Covid-19 telah dinyatakan sebagai pandemi. Penularan Covid-19 menyebar melalui droplet. Pemerintah Indonesia telah melakukan upaya pencegahan penyebaran Covid-19, salah satunya dengan penerapan karantina yang diatur melalui Surat Edaran Nomor 8 Tahun 2021 tentang Protokol Kesehatan Perjalanan Internasional. Kasus kaburnya Selegram Rachel Vennya dari karantina sempat menjadi trending topik di Twitter. Banyak pengguna Twitter di Indonesia yang memberikan pendapat dan komentar terkait kasus ini. Oleh karena itu, perlu dilakukan penelitian tentang sentimen publik terhadap kasus kaburnya Selegram Rachel Vennya dari karantina. Data yang digunakan diambil dari komentar warganet dari media sosial yaitu Twitter, berupa komentar positif dan negatif; algoritme yang digunakan adalah Random Forest (RF) dan Support Vector Machine (SVM). Penelitian ini bertujuan untuk membandingkan metode klasifikasi terhadap sentimen publik terkait kasus kaburnya Selebgram Rachel Vennya dari karantina menggunakan metode Random Forest dan SVM. Hasil klasifikasi menunjukkan bahwa algoritme Random Forest memiliki nilai akurasi sebesar 94%. Sedangkan, hasil klasifikasi algoritme SVM mendapatkan nilai akurasi sebesar 93%. Sehingga dapat disimpulkan, analisis sentimen Twitter pada studi kasus kaburnya Selegram Rachel Venya dari karantina bahwa algoritme Random Forest mendapatkan hasil terbaik.

Kata kunci: Analisis sentimen, Covid-19, karantina, NLP, Random Forest, SVM.

1. PENDAHULUAN

Pada akhir tahun 2019 hingga awal 2022, peristiwa pandemi yang disebabkan oleh *Coronavirus* atau yang biasa disebut dengan Covid-19 masih mencemaskan seluruh warga dunia. Covid-19 dapat menular melalui mulut dan hidung saat orang yang terinfeksi berbicara, batuk, maupun bersin, pada ruangan tertutup atau keramaian ketika melakukan kontak dekat dengan orang yang terinfeksi [1]. Selain itu, Covid-19 dapat hidup berhari-hari pada area terbuka dalam kondisi atmosfer yang menguntungkan [2].

Pemerintah Indonesia telah banyak melakukan upaya untuk menurunkan tingkat penyebaran Covid-19. Salah satu upaya yang sudah dilakukan pemerintah Indonesia yaitu dengan mengeluarkan tiga peraturan. Ketiga peraturan tersebut yaitu: (1) Perppu Nomor 1 Tahun 2020; (2) Perppu Nomor 21 Tahun 2020; (3) Keputusan Presiden Nomor 11 Tahun 2020. Undang-Undang Karantina Kesehatan No.6 pada tahun 2018 (selanjutnya disebut UU No.6 Tahun 2018) menjadi dasar diberlakukannya tiga peraturan perundang-undangan untuk menekan penyebaran Covid-19. Salah satunya menerapkan karantina, melakukan karantina kesehatan merupakan upaya tanggung jawab dari pemerintah, untuk membantu penanganan warga negara dari segala jenis bahaya, jenis wabah, dan faktor risiko kesehatan yang dapat mengancam kesehatan masyarakat [3]. Karantina menjadi aktivitas yang harus dilakukan setelah seseorang melakukan bepergian dari luar Indonesia. Karantina bertujuan untuk memantau, mengendalikan, dan mengevaluasi guna mencegah peningkatan infeksi Covid-19 [4]. Penduduk yang mengingkari peraturan karantina kesehatan, maka akan mendapatkan hukuman yang tertera pada Undang-Undang Nomor 14 tentang Wabah Penyakit Menular 4/1984 dan Pasal 93 Undang-Undang Karantina Kesehatan 6/2018 [3].

Kaburnya *Selebgram* Rachel Vennya dari Wisma Atlet yang ramai dibicarakan di media sosial sejak 9 Oktober 2021 merupakan contoh warga yang tidak mematuhi Undang-Undang Karantina Kesehatan. Rachel Vennya diketahui telah berpergian dari Amerika Serikat. Sesampainya di Indonesia, Rachel Vennya hanya berdiam di Wisma Atlet selama tiga hari [5]. Berdasarkan Surat Edaran Nomor 8 Tentang Prokes (Protokol Kesehatan) Tata Cara Perjalanan Selama Pandemi Perjalanan Luar Negeri Tahun 2021 (Covid-19) Seluruh perjalanan luar negeri WNI maupun WNA diwajibkan untuk tes PCR dan diwajibkan 8 (delapan) hari karantina sebelum bertemu orang lain [4].

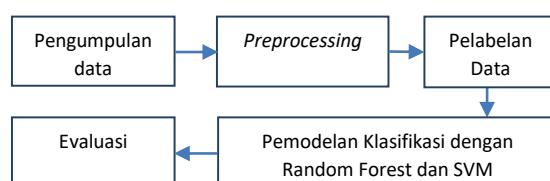
Kasus *Selebgram* Rachel Vennya melarikan diri dari karantina telah menjadi topik hangat di media sosial Twitter. Banyak orang Indonesia berkomentar kejadian tersebut. Opini atau sentimen publik, yang biasa disebut sebagai "cuitan" dalam terminologi Twitter, dapat berbentuk opini negatif

atau positif. Oleh karena itu, banyaknya opini komentar masyarakat maka diperlukan analisa untuk mengklasifikasikan opini yang beredar [6]. Analisis sentimen adalah cara untuk mendapatkan berbagai opini masyarakat tentang pelayanan publik, kasus, isu publik, serta kejadian permasalahan. Sehingga dari analisis sentimen dapat memperoleh berbagai klasifikasi berupa sentimen positif dan negatif [6].

Penelitian sebelumnya yang dilakukan pada tahun 2019 oleh Rachmad Mahendrajaya *et al.* Membahas tentang sentiment analisis masyarakat terhadap layanan Go-Pay menggunakan metode Lexicon Based dan Support Vector Machine yang menghasilkan akurasi sebesar 89,17% [7]. Penelitian lain yang menggunakan algoritme SVM yang dilakukan oleh Aliffia Kulsumarwati *et al.* pada tahun 2021 mengenai analisis sentimen menggunakan komentar pengguna Twitter tentang pelaksanaan Pilkada mendapatkan hasil akurasi sebesar 92%, dimana nilai *precision* 22%, dan *recall* 67% [8]. Pada dua penelitian tersebut algoritme SVM dapat melakukan analisis sentimen pada sosial media Twitter dengan hasil akurasi yang cukup tinggi. Penelitian selanjutnya yang melakukan perbandingan metode SVM dan *Random Forest* dilakukan oleh M. R. Adrian *et al.* pada tahun 2021 untuk analisis sentimen masyarakat pada Twitter mengenai penerapan PSBB. Pada penelitian ini didapatkan akurasi pada pengujian menggunakan *confusion matrix* pada algoritme *Random Forest* sebesar 0,578, sedangkan pada algoritme SVM sebesar 0,557 [9]. Penelitian lain tentang analisis sentimen terhadap aplikasi Ruangguru algoritme Naive Bayes, *Random Forest* dan Support Vector Machine yang dilakukan oleh Evita Fitri *et al.* pada tahun 2020. Penelitian tersebut bertujuan untuk mendapatkan metode terbaik yang dapat digunakan dalam analisis sentimen dengan hasil yang didapatkan dalam pengujian menggunakan *Random Forest* diperoleh akurasi sebesar 97,16%, sedangkan akurasi yang diperoleh dari SVM sebesar 96,01% dan akurasi menggunakan Naive Bayes didapatkan sebesar 94,16% [10].

Berdasarkan latar belakang dan penelitian sebelumnya, pada penelitian ini akan dilakukan perbandingan metode mana yang lebih baik dalam mengklasifikasikan sentimen masyarakat mengenai kasus kaburnya *Selebgram* Rachel Vennya dari karantina antara metode *Random Forest* dan Support Vector Machine (SVM).

2. METODE PENELITIAN



Gambar 1. Diagram alir penelitian

Diagram alir penelitian ini ditunjukkan pada Gambar 1.

2.1. Pengumpulan Data

Tahap pertama dari metodologi penelitian ini merupakan tahap pengumpulan data yang diambil dari hasil *crawling* di media sosial Twitter, menggunakan kata kunci “rachel” dengan rentang waktu dari tanggal 1 Oktober 2021 hingga 20 Desember 2021. *Crawling data* menggunakan Bahasa pemrograman Python melalui *emulator* Google Colab. Proses *Crawling* tanpa menggunakan API, namun menggunakan *library* Twint. Data yang didapatkan sebanyak 1000 data *tweets*. Hasil *crawling data* ditunjukkan pada Gambar 2.

	Created	Tweets	User
0	2021-12-18 11:28:26	b'Melupakan aku itu hakmu seutuhnyaTetap men...	Rachel_Seren
1	2021-12-18 11:14:43	b'@poecular Kalo sopan mah rachel venny'	Juan.
2	2021-12-18 11:13:25	b'@nanamint0n Rachel sama trias wdn juga ya?'	Srtx
3	2021-12-18 11:03:24	b'@A100888 Buset belum ada yg bener/nRachel ...	flo
4	2021-12-18 11:00:02	b'@bigcozyb0i Gue sampe berhanti baca ga kuat ...	

Gambar 2. Hasil pengumpulan data

2.2. Preprocessing

Preprocessing dilakukan untuk membentuk *dataset* yang siap dianalisis. Pada penelitian ini, tahap *preprocessing* dilakukan dengan empat langkah, yaitu:

1. Data Cleansing

Data *cleansing* adalah membersihkan data dari *noise* seperti tanda baca dan karakter lain yang tidak penting [11]. *Data cleansing* dimulai dari menghapus *RT*, *hashtag*, karakter yang tidak penting dan lain-lain. Kemudian semua huruf besar dilakukan konversi menjadi huruf kecil (*case folding*). Mengubah *tweet* juga dapat meningkatkan konsistensi data.

2. Tokenizing

Tokenizing merupakan tahapan untuk memotong sebuah dokumen atau kalimat menjadi bagian-bagian, yang disebut dengan *token*. Tahap *tokenizing* akan memisahkan frase, kata, simbol, dan entitas penting lainnya.

3. Remove Stopword

Stopword merupakan kata-kata yang tidak unik dalam dokumen. Contohnya adalah "oleh", "pada", "di", "karena", "sebauh", dll. *Stopwords* akan dihapus untuk meningkatkan kinerja analisis sentimen.

4. Stemming

Stemming merupakan sebuah proses atau tahapan untuk menguraikan suatu kata menjadi bentuk kata dasarnya.

2.3. Pelabelan Data

Model analisis sentimen pada penelitian ini termasuk kedalam *Supervised Learning*, dimana *dataset* memiliki target atau label [12]. Pada

penelitian ini menentukan label berdasarkan sentimen positif dan sentimen negatif menggunakan indonesia *sentiment lexicon* (source: <https://github.com/fajri91/InSet>).

2.4. Klasifikasi

Penelitian ini akan membandingkan algoritme Random Forest dengan algoritme SVM. Algoritme Random Forest merupakan metode klasifikasi dan regresi yang tersusun dari sekumpulan pohon keputusan yang berfungsi sebagai *base classifier* yang dibangun dan dikombinasikan [13]. *Bootstrap* sampling merupakan salah satu aspek penting dari metode Random Forest yang berfungsi untuk membuat pohon prediksi, dimana setiap pohon keputusan menggunakan prediksi menggunakan prediktor acak dan rata-rata klasifikasi mayoritas dan regresi. Random Forest sendiri yang membuat prediksi dengan menggabungkan hasil dari setiap pohon keputusan [14]. Sedangkan Support Vector Machine (SVM) merupakan metode pembelajaran terlatih yang membuat *hyperplanes* dalam proses regresi, klasifikasi, dan deteksi anomali. Salah satu kegunaan metode tersebut adalah untuk mengelompokkan teks dan *hypertext* [15].

2.5. Evaluasi

Tahap evaluasi pada penelitian ini terdiri dari proses pengujian yang dilakukan menggunakan *Confusion Matrix*. Proses pengujian menggunakan *Confusion Matrix* digunakan untuk mengecek kinerja algoritme. Tabel *Confusion Matrix* ditunjukkan pada Tabel 1. Setiap kolom matriks mewakili kelas yang diprediksi, dan matriks baris mewakili kelas yang sebenarnya. Klasifikasi memerlukan perhitungan data table matriks menggunakan empat ukuran yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*, untuk mewakili hasil klasifikasi yang ada. Pada tahap evaluasi akan menghasilkan nilai *recall*, *precision*, dan *precision* [16].

Tabel 1. *Confusion Matrix*

Klasifikasi		Prediksi	
		Positive	Negative
Aktual	Positive	True Positive (TP)	True Negative (TN)
	Negative	False Positive (FP)	False Negative (FN)

3. HASIL DAN PEMBAHASAN

Preprocessing memiliki peran yang penting pada saat melakukan pemodelan analisis sentimen dikarenakan keadaan dari teks mempengaruhi hasil akurasi. Terdapat 4 tahapan *text-preprocessing* yang dilakukan yaitu *cleansing*, *stopword-removing*, *tokenizing* dan *stemming*. Keempat tahapan ini dilakukan secara berurutan. Proses *cleansing* dilakukan untuk mengubah seluruh teks data ulasan

menjadi huruf kecil atau *lowercase*, membuang karakter selain huruf, dikarenakan data pada *tweets* tidak konsisten dalam penggunaan huruf kapital dan terdapat angka, tanda baca, maupun *emoticon* seperti contoh pada Tabel 2. Tahap kedua *tokenizing* yaitu mengurai kalimat *tweets* menjadi kata per kata seperti contoh pada Tabel 3.

Tabel 2. Hasil *Cleansing*

Tweets	Hasil Cleansing
b'@GUSDURians kurang sopan kakeknya mungkin? saran saya semua terdakwa bisa belajar sopan santun ke rachel venny	b kurang sopan kakeknya mungkin saran saya semua terdakwa bisa belajar sopan santun ke rachel venny
\xf0\x9f\x98\x8f	xfxfxf

Tabel 3. Hasil *Tokenizing*

Hasil Cleansing	Hasil Tokenizing
b kurang sopan kakeknya mungkin saran saya semua terdakwa bisa belajar sopan santun ke rachel venny	['b', 'kurang', 'sopan', 'kakeknya', 'mungkin', 'saran', 'saya', 'semua', 'terdakwa', 'bisa', 'belajar', 'sopan', 'santun', 'ke', 'rachel', 'venny', 'xfxfxf']

Tahap ketiga, *stopword removing* menggunakan *corpus NLTK*. *Stopword-removing* yaitu membuang kata-kata yang tidak mempunyai arti penting seperti contoh pada Tabel 4. Misalnya kata sambung dan, dengan, yang, dan sejenisnya. Kemudian tahap terakhir adalah melakukan *stemming* untuk mengubah suatu kata menjadi bentuk kata dasarnya seperti contoh pada Tabel 5.

Tabel 4. Hasil *Stopword-Removing*

Hasil Tokenizing	Hasil Stopword-Removing
['b', 'kurang', 'sopan', 'kakeknya', 'mungkin', 'saran', 'saya', 'semua', 'terdakwa', 'bisa', 'belajar', 'sopan', 'santun', 'ke', 'rachel', 'venny', 'xfxfxf']	['b', 'sopan', 'kakeknya', 'saran', 'terdakwa', 'belajar', 'sopan', 'santun', 'rachel', 'venny', 'xfxfxf']

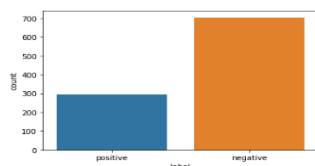
Tabel 5. Hasil *Stemming*

Hasil Stopword-Removing	Hasil Stemming
['b', 'sopan', 'kakeknya', 'saran', 'terdakwa', 'belajar', 'sopan', 'santun', 'rachel', 'venny', 'xfxfxf']	['b', 'sopan', 'kakek', 'saran', 'dakwa', 'ajar', 'sopan', 'santun', 'rachel', 'venny', 'xfxfxf']

Setelah dilakukan *preprocessing* data diberikan label yaitu kelas positif atau negatif. Contoh hasil pelabelan terdapat pada Gambar 3.

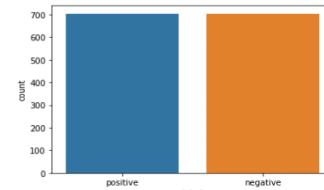
	Tweets	text_preprocessed	label
0	b'Melupakan aku itu hakmu seutuhnyaTetap men...	bmelupakan hak seutuhnyaTetap cinta keputusan...	positive
1	b'@poecular Kalo sopan mah rachel venny	b kalo sopan mah rachel venny	negative
2	b'@nanamintDn Rachel sama trias wdn juga ya?	b rachel trias wdn ya	positive
3	b'@A100888 Buset belum ada yg benerrnRachel ...	b buset bom yg benerrnRachel venny	negative
4	b'@bigcozy00l Gue sampe bernanti baca ga kuat ...	b gue sampe bernanti baca ga kuat liat rachel ...	positive

Gambar 3. Hasil pelabelan



Gambar 4. Proporsi kelas hasil pelabelan

Proporsi jumlah data yang termasuk kelas positif dan negatif ditunjukkan oleh Gambar 4. Jika dilihat dari jumlah kelas, terlihat kedua kelas sentimen termasuk *imbalanced* (tidak seimbang). Karena data tidak seimbang maka dilakukan penyeimbangan data, untuk meningkatkan akurasi saat membuat model. Pada penelitian ini menggunakan teknik *oversampling* untuk menyeimbangkan data. Hasil *oversampling* ditunjukkan pada Gambar 5.



Gambar 5. Hasil Penyeimbangan data

Sebelum dilakukan pemodelan, dilakukan tahap vektorisasi untuk mengubah data berupa kata (token) menjadi numerik. Setelah pembobotan kata dalam bentuk numerik, selanjutnya dapat dilakukan pemodelan analisis sentimen dengan algoritme klasifikasi *Random Forest* dan *SVM* untuk melihat manakah model yang terbaik.

Berikut hasil evaluasi dari algoritme *Random Forest* dan *SVM* yang dipakai. Indikator yang digunakan yaitu nilai *precision*, *recall*, *f1-score*, dan *accuracy*.

Training accuracy Score	: 1.0			
Test accuracy Score	: 0.9375			
	precision	recall	f1-score	support
negative	0.96	0.92	0.94	187
positive	0.91	0.96	0.93	165
accuracy			0.94	352
macro avg	0.94	0.94	0.94	352
weighted avg	0.94	0.94	0.94	352

Gambar 6. Hasil evaluasi algoritme *Random Forest*

Training accuracy Score	: 1.0			
Test accuracy Score	: 0.9318181818181818			
	precision	recall	f1-score	support
negative	0.97	0.90	0.94	193
positive	0.89	0.97	0.93	159
accuracy			0.93	352
macro avg	0.93	0.94	0.93	352
weighted avg	0.94	0.93	0.93	352

Gambar 7. Hasil evaluasi algoritme *SVM*

Berdasarkan hasil evaluasi *Random Forest* yang ditunjukkan pada Gambar 6. Algoritme *Random Forest* mendapatkan nilai *precision*, *recall*, *f1-score*, dan *accuracy* sebesar 94%. Sedangkan hasil evaluasi algoritme *SVM* yang ditunjukkan pada gambar 7, mendapatkan nilai *accuracy* sebesar 93%, rata rata nilai *precision* sebesar 93%, *recall* sebesar 94% dan *f1-score* sebesar 93%.

Berdasarkan hasil evaluasi pada algoritme *Random Forest* dan *SVM*, kedua algoritme tersebut menunjukan hasil yang sangat baik pada klasifikasi sentimen Twitter dengan studi kasus kabarnya

Selebgram Rachel Vennya dari karantina. Tetapi algoritme Random Forest mendapatkan hasil lebih unggul dari pada algoritme SVM.

4. KESIMPULAN

Algoritme Random Forest mendapatkan hasil yang lebih optimal dalam melakukan analisis sentimen. Algoritme Random Forest mendapatkan nilai *precision*, *recall*, *f1-score*, dan *accuracy* sebesar 94%. Sedangkan algoritme SVM mendapatkan nilai *accuracy* sebesar 93%, rata rata nilai *precision* sebesar 93%, *recall* sebesar 94% dan *f1-score* sebesar 93%. Dari studi kasus kaburnya *Selebgram* Rachel Vennya dari karantina dapat diketahui bahwa algoritme Random Forest dan algoritme SVM mendapatkan hasil yang sangat baik, tetapi algoritme Random Forest memperoleh hasil yang lebih baik dibandingkan dengan algoritme SVM.

DAFTAR PUSTAKA

- [1] K. K. Sahu, A. K. Mishra, and A. Lal, "Comprehensive update on current outbreak of novel coronavirus infection (2019-nCoV)," *Ann. Transl. Med.*, vol. 8, no. 6, pp. 393–393, 2020, doi: 10.21037/atm.2020.02.92.
- [2] L. Morawska and J. Cao, "Airborne transmission of SARS-CoV-2: The world should face the reality," *Environ. Int.*, vol. 139, no. April, p. 105730, 2020, doi: 10.1016/j.envint.2020.105730.
- [3] Undang Undang Nomor 6 tahun 2018 tentang Kekarantinaan Wilayah, "Undang Undang Nomor 6 tahun 2018 tentang Kekarantinaan Wilayah," *Natl. Stand. Agency Indones.*, pp. 31–34, 2018, [Online]. Available: <https://jdih.bsn.go.id/produk/detail/?id=730&jns=2>
- [4] BNPB, "Addendum Surat Edaran Protokol Kesehatan Kesehatan Perjalanan Internasional pada Masa Pandemi Covid-19." 2021.
- [5] Suara.com, "Kronologi Rachel Vennya Kabur dari Wisma Atlet," *Suara.Com*, 2021.
- [6] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013, doi: 10.1145/2436256.2436274.
- [7] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [8] A. Kulsumarwati, I. Purnamasari, and B. A. Darmawan, "Penerapan SVM dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, pp. 101–109, 2021, doi: 10.37012/jtik.v7i2.641.
- [9] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021.
- [10] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [11] R. Patel and K. Passi, "Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning," *IoT*, vol. 1, no. 2, pp. 218–239, 2020, doi: 10.3390/iot1020014.
- [12] I. Muhammad and Z. Yan, "Supervised Machine Learning Approaches: a Survey," *ICTACT J. Soft Comput.*, vol. 05, no. 03, pp. 946–952, 2015, doi: 10.21917/ijsc.2015.0133.
- [13] L. Breiman, "Random forests," *Random For.*, pp. 1–122, 2001, doi: 10.1201/9780429469275-8.
- [14] 2018) (Al Amrani, Lazaar, El Kadiri., "Random Forest and Support Vector Machine based Hybrid Approach to SA -- RF.pdf," *The First International Conference on Intelligent Computing in Data Sciences*. pp. 511–520, 2018.
- [15] J. T. Lalis, "A new multiclass classification method for objects with geometric attributes using simple linear regression," *IAENG Int. J. Comput. Sci.*, vol. 43, no. 2, pp. 198–203, 2016.
- [16] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, 2019, doi: 10.3390/e21060589.