

Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode *Random Forest Classifier* (Studi Kasus: Vaksin Sinovac)

Muhammad Yusril Aldean ^{#1} Paradise ^{#2}, Novanda Alim Setya Nugraha ^{#3}

^{#1,2,3} Teknik Informatika

^{#1,2,3} Institut Teknologi Telkom Purwokerto

JL. D.I Panjaitan 128 Purwokerto Selatan, Jawa Tengah, Indonesia

¹ 18102062@iittelkom-pwt.ac.id

² paradise@ittelkom-pwt.ac.id

³ novanda@ittelkom-pwt.ac.id

Penulis Korespondensi : paradise@ittelkom-pwt.ac.id

Abstrak

Pada tahun 2019 terjadi musibah yang melanda berbagai negara didunia termasuk Indonesia. Terjadi penyebaran virus secara cepat dan menyeluruh, yaitu Virus Covid-19. Kasus Covid-19 pertama di dunia terdeteksi di Kota Wuhan, Provinsi Hubei, China. Penyakit ini disebabkan oleh virus sindrom pernafasan akut Coronavirus 2 (SARS-CoV-2). Indonesia sendiri sudah melakukan langkah vaksinasi untuk Virus Covid-19, dengan menggunakan beberapa jenis vaksin yang salah satunya adalah Vaksin Sinovac. Program vaksinasi yang dilakukan di Indonesia menuai banyak pro dan kontra khususnya dari masyarakat. Banyak dari masyarakat yang menyampaikan pendapatnya melalui media sosial berbasis teks, salah satu media sosial yang sering digunakan adalah Twitter. Sehingga sentimen masyarakat yang terdapat di media sosial dapat menjadi tolak ukur bagaimana informasi melalui media sosial yang diterima oleh masyarakat adalah hal positif ataupun hal negatif, sehingga dapat dievaluasi bersama. Pada penelitian ini dibuat sebuah metode *machine learning* untuk menganalisis sentimen masyarakat pada program vaksinasi menggunakan Vaksin Sinovac. Penelitian ini menggunakan *tweet* sebanyak 1500 data dengan pembagian 2 kategori yaitu positif dan negatif. Pengoalahan data yang digunakan pada penelitian ini adalah dengan menggunakan Algoritme *TF-IDF* serta penyeimbangan data menggunakan *SMOTE*. Model yang dibuat akan dilatih dengan Algoritme *Random Forest Classifier* dan akan divalidasi menggunakan *K-fold Cross Validation* dan *Confusion Matrix*. Hasil pada penelitian ini adalah sentimen masyarakat terhadap Vaksinasi Sinovac adalah positif dan model dapat memprediksi sentiment sebuah *tweet* dengan akurasi mencapai 79% dan nilai *Precision* sebesar 85%, *Recall* sebesar 90% dan *F1 Score* sebesar 88%..

Kata Kunci: *Twitter, Vaksin Sinovac, Machine Learning, Analisis Sentimen, Random Forest Classifier*

I. PENDAHULUAN

Pada Pandemi wabah Virus Covid-19 sudah melanda berbagai macam negara termasuk Indonesia. Covid-19 memiliki tingkat penyebaran virus yang sangat cepat, sehingga harus ditangani secara cepat dan seksama oleh pemimpin negara dan rakyat[1]. Strategi dalam penanganan Covid-19 salah satunya adalah dengan program vaksinasi. Proses vaksin seringkali menuai komentar masyarakat dari yang bertema positif maupun negatif[2]. Masyarakat yang pro terhadap program tersebut akan menuliskan dukungannya dan masyarakat yang kontra terhadap vaksinasi akan mengkritisi program tersebut. Pemerintah secara konsisten melakukan pengembangan dalam penanganan dan sudah pada tahap vaksinasi secara menyeluruh sehingga sejak tanggal 5 Juli 2021, sebanyak 32.301.268 masyarakat sudah melakukan vaksinasi pertama dan sebanyak 14.035.934 sudah melakukan vaksinasi kedua[3].

Penelitian ini akan menggunakan Vaksin Sinovac sebagai data dikarenakan vaksin tersebut merupakan vaksin pertama yang digunakan oleh warga Indonesia[4]. Kabar vaksinasi massal yang dilakukan pemerintah menimbulkan pro dan kontra dari sisi masyarakat. Banyak masyarakat yang merasakan pro dengan berasumsi bahwa semakin cepat dilaksanakannya vaksinasi, maka semakin cepat juga pandemi ini akan selesai[5]. Berbeda dengan beberapa masyarakat yang memilih pro dengan program vaksinasi, masyarakat juga memiliki kontra terhadap vaksinasi ini. Masyarakat beranggapan bahwa vaksinasi ini bersifat memaksa dan secara tidak langsung menjadi suatu kewajiban bagi masyarakat khususnya bagi yang tidak mau divaksin[5]. Maka dari itu penelitian analisis sentimen ini dibuat karena untuk mengerti bagaimana opini masyarakat terhadap program vaksinasi khususnya pada Vaksin Sinovac. Data opini penelitian ini diambil dari media sosial Twitter.

Media sosial Twitter menjadi alat berkomunikasi yang sangat populer digunakan dimasyarakat pengguna internet[6]. Statistik menyatakan bahwa pada April 2010, Twitter memiliki pengguna sebanyak 106 juta akun dan sebanyak 180 juta pengunjung untuk setiap bulannya. Jumlah peningkatan pengguna Media Sosial Twitter disebutkan meningkat 300.000 user pada setiap harinya. Penelitian ini menggunakan *machine learning* dengan Algoritma *TF-IDF* untuk pemrosesan data sehingga data dapat diolah dalam bentuk matriks dan *Random Forest Classifier* untuk prediksi, Hasil prediksi akan dilampirkan dalam bentuk *Confusion Matrix*. *random forest classifier* adalah metode standar untuk penelitian berbasis klasifikasi dan regresi yang termasuk dari *ensemble learning*[7]. Model *Random Forest Classifier* banyak digunakan untuk menyelidiki sebuah klasifikasi karena model ini merupakan sebuah perkembangan dari model *tree*[7]

II. DASAR TEORI

A. Covid-19

Covid-19 adalah penyakit yang disebabkan oleh Virus Corona dan mulai muncul pada tahun 2019 di Wuhan, China dan menyebar keseluruh bagian negara di seluruh dunia[8]. Gejala utama penyakit Covid-19 yaitu batuk, demam, dan sesak napas. Covid-19 telah membuat ekonomi negara berubah, termasuk Indonesia dengan dibuktikan adanya pembuatan anggaran dalam rangka penanganannya, termasuk untuk desa-desa. Penamaan *2019-nCoV* disematkan *World Health Organization (WHO)*[9]. Pentingnya penanganan wabah virus ini dikarenakan dapat membuat masyarakat menjadi resah dan dapat meningkatkan kejahatan khususnya karena terjadinya penurunan ekonomi yang sedang terjadi[10].

B. Vaksin Sinovac

Vaksin merupakan sebuah cara untuk membuat tubuh manusia memiliki kekebalan terhadap virus tertentu untuk mengurangi efek samping apabila terjangkit virus tersebut[11]. Perkembangan vaksin yang aman dan efektif pada saat menyebarnya Virus Covid-19 ini sangat penting. Karena diharapkan mampu menghambat penyebaran dan mencegah terulangnya kembali[11]. Pada saat pandemi Virus Covid-19, Pemerintah Indonesia pada tanggal 13 Januari 2021 memesan 125 Juta Vaksin Covid-19 yang dikembangkan oleh Sinovac berasal dari China[12].

C. Analisis Sentimen

Analisis sentimen merupakan suatu aktivitas menganalisa suatu perasaan, emosi dan opini seseorang yang di ekspresikan dalam teks.[13]. Analisis sentimen juga merupakan ilmu yang berguna untuk menganalisis pendapat seseorang, sentiment seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis[14]. Adapun dalam analisis sentimen pada umumnya melakukan suatu proses pengelompokkan dari teks yang ada didalam baik dokumen, kalimat atau suatu pendapat sehingga menghasilkan suatu nilai yang mengartikan apakah dokumen, kalimat dan pendapat tersebut bernilai positif atau negatif[15].

D. Twitter

Media massa merupakan perkembangan alat yang dibuat manusia dan semakin hari semakin berkembang dan menyebar. Media massa menjadi sarana untuk menyebarkan pesan/informasi kepada masyarakat[16]. Twitter adalah media sosial yang populer dimasyarakat dan banyak digunakan pada saat ini. Twitter merupakan peringkat kedua menjadi media sosial teraktif di Indonesia[17].

Twitter adalah media sosial mikroblog yang digunakan pengguna untuk mengirimkan pesan dengan batas penggunaan hingga 280 karakter. Twitter digunakan pengguna untuk melakukan penilaian dan mengeluarkan opini dan pendapat mengenai sesuatu hal, mengunggah dan mengukur dengan opini yang berbeda-beda[18]. Twitter memiliki fitur, tanda # atau tagar menunjukkan topik-topik yang sedang dibahas. Fungsi tagar dalam Twitter sebagai media pencarian dan menampilkan informasi lebih cepat dan mudah, sehingga dapat mencari topik yang sedang ramai[18].

E. *Scrapping*

Sebuah cara pengambilan data dimana menggunakan sebagai *interface* sebagai penghubung ke sebuah web, sehingga dapat diartikan bahwa pengambilan data hanya dilakukan pada tampilan yang ada pada web tersebut[19]. *Scrapping* bisa digunakan dengan beberapa metode dan teknik, *Scrapping* juga dapat mengubah data yang tidak terstruktur menjadi terstruktur sehingga dapat disimpan kedalam *database*[19]. Inti dari *scrapping* ini adalah mengumpulkan data, menyimpan data dan memvalidasi data, sehingga data dapat diinterpretasikan menjadi informasi yang baik[19].

F. *Labelling*

Data yang didapatkan dari proses *scrapping*, dilakukan proses *labelling* dengan membuat kolom baru yang berisikan *value* dari *tweet* tersebut. Jika *tweet* tersebut bermaksud negatif atau bertanggung negatif maka akan bernilai 0, dan jika bermaksud positif atau bertanggung positif maka akan bernilai 1

G. *TF-IDF Tokenizing*

Sebuah metode dimana merubah sebuah kalimat yang berisi kata-kata menjadi sebuah matriks. Sehingga data yang dibuat dapat dimasukkan sebagai dataset dalam bentuk list[20]. Token kemudian dimasukkan kedalam sebuah variabel dan diubah menjadi sebuah Tensor berisikan angka yang diproses melalui sebuah algoritma[21]. Perhitungan bobot kata dimulai dengan menghitung nilai TF dengan bobot masing masing kata adalah 1, dan IDF dihitung dengan cara

$$TF - IDF(tk, dj) = TF(tk, dj) * IDF(tk, dj) \quad (1)$$

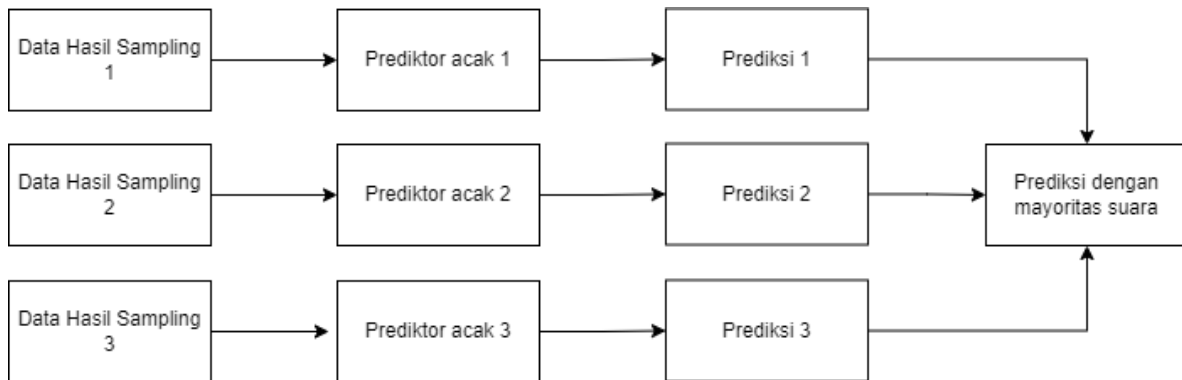
IDF yang dicari adalah dengan pembagian antara *td* yang bernilai jumlah keseluruhan dokumen yang ada, dan *df* adalah kemunculan kata pada data[22].

H. *SMOTE*

Pada penelitian dengan data yang banyak, seringkali terjadinya kelas yang tidak seimbang dikarenakan data yang kurang cocok ataupun data yang tidak terstruktur dengan baik[23]. Sehingga dibutuhkan sebuah metode untuk membebani permasalahan ini. Metode *SMOTE* adalah suatu teknik *oversampling* dimana sampel sintetis dihasilkan untuk kelas minoritas atau dengan data yang sedikit[23]. Beberapa kondisi dapat digunakan bahkan hingga data minor dapat setara dengan data mayor[24].

I. Random Forest Classifier

Algoritme *Random Forest Classifier* adalah pengembangan dari model Algoritme *Decision Tree*, dimana setiap pohon fikiran dilatih dengan sampel individu[7]. Model *Random Forest Classifier* yang menghasilkan banyak *tree* dan dengan cara yang sama. Seiring bertambahnya data, maka *tree* ikut berkembang[7]. *Random Forest Classifier* merupakan algoritme dengan membuat pohon klasifikasi dan regresi yang nodenya dipisahkan oleh algoritme yang dioptimasi sebagai fungsi untuk meminimalkan *squared-error loss*[25].



Gambar 1. Alur perhitungan *Random Forest Classifier*

Random Forest dibangun dengan menggunakan pemilihan atribut secara acak. Metode *CART* (*Classification and Regression Tree*) digunakan untuk membuat pohon keputusan, sehingga pohon keputusan tersebut tumbuh mencapai ukuran maksimum dan tidak dipangkas sehingga dihasilkan kumpulan pohon yang kemudian disebut *forest*[15].

Waktu komputasi yang dibutuhkan oleh Algoritme *Random Forest Classifier* bekerja mengklasifikasi adalah :

$$T\sqrt{MN\log(N)} \quad (2)$$

Nilai T merupakan banyaknya pohon, M adalah banyaknya peubah yang digunakan pada subsample, N adalah banyaknya pengujian[25].

J. K-fold Croos Validation

Model *Kfold Cross Validation* adalah sebuah cara untuk mengevaluasi suatu model pembelajaran mesin[23]. *Kfold Cross Validation* membagi *dataset* sebanyak nilai *k* partisi yang tidak tumpang tindih yang nantinya akan dilatih sebanyak *k-1* dan sisanya akan divalidasi pada partisi berikutnya[25]. Validasi berguna untuk mengevaluasi model yang sudah dibuat dengan hasil apakah sudah memenuhi ekspektasi untuk akurasinya[26].

K. Confusion Matrix

Confusion Matrix merupakan sebuah cara untuk mempresentasikan hasil daripada akurasi dari model yang sudah dibuat[17]. *Confusion Matrix* merangkum performa dalam mengklasifikasi sesuai dengan banyaknya kategori yang diklasifikasi dari nilai benar dari suatu kelas dari objek yang diprediksi[25]

TABEL I. *Confusion Matrix*

Confusion Matrix	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>TP</i>	<i>TN</i>
<i>Actual Negative</i>	<i>FP</i>	<i>FN</i>

Dimana :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

TP = jumlah data positif yang terklasifikasi benar
TN = jumlah data positif yang terklasifikasi salah
FP = jumlah data positif yang terklasifikasi salah
FN = jumlah data negatif yang terklasifikasi salah.

L. Precision, Recall dan F1 Score

Preciision merupakan sebuah definisi dari ratio antara *True Positive* dan *Predicted Positive* dan *Recall* adalah rasio antara *True Positives* dan *Actual Positives*[17]. Berikut perhitungan *Precision* dan *Recall*:

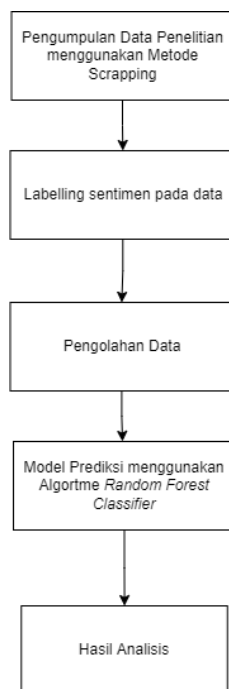
$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} = \frac{2 \cdot TP}{2 \cdot TP+FN+FP} \quad (6)$$

III. METODE PENELITIAN

Variabel pada penelitian ini terdiri dari 2 data. Data pertama yang digunakan sebagai variabel adalah isi dari setiap *tweet* yang sudah diunggah oleh pengguna. *tweet* yang diunggah berisikan tanggapan tentang Vaksin Sinovac. Data kedua yang digunakan adalah data yang dibuat sendiri yang berisikan angka 1 dan 0, angka 1 berartikan bahwa *tweet* tersebut bersikian konten setuju dengan penggunaan Vaksin Sinovac, sedangkan angka 0 berarti bahwa *tweet* tersebut berisi konten tidak setuju dengan penggunaan Vaksin Sinovac.



Gambar 2. Alur Penelitian

A. Pengumpulan Data

Data *scrapping* pada penelitian ini berasal dari media sosial berbasis teks dan gambar yaitu “Twitter” dengan format file *.xlsx* berisikan teks yang pengguna tulis khususnya berkaitan dengan Vaksin Sinovac serta menggunakan tagar #VaksinSinovac sebanyak 1500 data

B. Pemrosesan Data

Pengolahan data berawal dilakukan secara manual dengan membaca semua *tweet* terlebih dahulu dan memberikan sentimen 1 untuk positif dan sentiment 0 untuk negatif kepada setiap *tweet* yang ada, kemudian masuk ke bagian pembobotan kata *TF-IDF Tokenizer* serta *transformasi* kata menjadi sebuah matriks. Kemudian data yang sudah tersedia dimasukkan kedalam Algoritme *SMOTE* untuk penyeimbangan data sebelum dilakukannya pelatihan dan pengujian

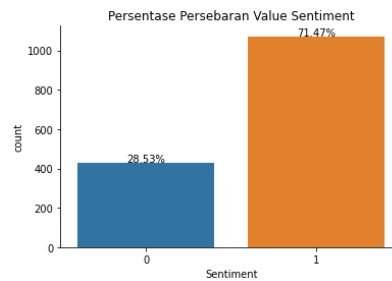
C. Klasifikasi menggunakan Metode Random Forest Classifier

Model pembelajaran mesin yang akan dibuat menggunakan Algoritme *Random Forest Classifier*. Setelah semua selesai, model dievaluasi menggunakan *Kfold Cross Valdiation* dan Hasil akan dijabarkan dalam bentuk *Confusion Matrix* dan nilai *Precision*, *Recall* dan *F1 Score*. Hasil dan kesimpulan yang akan didapatkan setelah model terselesaikan dan hasil evaluasi beserta semua nilai sudah didapatkan.

IV. RESULTS AND DISCUSSION

A. Hasil perhitungan model

Pada penelitian ini menggunakan uji coba data sebanyak 1500 data yang diambil pada tanggal 8 Febuari 2022, maka data yang digunakan akan seperti Gambar 2 berikut :



Gambar 3. Persebaran data sentimen

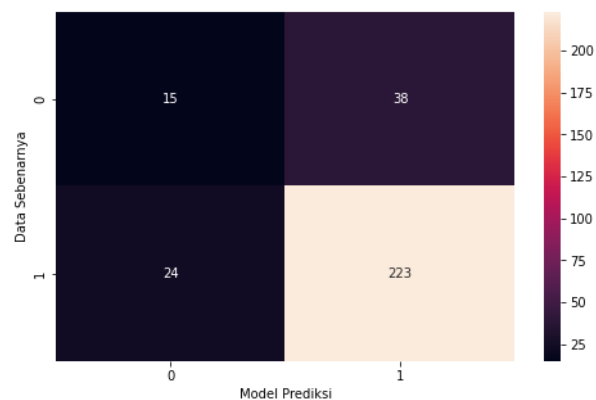
Terdapat persebaran data yang tidak merata sehingga menurut data yang tersedia tersebut lebih banyak pada sentimen positif dengan jumlah 71.33 %, dan untuk sentiment negatif hanya ada di 28.67% data. Sehingga pada penelitian ini digunakan Algoritme *SMOTE* sebagai penyeimbang data. Pembagian data menjadi data *train* dan data *test* sebanyak 80% untuk *train* dengan jumlah 1200 data dan untuk data *test* sebanyak 300 data, kemudian akan diubah menjadi sebuah matrik dengan menggunakan algoritma *Tokenizer*. Model *Random Forest Classifier* menghasilkan akurasi yang akan dilakukan validasi sebanyak 10 kali. Maka hasil akurasi akan seperti Gambar 3 berikut :

TABEL II. Hasil Uji Akurasi

Pengujian Ke -	Akurasi
1	88.48%
2	86.06%
3	84.24%
4	86.06%
5	85.45%
6	84.24%
7	90.30%
8	85.45%
9	87.27%

B. Pengujian *Confusion Matrix*

Pada model ini disajikan juga hasil uji data dalam bentuk *Confusion Matrix*, Matrik ini sendiri merupakan dasar dari perhitungan untuk *Precision*, *Recall* dan *F1 Score* dengan perhitungan yang tersedia pada bagian sebelumnya dengan hasil seperti pada Gambar 4 berikut :



Gambar 4 Hasil *Confusion matrix*

Pada model ini dapat hitung dari hasil *Confusion* Matrik dan menghasilkan *F1-Score* sebesar 88%, Nilai *Recall* sebesar 90% dengan jumlah Nilai *Precision* sebesar 85% dari perhitungan sebelumnya serta akurasi sebesar 79%.

V. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa kesimpulan bahwa penggunaan model *Machine Learning* menggunakan Algoritme *Random Forest Classifier* dan Pengolahan data menggunakan Algoritme *SMOTE* dihasilkan akurasi sebesar 79% dan Analisis Sentimen terhadap Vaksin Sinovac menghasilkan prediksi dengan 13% berisikan komentar negatif dan 87% berisikan komentar positif. Penelitian selanjutnya dapat menambahkan jumlah data pada model yang digunakan

DAFTAR PUSTAKA

- [1] R. H. Syah, "Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran," *SALAM J. Sos. dan Budaya Syar-i*, vol. 7, no. 5, 2020, doi: 10.15408/sjsbs.v7i5.15314.
- [2] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," vol. 1, no. 1, pp. 10–12, 2021.
- [3] T. Tamara, "Gambaran Vaksinasi COVID-19 di Indonesia pada Juli 2021," *Medula*, vol. 11, no. 1, pp. 180–183, 2021.
- [4] N. M. Nasir, I. S. Joyosemito, B. Boerman, and I. Ismaniah, "Kebijakan Vaksinasi COVID-19: Pendekatan Pemodelan Matematika Dinamis Pada Efektivitas Dan Dampak Vaksin Di Indonesia," *J. Pengabd. Kpd. Masy. UBJ*, vol. 4, no. 2, pp. 191–204, 2021, doi: 10.31599/jabdimas.v4i2.662.
- [5] G. Farina and H. Fikri, "Aspek Hukum Pelaksanaan Vaksinasi Covid-19 Di Indonesia," *J. RechtsVinding*, vol. 10, no. April, p. 1263, 2021, doi: 10.24912/psenapenmas.v0i0.15162.
- [6] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [7] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.
- [8] M. K. Sari, "Sosialisasi tentang Pencegahan Covid-19 di Kalangan Siswa Sekolah Dasar di SD Minggiran 2 Kecamatan Papar Kabupaten Kediri," *J. Karya Abdi*, vol. 4, no. 1, pp. 80–83, 2020.
- [9] Sarip, A. Syarifudin, and A. Muaz, "Dampak Covid-19 Terhadap Perekonomian Masyarakat Dan Pembangunan Desa," *J. Penelit. Ekon. Syariah*, vol. 21, no. 1, pp. 1–9, 2020.
- [10] I. Abdullah, "COVID-19: Threat and Fear in Indonesia," *Psychol. Trauma Theory, Res. Pract. Policy*, vol. 12, no. 5, pp. 488–490, 2020, doi: 10.1037/tra0000878.
- [11] I. P. Sari and S. Sriwidodo, "Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin COVID-19," *Maj. Farmasetika*, vol. 5, no. 5, p. 204, 2020, doi: 10.24198/mfarmasetika.v5i5.28082.
- [12] I. R. S. Turnip, "Kehalalan Vaksin Covid-19 Produksi Sinovac Dalam Fatwa Mui Dan Implementasi Vaksinasinya Pada Tenaga Kesehatan Di Puskesmas Tanjung Morawa, Deli Serdang (Perspektif Qawaidh Fiqhiyyah)," *J. Huk. Islam dan Pranata Sos. Islam*, vol. 9, no. 01, pp. 59–83, 2021, doi: 10.30868/am.v9i01.1250.
- [13] K. Ramadhan and K. Muslim, "Analisis Sentimen Terhadap Toko Online Menggunakan Naive Bayes Pada Media Sosial Twitter," *eProceedings Eng.*, vol. 5, no. 3, pp. 8141–8151, 2018.
- [14] N. Saputra, T. B. Adji, and A. E. Permasari, "Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM," *J. Din. Inform.*, vol. 5, no. November, p. 12, 2015.
- [15] 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.
- [16] D. S. Adhlarso, P. Utari, and Y. Slamet, "Pemberitaan Hoax di Media Online Ditinjau dari Konstruksi Berita dan Respon Netizen," *J. Ilmu Komun.*, vol. 15, no. 3, p. 215, 2017, doi: 10.31315/jik.v15i3.2173.
- [17] M. Y. Aldean, M. D. Hilmawan, R. Indriyati, and J. Lasama, "Analisa Relevansi Tweet terhadap Hashtag dengan Metode Logistic Regression," *CENTIVE*, vol. 2, no. 1, pp. 32–38, 2019.
- [18] A. A. Budiman and S. Widiksono, "Aplikasi Pengolahan Data Untuk Menganalisa Penggunaan Hashtag Pada Twitter," vol. 8, no. 2, 2018.
- [19] L. C. Dewi, Meiliana, and A. Chandra, "Social media web scraping using social media developers API and regex," *Procedia Comput. Sci.*, vol. 157, pp. 444–449, 2019, doi: 10.1016/j.procs.2019.08.237.
- [20] J. Gražen, M. Bertamini, and M. Volk, "Cutter - A universal multilingual tokenizer," *CEUR Workshop Proc.*, vol. 2226, no. June, pp. 75–81, 2018.
- [21] B. P. Zen, I. Susanto, and D. Finaliamartha, "TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News," vol. 6, no. 1, pp. 69–78, 2021.
- [22] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.
- [23] P. Arsi, L. N. Hidayati, and A. Nurhakim, "Komparasi Model Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform Instagram," vol. 6, pp. 459–466, 2022, doi: 10.30865/mib.v6i1.3509.
- [24] C. E. Puspita, O. N. Pratiwi, and E. Sutoyo, "PERBANDINGAN ALGORITMA KLASIFIKASI SUPPORT VECTOR MACHINE DAN NAIVE BAYES PADA IMBALANCE DATA," *Jurteksi (Jurnal Teknol. dan Sist. Informasi)*, vol. VIII, no. 1, pp. 11–18, 2021.
- [25] J. Lasama, A. P. E. P., and A. Prasetiadi, "Prediksi Tsunami Pada Gempa Menggunakan Random Forest Classifier," *Conf. Electr. Eng. Telemat. Ind. Technol. Creat. Media*, vol. 2, no. 1, pp. 39–47, 2019.
- [26] 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.