

Analisis Sentimen Masyarakat Indonesia Terhadap Vaksin Booster COVID-19

Dionisia Bhisetya Rarasati¹, Angelina Pramana Thenata², Afiyah Salsabila Arief³

^{1,2,3}Program Studi Informatika Fakultas Teknologi dan Desain Universitas Bunda Mulia, Tangerang 15143 INDONESIA (tel.: 021 80821428; email: ¹dionisia.rarasati@gmail.com, ²angelina.pramana@outlook.com, ³afiyahsalsabila@outlook.com)

[Diterima: 10 April 2023, Direvisi: 4 Agustus 2023]

Corresponding Author: Dionisia Bhisetya Rarasati

INTISARI — Pandemi COVID-19 masih terjadi di berbagai negara, termasuk Indonesia. Hal ini disebabkan oleh virus corona yang berevolusi menghasilkan berbagai varian virus, seperti Delta dan Omicron. Adapun orang Indonesia yang terkonfirmasi positif pada 9 Februari 2022 berjumlah 4.626.936. Pemerintah Indonesia telah mencegah penyebaran varian virus tersebut dengan memperkenalkan vaksin *booster* kepada masyarakat. Namun, vaksin *booster* menimbulkan berbagai sentimen di masyarakat Indonesia. Sentimen ini perlu segera diketahui pemerintah untuk mengoptimalkan upaya pencegahan penyebaran virus COVID-19. Berdasarkan permasalahan tersebut, diusulkan penerapan teknologi pembelajaran mesin untuk mengembangkan sistem yang dapat menganalisis sentimen masyarakat Indonesia terhadap vaksin *booster*. Penelitian ini memiliki beberapa tahapan, yaitu pengumpulan data, pelabelan data, prapemrosesan teks, ekstraksi fitur, dan penerapan algoritma *support vector machine* (SVM) dengan berbagai *kernel*, yaitu kernel linear, kernel Gaussian *radial basis function* (RBF), dan kernel polinomial. Selanjutnya, hasil dari sistem diuji akurasi menggunakan *10-fold cross validation* dan *confusion matrix*. *Dataset* dalam penelitian ini adalah 681 *tweet* dengan tagar “vaksinbooster”. *Dataset* terdiri atas dua kelas, yaitu negatif (0) dan positif (1). Hasil menunjukkan bahwa data ini positif untuk vaksin *booster*, terlihat dari jumlah *tweet* positif sebanyak 554 dibandingkan negatif sebanyak 127 *tweet*. Selain itu, *dataset* dibagi menjadi data latih sebanyak 545 data dan data uji sebanyak 136 data. Selanjutnya, dari hasil pengujian penelitian ini, didapatkan bahwa tingkat akurasi tertinggi diperoleh dari penggunaan algoritma SVM dengan kernel polinomial yang diuji dengan *10-fold cross validation*, yaitu sebesar 79,22%.

KATA KUNCI — COVID-19, Vaksin *Booster*, Analisis Sentimen, *Support Vector Machine*.

I. PENDAHULUAN

Virus corona, yang biasa dikenal dengan COVID-19, masih menyebar hampir di seluruh dunia, termasuk di Indonesia. Beragam varian virus COVID-19 menyebar di Indonesia: Alpha, Delta, Beta, Kappa, dan Omicron. Adapun orang Indonesia yang terkonfirmasi positif pada 9 Februari 2022 adalah sebanyak 4.626.936 orang dan jumlah ini terus bertambah di Indonesia [1]. Hal ini membuat pemerintah makin berupaya menekan penyebaran berbagai varian virus COVID-19 dengan melakukan vaksinasi massal kepada masyarakat Indonesia.

Vaksinasi di Indonesia sendiri terdiri atas dua dosis wajib dan saat ini dosis terbaru adalah vaksinasi *booster*. Pemerintah Indonesia berupaya menyosialisasikan dan mengajak masyarakat Indonesia untuk melakukan vaksinasi ulang guna menekan lonjakan kasus COVID-19. Namun, upaya pemerintah tersebut menimbulkan berbagai sentimen di kalangan masyarakat Indonesia. Sentimen tersebut perlu segera diketahui oleh pemerintah untuk mengoptimalkan upaya pencegahan varian virus COVID-19. Berbagai sentimen tersebut biasanya dituangkan ke dalam media sosial Twitter. Di media sosial tersebut pengguna dapat saling bertukar informasi melalui media teks, gambar, dan video [2].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem yang dapat menganalisis sentimen masyarakat Indonesia di media sosial Twitter terhadap vaksin *booster*. Analisis ini akan mengklasifikasikan kelompok positif dan negatif menggunakan algoritma *support vector machine* (SVM). Algoritma SVM akan mencirikan data masukan dengan mengembangkan *hyperplane* N-dimensi yang mengisolasi informasi menjadi dua jenis klasifikasi (positif dan negatif) [3]. Selanjutnya, hasil sistem diuji akurasi menggunakan *10-fold cross validation* [4] dan *confusion*

matrix [5]. Hasil analisis sentimen diharapkan dapat menjadi masukan bagi pemerintah atau pihak yang berkepentingan untuk memutuskan langkah selanjutnya dalam menekan penyebaran virus COVID-19.

Pada Bagian II, akan dibahas beberapa metode yang digunakan. Lalu, Bagian III membahas lebih lanjut metodologi yang dirumuskan untuk mencapai tujuan penelitian. Beberapa contoh dari proses metodologi juga disajikan dalam bagian ini. Kemudian, pada Bagian IV dibahas hasil beserta penerapan metodologi yang telah dirumuskan. Terakhir, Bagian V berisi kesimpulan dari penelitian yang dilakukan.

II. ANALISIS SENTIMEN

Analisis sentimen dilakukan untuk mengetahui opini publik/pengguna terhadap suatu topik. Proses analisis sentimen umumnya menggunakan data dari internet dan media/platform sosial. Sebuah kajian yang datanya bersumber dari media sosial Facebook mengkaji analisis sentimen publik terhadap calon presiden (capres) dan calon wakil presiden (cawapres) Indonesia tahun 2019 [6]. Hasil riset tersebut menyatakan bahwa capres dan cawapres Jokowi-Ma'ruf mendapatkan sentimen positif 56,76% dan sentimen negatif 43,24%, sedangkan Prabowo-Sandi mendapat sentimen positif 24,21% dan sentimen negatif 75,79%. Hasil tersebut ditemukan menggunakan algoritma *naïve Bayes* [6].

Studi analisis sentimen terhadap COVID-19 telah dilakukan pada tahun 2019 [7]. Dengan data yang bersumber dari Twitter, penelitian dilakukan menggunakan algoritma *k-nearest neighbors* (KNN) dan *naïve Bayes*. Uji sentimen menghasilkan tingkat akurasi sebesar 63,21% untuk *naïve Bayes* dan 58,10% untuk KNN. Kemudian, presisi yang didapat adalah 59,11% untuk *naïve Bayes* dan 53,10% untuk KNN. Selain itu, ditemukan kecenderungan sentimen masyarakat di

Twitter yang bersifat positif. Hal ini dibuktikan dengan jumlah sentimen positif sebanyak 610 dan negatif sebanyak 488 [7].

Sebuah penelitian yang dilakukan pada tahun 2020 membandingkan algoritma *support vector machine* (SVM), *random forest* (RF), dan *stochastic gradient descent* (SGD) untuk mengklasifikasikan kinerja (baik atau buruk) *programmer* selama beraktivitas dengan media sosial [8]. Diperoleh hasil yang akurat dengan validasi silang SVM (81,3%), RF (74,4%), dan SGD (80,1%). Hasil ini menunjukkan bahwa algoritma SVM memiliki kinerja yang lebih baik dari kedua algoritma lainnya dalam mengklasifikasikan kinerja *programmer* (baik atau buruk) selama beraktivitas menggunakan media sosial [8]. Selain itu, sebagian besar deskripsi penelitian terkait hanya mengkaji analisis sentimen menggunakan algoritma *naïve Bayes* terhadap opini publik terkait pemilihan presiden dan wabah COVID-19.

Namun, opini publik tentang vaksinasi *booster* COVID-19, yang merupakan upaya pemerintah untuk menekan lonjakan kasus COVID-19, belum dialami. Di sisi lain, algoritma SVM memiliki kinerja yang baik dalam pengelompokan kumpulan data. Oleh karena itu, penelitian ini menerapkan teknologi pembelajaran mesin untuk menganalisis sentimen masyarakat terhadap vaksinasi *booster* COVID-19 dari media sosial Twitter dengan menggunakan algoritma SVM.

III. PEMBELAJARAN MESIN

Pembelajaran mesin merupakan bagian dari kecerdasan buatan yang banyak digunakan untuk memecahkan berbagai masalah melalui pembelajaran yang bersumber dari data dalam berbagai bentuk, seperti teks, angka, gambar, dan video. Proses pembelajaran dari data yang diperoleh dari upaya tersebut dilakukan melalui dua tahap, yaitu pelatihan dan pengujian [9]. Penemuan pengetahuan yang menarik diperoleh dari data berupa teks. Pembelajaran mesin bekerja dengan berbagai algoritma, seperti *decision trees*, *naïve Bayes*, KNN, RF, dan SVM [10].

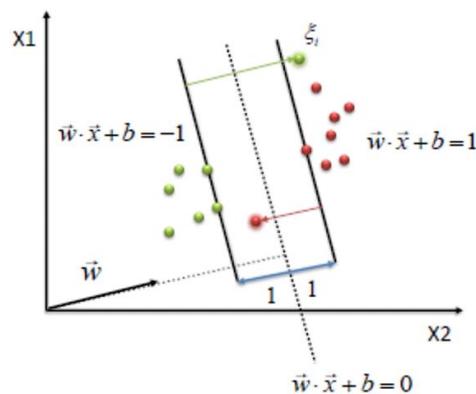
IV. SUPPORT VECTOR MACHINE

Algoritma SVM bertujuan untuk menemukan *hyperplane* terbaik yang dapat dengan sempurna memisahkan dua kelas dengan margin terluas. Margin digambarkan sebagai jarak antara *hyperplane* tersebut dengan *support vector* terdekat dari masing-masing kelas, sedangkan *support vector* dapat digambarkan sebagai titik data terjauh dari setiap kelas pada *hyperplane* tersebut [5]. Teori SVM telah berkembang sejak tahun 1960-an, tetapi baru diperkenalkan oleh Vapnik, Boser, dan Guyon pada tahun 1992. Seperti yang ditunjukkan, SVM berfungsi sebagai metode untuk menghasilkan *hyperplane* dari kumpulan data menjadi dua kelas secara linear. Pada Gambar 1, *hyperplane* di sini merupakan istilah umum untuk semua dimensi [11]. Misalnya, untuk kumpulan data satu dimensi, *hyperplane* dapat bermanifestasi sebagai titik; jika himpunan berbentuk dua dimensi, *hyperplane*-nya berupa garis lurus [12].

Gambar 1 menunjukkan bahwa sepasang *hyperplane* paralel dapat memisahkan dua kelas. Bidang batas pertama menjadi batas kelas pertama. Sebaliknya, bidang batas kedua adalah batas kelas kedua, sehingga diperoleh (1) dan (2), dengan w adalah bidang normal dan b adalah posisi bidang relatif terhadap pusat koordinat [11].

$$x_i \cdot w + b \geq +1, \text{ jika } y_i = +1 \quad (1)$$

$$x_i \cdot w + b \leq -1, \text{ jika } y_i = -1. \quad (2)$$



Gambar 1. Hyperplane SVM.

Adapun pada algoritma ini, pembelajaran *kernel* yang paling banyak digunakan meliputi *kernel* linear, RBF Gaussian, dan polinomial. Algoritma SVM dengan *kernel* ini menemukan *hyperplane* dengan pemetaan data dari ruang fitur ke ruang *kernel* berdimensi lebih tinggi. Ini mengarah pada pencapaian pemisahan nonlinier di ruang *kernel*. Selain itu, *kernel* linear dapat dinyatakan sebagai (3) [13].

$$\kappa(x_i, x_j) = x_i^T x_j. \quad (3)$$

SVM dapat diaktifkan oleh fungsi *kernel* yang sesuai, seperti RBF Gaussian nonlinear, jika ada masalah klasifikasi yang tidak dapat dipisahkan secara linear. Persamaan *kernel* ditunjukkan pada (4), dengan σ menandakan lebar *kernel* [13].

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (4)$$

Namun, dalam *kernel* RBF Gaussian, yang memiliki parameter σ sebagai lebar *kernel*, SVM akan mengalami *overfitting* di semua *instance* pelatihan jika parameternya mendekati nol. Penetapan nilai yang lebih signifikan ke σ dapat mengakibatkan *underfitting*, yaitu semua *instance* diklasifikasikan ke dalam satu kelas. Karena itu, harus dipilih nilai yang benar untuk *kernel*. Derajat *kernel* polinomial mengontrol derajat yang lebih tinggi, yang memungkinkan batas keputusan yang lebih fleksibel daripada batas linear dan fleksibilitas lebar pengklasifikasi. Persamaan *kernel* ditunjukkan pada (5), dengan p menandakan derajat polinomial *kernel* [13].

$$\kappa(x_i, x_j) = \left(1 + x_i^T x_j\right)^p. \quad (5)$$

Sementara itu, saat melatih data menggunakan SVM dan memilih fungsi *kernel*, beberapa keputusan juga perlu diambil dalam menyiapkan data, antara lain dengan memberi label dan mengatur parameter SVM untuk memberikan hasil yang optimal.

V. AKURASI

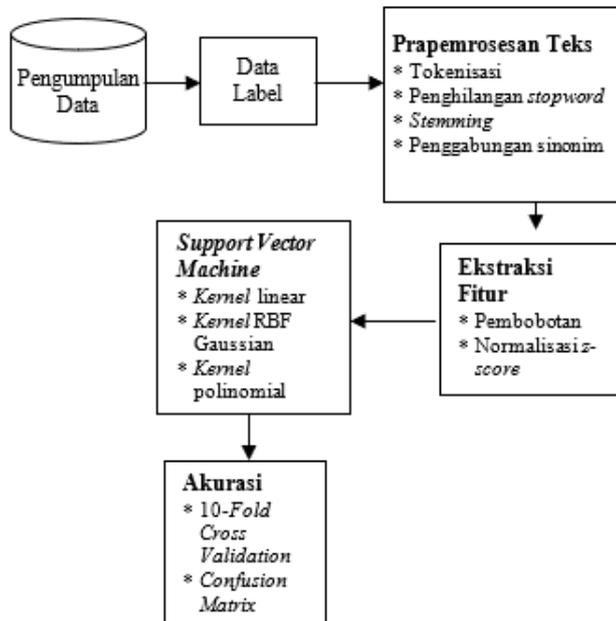
Berikut ini dijelaskan implementasi tahap-tahap pengujian.

A. CROSS VALIDATION

Cross validation merupakan salah satu metode yang dapat digunakan untuk mencari validitas model pembelajaran mesin. Metode ini bekerja dengan mempartisi set pembelajaran menjadi k subset dan, dengan demikian, diperoleh k *fold* (lipatan). Pada setiap *fold*, sebanyak $k-1$ subset akan digunakan sebagai set pelatihan dan sisanya digunakan sebagai set validasi. Prosedur ini terus diulang sampai seluruh subset telah menjadi

TABEL I
CONFUSION MATRIX 2 × 2

Nilai Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	True positive (TP)	False positive (FP)
Negatif	False Negative (FN)	True negative (TN)



Gambar 2. Alur penelitian.

set validasi. Disarankan menggunakan nilai 10 sebagai nilai terbaik untuk k saat mencoba memvalidasi model, atau yang dikenal luas sebagai *10-fold cross validation* [4].

B. CONFUSION MATRIX

Confusion matrix merupakan matriks berukuran $N \times N$, dengan N adalah jumlah kelas yang digunakan dalam klasifikasi. Matriks ini membandingkan nilai prediksi model dengan nilai sebenarnya. *Confusion matrix* yang digunakan dalam penelitian ini menggunakan matriks berukuran 2×2 , seperti terlihat pada Tabel I [5].

Tabel I menjelaskan bahwa nilai *true positive* (TP) diperoleh jika nilai prediksi dan nilai aktualnya positif. Sebaliknya, nilai *true negative* (TN) diperoleh jika nilai prediksi dan nilai sebenarnya negatif. Selanjutnya, nilai *false positive* (FP) diperoleh jika prediksinya positif, tetapi nilai aktualnya negatif; sedangkan nilai *false negative* (FN) diperoleh jika prediksi negatif, tetapi nilai aktualnya positif. Sementara itu, rumus akurasi dalam pengujian ini ditunjukkan pada (6). Persamaan tersebut menyatakan bahwa makin besar nilai TN dan TP yang ditemukan, makin tinggi tingkat akurasi [5].

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

VI. METODOLOGI

Penelitian dilakukan dengan alur penelitian yang ditunjukkan pada Gambar 2.

A. TAHAP PENGUMPULAN DATA DAN TAHAP PELABELAN

Data yang digunakan dalam penelitian ini bersumber dari media sosial Twitter berbahasa Indonesia, sejalan dengan tujuan penelitian untuk mengembangkan sistem yang dapat menganalisis sentimen masyarakat Indonesia di media sosial

TABEL II
TAHAPAN PROSES *TOKENIZING*

Tahapan	Hasil
Tweet awal	DPP Projo mengadakan vaksin booster Covid-gratis untuk rakyat selama lima hari untuk masyarakat DKI Jakarta
Pengubahan ke huruf kecil	dpp projo mengadakan vaksin booster covid-gratis untuk rakyat selama lima hari untuk masyarakat dki jakarta
Penghapusan tanda baca	dpp projo mengadakan vaksin booster covid gratis untuk rakyat selama lima hari untuk masyarakat dki jakarta
Pemisahan kata sebagai <i>token</i> individu	“dpp” “projo” “mengadakan” “vaksin” “booster” “covid” “gratis” “untuk” “rakyat” “selama” “lima” “hari” “untuk” “masyarakat” “dki” “jakarta”

TABEL III
TAHAPAN PROSES MENGHILANGKAN *STOPWORD*

Tahapan	Hasil
Sebelum <i>stopword</i> dihapus	“dpp” “projo” “mengadakan” “vaksin” “booster” “covid” “gratis” “untuk” “rakyat” “selama” “lima” “hari” “untuk” “masyarakat” “dki” “jakarta”
Setelah <i>stopword</i> dihapus	“dpp” “projo” “mengadakan” “vaksin” “booster” “covid” “gratis” “rakyat” “lima” “masyarakat” “dki” “jakarta”

Twitter terhadap vaksin *booster*. Data tersebut dikumpulkan dari sumber data menggunakan skrip Python dengan kata kunci “vaksinbooster”. Data *tweet* yang digunakan adalah 681 data berbahasa Indonesia dari tanggal 12 Januari 2022 hingga 12 April 2022. Kemudian, tahap berikutnya adalah pelabelan, yaitu pemberian label kelas berdasarkan dua kelas (positif dan negatif).

B. TAHAP PRAPEMROSESAN TEKS

Tahap ini menyiapkan data agar mesin dapat menganalisisnya dengan mudah [10]. Ada beberapa langkah dalam tahap ini, yaitu *tokenizing*, menghilangkan *stopwords*, dan *stemming*. Langkah-langkah ini dilakukan untuk menghilangkan derau (*noise*) pada *tweet*, seperti tautan URL, dan *username*. Selain itu, sistem akan mengubah kata atau singkatan bahasa Indonesia nonformal menjadi kata yang mematuhi Kamus Besar Bahasa Indonesia (KBBI). Sistem juga akan mengekstrak kata yang diawali dengan tagar.

1) *TOKENIZING*

Tahap pertama dalam prapemrosesan adalah *tokenizing* [12]. Beberapa proses dilakukan pada tahap ini, yaitu pengubahan ke huruf kecil, penghapusan tanda baca, dan memecah kalimat menjadi kata-kata terpisah [14]. Prosesnya dapat dilihat pada Tabel II.

2) MENGHILANGKAN *STOPWORD*

Tahap kedua adalah menghilangkan *stopword*, yaitu kata-kata dengan frekuensi tinggi dalam teks yang tidak memiliki arti khusus atau tidak penting. Contoh *stopword* dalam bahasa Indonesia adalah “dan”, “kita”, “sebuah”, dan “adalah” [15]. Proses ini dapat dilihat pada Tabel III, yaitu mencari *stopword* dalam *tweet* dan jika ditemukan, *stopword* tersebut akan dihapus.

3) *STEMMING*

Tahap ketiga adalah *stemming*, yaitu menghilangkan imbuhan untuk mencari kata dasar atau *root word*. Hal ini bertujuan untuk mengurangi jumlah kata yang diproses dalam

TABEL IV
 TAHAPAN PROSES STEMMING

Tahapan	Contoh
Token	mengadakan
Imbuhan	meng-ada-kan
Hasil stemming	ada

penambahan teks (*text mining*), sehingga waktu pemrosesan berkurang dan memori yang digunakan tidak terlalu banyak [16]. Hasil dari proses *stemming* ditunjukkan pada Tabel IV.

4) PENGGABUNGAN SINONIM

Tahap terakhir dalam prapemrosesan adalah menggabungkan sinonim atau bentuk bahasa dengan makna yang mirip. Pada tahap ini, jika kata yang berbeda memiliki arti sama, kata tersebut akan digabungkan. Sebagai contoh, kata “saya” dan “aku” memiliki arti yang sama. Tahapan ini bertujuan untuk meminimalkan jumlah kata dalam sistem tanpa menghilangkan frekuensi [3].

C. TAHAP EKSTRAKSI FITUR

Ekstraksi fitur merupakan tahapan selanjutnya setelah prapemrosesan. Karena data tekstual (*tweet* dalam hal ini) adalah opini publik, penulisannya sering kali tidak terstruktur. Oleh karena itu, perlu dilakukan transformasi data tekstual menjadi data terstruktur agar algoritma pembelajaran mesin dapat segera bekerja [17]. Terdapat dua proses dalam tahap ini, yaitu pembobotan dan normalisasi *z-score*.

1) PEMBOBOTAN

Pembobotan merupakan tahapan yang mencerminkan tingkat pentingnya sebuah kata dalam dokumen. Metode yang digunakan untuk melakukan pembobotan pada penambahan teks adalah *term frequency - inverse document frequency* (TF-IDF), yang rumusnya dituliskan pada (7). Bobot akan bertambah dengan naiknya frekuensi kemunculan *term* dalam teks. Namun, peningkatan ini juga sejalan dengan frekuensi kemunculan istilah-istilah yang termasuk dalam domain penelitian yang diminati [18].

$$W_{t,d} = tf_{t,d} \times idf_t \tag{7}$$

dengan $W_{t,d}$ menunjukkan bobot, $tf_{t,d}$ menunjukkan frekuensi istilah dari kata (*term frequency*, TF), dan idf_t menunjukkan frekuensi dokumen terbalik (*inverse document frequency*, IDF).

2) NORMALISASI Z-SCORE

Proses normalisasi *z-score* merupakan tahapan yang dilakukan setelah nilai bobot diperoleh. Karena perbedaan nilai *range* yang signifikan akan berdampak pada permasalahan klasifikasi nantinya, harus dilakukan normalisasi [19]. Untuk melakukan normalisasi *z-score*, dapat digunakan (8) [20].

$$Z = \frac{x - \bar{x}}{s} \tag{8}$$

dengan Z adalah normalisasi *z-score*, x adalah nilai dari data, \bar{x} adalah rerata data, dan s adalah standar deviasi.

D. TAHAP SUPPORT VECTOR MACHINE (SVM)

Sistem akan mengelompokkan *tweet* menjadi dua kluster, yaitu positif dan negatif. Sebuah *hyperplane* akan mengelompokkan setiap *tweet*. Dalam penggunaan algoritma SVM, terdapat tiga *kernel* yang dapat digunakan untuk mencari *hyperplane* terbaik, yaitu *kernel* RBF Gaussian, *kernel* linear, dan *kernel* polinomial [13]. Oleh karena itu, pada penelitian ini, akurasi ketiga *kernel* tersebut dibandingkan satu sama lain dan *kernel* terbaik yang ditemukan akan digunakan sebagai hasil analisis sentimen berdasarkan hasil akurasinya.

Akurasi Confusion Matrix

Total Data Latih : 545

Total Data Uji : 136

n=136	Aktual : Tidak	Aktual : Ya	
Prediksi : Tidak	8	7	15
Prediksi : Ya	29	92	121
	99	37	

Precision	0,93
Recall	0,76
F Measure	0,84
Error Rate	0,26
Akurasi	74%

Gambar 3. Akurasi confusion matrix dengan kernel linear.

10-Fold Cross Validation

	Data Uji	Data Latih
K-1	79,71%	612 69
K-2	82,61%	612 69
K-3	75,36%	612 69
K-4	79,71%	612 69
K-5	63,77%	612 69
K-6	75,36%	612 69
K-7	85,51%	612 69
K-8	72,46%	612 69
K-9	69,57%	612 69
K-10	76,67%	621 60

AKURASI : 76,07%

Gambar 4. Akurasi 10-fold cross validation dengan kernel linear.

E. TAHAP AKURASI

Hasil akan diuji akurasinya dengan *10-fold cross validation* dan *confusion matrix*. Hasil pengujian kedua metode tersebut dibandingkan dan hasil terbaik akan digunakan untuk menguji akurasi analisis sentimen pada penelitian ini.

VII. HASIL DAN PEMBAHASAN

Dataset yang terkumpul di media sosial Twitter dengan tagar “vaksinbooster” berjumlah 681 *tweet*. Data ini terdiri atas dua kelas, yaitu negatif (0) dan positif (1). Selain itu, data menunjukkan bahwa sentimen masyarakat terhadap vaksin *booster* cenderung positif, yang terlihat dari jumlah *tweet* positif sebanyak 554 dan *tweet* negatif sebanyak 127 *tweet*. *Dataset* dibagi menjadi data latih sebanyak 545 dan data uji sebanyak 136. Selanjutnya, terhadap *dataset* tersebut dilakukan prapemrosesan teks yang menghasilkan larik kata-kata penting dari sebuah *tweet*. Lalu, dilakukan tahap ekstraksi fitur yang menghasilkan nilai bobot setiap kata dari tahap prapemrosesan teks. Nilai bobot hasil ekstraksi kemudian diolah sebagai masukan menggunakan algoritma SVM yang diproses dengan tiga *kernel*, yaitu *kernel* linear (lihat Gambar 3 dan Gambar 4), *kernel* RBF Gaussian (lihat Gambar 4), dan *kernel* polinomial (lihat Tabel V dan Tabel VI).

Gambar 3 menjelaskan hasil pengujian algoritma SVM dengan *kernel* linear pada analisis sentimen vaksin *booster*. Pengujian dengan metode *confusion matrix* memberikan hasil data prediksi positif mengikuti fakta positif sebanyak 92 data, prediksi negatif mengikuti realitas negatif sebanyak delapan data prediksi positif. Namun realitas negatif ditemukan sebanyak 29 data dan data prediksi negatif, sedangkan realitas positif ditemukan sebanyak tujuh data. Oleh karena itu, hasil tingkat akurasi dengan metode *confusion matrix* adalah sebesar

TABEL V
HASIL *CONFUSION MATRIX*

$n = 136$	Aktual: No	Aktual: Yes	Jumlah
Prediksi: No	0	0	0
Prediksi: Yes	37	99	136
	37	99	

TABEL VI
HASIL *10-FOLD CROSS VALIDATION*

Fold	Akurasi	Pelatihan	Pengujian
k-1	88,41%	612	69
k-2	82,61%	612	69
k-3	81,16%	612	69
k-4	81,16%	612	69
k-5	75,36%	612	69
k-6	78,26%	612	69
k-7	86,96%	612	69
k-8	75,36%	612	69
k-9	69,57%	612	69
k-10	73,33%	621	60

74%. Sementara itu, pengujian dengan metode *10-fold cross validation* mendapatkan tingkat akurasi yang lebih tinggi, yaitu sebesar 76,07%, seperti dapat dilihat pada Gambar 4.

Gambar 5 menunjukkan hasil pengujian algoritma SVM dengan *kernel* RBF Gaussian pada analisis sentimen vaksin *booster*. Pengujian dengan metode *confusion matrix* memberikan hasil data prediksi positif yang sesuai dengan realitas positif sebanyak 99 dan data prediksi negatif yang sesuai dengan realitas negatif sebanyak 0. Sementara itu, data prediksi positif, tetapi realitas negatif ditemukan sebanyak 37 data dan data prediksi negatif, tetapi realitas positif ditemukan sebanyak 0 data. Oleh karena itu, metode pengujian *confusion matrix* mendapatkan hasil akurasi sebesar 73%. Di sisi lain, pengujian dengan metode *10-fold cross validation* mendapatkan tingkat akurasi yang lebih tinggi, yaitu sebesar 78,2%, seperti ditunjukkan pada Gambar 6.

Penelitian pada tahun 2019 tentang klasifikasi *human development index* (HDI) menggunakan SVM dengan *kernel* RBF Gaussian mendapatkan akurasi sebesar 98,1%, lebih tinggi dibandingkan dengan penerapan *kernel* linear yang memiliki akurasi sebesar 95,1% [21].

Selanjutnya, baik Tabel V maupun Tabel VI memaparkan hasil pengujian algoritma SVM dengan *kernel* polinomial pada analisis sentimen vaksin *booster*. Pengujian dengan metode *confusion matrix* memberikan hasil yang sama dengan pengujian algoritma SVM menggunakan *kernel* RBF Gaussian, yaitu memiliki tingkat akurasi sebesar 73%. Hasil ini dapat dilihat pada Tabel V. Namun, pengujian dengan menggunakan metode *10-fold cross validation* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan *confusion matrix*, yaitu rata-rata sebesar 79,22%, seperti dapat dilihat pada Tabel VI. Dengan demikian, penerapan algoritma SVM dengan *kernel* terbaik terdapat pada *kernel* polinomial, yang memiliki tingkat akurasi paling tinggi menggunakan metode uji *10-fold cross validation*, yaitu sebesar 79,22%.

Sebuah penelitian pada tahun 2020 menganalisis sentimen layanan OVO di Twitter menggunakan algoritma SVM. Penelitian tersebut menemukan bahwa penggunaan *kernel* polinomial pada proses SVM memiliki tingkat akurasi yang tinggi, yaitu 89,09% [22]. Selain itu, penelitian terkait analisis sentimen kebijakan pembatasan sosial berskala besar (PSBB) di Twitter menemukan bahwa algoritma SVM dengan *kernel*

Akurasi *Confusion Matrix*

Total Data Latih : 545
Total Data Uji : 136

n=136	Aktual : Tidak	Aktual : Ya	
Prediksi : Tidak	0	7	0
Prediksi : Ya	37	99	136
	99	37	

Precision	1
Recall	0,73
F Measure	0,84
Error Rate	0,27
Akurasi	73%

Gambar 5. Akurasi *confusion matrix* dengan *kernel* RBF Gaussian.

10-Fold Cross Validation

	Data Uji	Data Latih	
K-1	88,41%	612	69
K-2	82,61%	612	69
K-3	76,81%	612	69
K-4	81,16%	612	69
K-5	69,57%	612	69
K-6	78,26%	612	69
K-7	86,96%	612	69
K-8	75,36%	612	69
K-9	69,57%	612	69
K-10	73,33%	621	60

AKURASI : 78,2%

Gambar 6. Akurasi *10-fold cross validation* dengan *kernel* RBF Gaussian.

polinomial memiliki tingkat akurasi sebesar 94,55% [23]. Berdasarkan hasil penelitian ini, dapat dinyatakan bahwa algoritma SVM dengan *kernel* polinomial merupakan yang paling baik dalam menganalisis sentimen vaksin *booster* di media sosial Twitter.

VIII. KESIMPULAN

Vaksinasi di Indonesia terdiri atas dua dosis wajib dan saat ini dosis terbaru adalah vaksinasi *booster*. Namun, vaksin *booster* menimbulkan berbagai sentimen di masyarakat Indonesia. Oleh karena itu, diusulkan penerapan teknologi pembelajaran mesin untuk mengembangkan sistem yang dapat menganalisis sentimen masyarakat Indonesia terhadap vaksin *booster* dengan menggunakan algoritma SVM. Penelitian ini mengumpulkan 681 data Twitter Indonesia dengan tagar “vaksinbooster” yang terbagi menjadi negatif (0) dan positif (1). Selain itu, *dataset* dibagi menjadi data latih sebanyak 545 data dan data uji sebanyak 136 data. Data menunjukkan bahwa sentimen masyarakat terhadap vaksin *booster* cenderung positif, terlihat dari jumlah *tweet* positif sebanyak 554 *tweet* dan *tweet* negatif sebanyak 127 *tweet*.

Hasil pengujian sistem menggunakan *10-fold cross validation* menemukan bahwa algoritma SVM dengan *kernel* linear mendapatkan tingkat akurasi sebesar 76,07%, sedangkan *kernel* RBF Gaussian memperoleh tingkat akurasi sebesar 78,2%. Sementara itu, *kernel* polinomial memperoleh tingkat akurasi sebesar 79,22%. Pada pengujian sistem menggunakan metode *confusion matrix*, diperoleh hasil bahwa algoritma SVM dengan *kernel* linear mendapatkan tingkat akurasi sebesar 74%, sedangkan *kernel* RBF Gaussian memperoleh tingkat akurasi 73%. Selanjutnya, *kernel* polinomial memperoleh tingkat akurasi sebesar 73%. Dengan demikian, penerapan algoritma SVM dengan *kernel* terbaik terdapat pada *kernel* polinomial, yang memiliki tingkat akurasi tertinggi menggunakan metode uji *10-fold cross validation*.

KONFLIK KEPENTINGAN

Semua penulis menyatakan bahwa tidak ada konflik kepentingan dalam penelitian ini.

KONTRIBUSI PENULIS

Konseptualisasi, Dionisia Bhisetya Rarasati; metodologi, Dionisia Bhisetya Rarasati; penulisan—penyusunan draf asli, Angelina Pramana Thenata dan Afiyah Salsabila Arief; penulisan—peninjauan dan penyuntingan, Dionisia Bhisetya Rarasati, Angelina Pramana Thenata, dan Afiyah Salsabila Arief.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Kemenristekdikti yang telah membiayai penelitian ini melalui Hibah Penelitian Dosen Pemula tahun anggaran 2022 dengan nomor kontrak: 155/E5/PG.02.00.PT/2022.

REFERENSI

- [1] (2022) "Peta Sebaran COVID-19," [Online], <https://covid19.go.id/peta-sebaran#>, tanggal akses: 14-Mar-2022.
- [2] D.A. Agustina, S. Subanti, dan E. Zukhronah, "Implementasi Text Mining pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," *Indones. J. Appl. Stat.*, Vol. 3, No. 2, hal. 109–122, Nov. 2020, doi: 10.13057/ijas.v3i2.44337.
- [3] D.B. Rarasati, "A Grouping of Song-Lyric Themes Using K-Means Clustering," *JISA (J. Inform., Sains)*, Vol. 3, No. 2, hal. 38–41, Des. 2020, doi: 10.31326/jisa.v3i2.658.
- [4] D. Berrar, "Cross-Validation," dalam *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, Vol. 1, S. Ranganathan, K. Nakai, C. Schönbach, Eds., Amsterdam, Belanda: Elsevier, 2019, hal. 542–545, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [5] Y. Sari, P.B. Prakoso, dan A.R. Baskara, "Road Crack Detection Using Support Vector Machine (SVM) and OTSU Algorithm," *2019 6th Int. Conf. Electr. Veh. Technol. (ICEVT)*, 2019, hal. 349–354, doi: 10.1109/ICEVT48285.2019.8993969.
- [6] B. Haryanto dkk., "Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates From Facebook Opinion Data," *Procedia Comput. Sci.*, Vol. 161, hal. 715–722, 2019, doi: 10.1016/j.procs.2019.11.175.
- [7] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai COVID-19 pada Twitter Menggunakan Metode Naïve Bayes dan K-NN," *Inti Nusa Mandiri*, Vol. 15, No. 1, hal. 23–28, Agu. 2020, doi: 10.33480/inti.v15i1.1347.
- [8] R. Umar, I. Riadi, dan Purwono, "Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial," *J. RESTI (Rekayasa Sist., Teknol. Inf.)*, Vol. 4, No. 2, hal. 329–335, Apr. 2020, doi: 10.29207/resti.v4i2.1770.
- [9] A. Roihan, P.A. Sunarya, dan A.S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," *IJCIT (Indones. J. Comput. Inf. Technol.)*, Vol. 5, No. 1, hal. 75–82, Mei 2020, doi: 10.31294/ijcit.v5i1.7951.
- [10] A.P. Thenata, "Text Mining Literature Review on Indonesian Social Media," *J. Eduk., Penelit. Inform.*, Vol. 7, No. 2, hal. 226–232, Agu. 2021, doi: 10.26418/jp.v7i2.47975.
- [11] F.S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist., Teknol. Inf.)*, Vol. 1, No. 1, hal. 19–25, Apr. 2017, doi: 10.29207/resti.v1i1.11.
- [12] S. Symeonidis, D. Effrosynidis, dan A. Arampatzis, "A Comparative Evaluation of Pre-Processing Techniques and Their Interactions for Twitter Sentiment Analysis," *Expert Syst. Appl.*, Vol. 110, hal. 298–310, Nov. 2018, doi: 10.1016/j.eswa.2018.06.022.
- [13] C. Savas dan F. DAVIS, "The Impact of Different Kernel Functions on the Performance of Scintillation Detection Based on Support Vector Machines," *Sens.*, Vol. 19, No. 23, hal. 1–16, Nov. 2019, doi: 10.3390/s19235219.
- [14] V.Y. Radygin dkk., "Application of Text Mining Technologies in Russian language for Solving the Problems of Primary Financial Monitoring for Solving the Problems of Primary Financial Monitoring," *Procedia Comput. Sci.*, Vol. 190, hal. 678–683, 2021, doi: 10.1016/j.procs.2021.06.078.
- [15] H. Najjichah, A. Syukur, dan H. Subagyo, "Pengaruh Text Preprocessing dan Kombinasinya pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia," *J. Teknol. Inf. Cyberku*, Vol. 15, No. 1, hal. 1–11, Jan. 2019.
- [16] D. Sebastian, "Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari Beberapa E-marketplace," *J. Tek. Inform., Sist. Inf.*, Vol. 5, No. 1, hal. 51–61, Apr. 2019, doi: 10.28932/jutisi.v5i1.1581.
- [17] H. Liu, P. Burnap, W. Alorainy, dan M.L. Williams, "A Fuzzy Approach to Text Classification with Two-Stage Training for Ambiguous Instances," *IEEE Trans. Comput. Soc. Syst.*, Vol. 6, No. 2, hal. 227–240, Apr. 2019, doi: 10.1109/TCSS.2019.2892037.
- [18] I. Yahav, O. Shehory, dan D. Schwartz, "Comments Mining With TF-IDF: The Inherent Bias and Its Removal," *IEEE Trans. Knowl., Data Eng.*, Vol. 31, No. 3, hal. 437–450, Mar. 2019, doi: 10.1109/TKDE.2018.2840127.
- [19] Henderi, W. Tri, dan R. Efan, "Comparison of Min-Max Normalization and Z-Score Normalization in the K-Nearest Neighbor (KNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIS Int. J. Inform., Inf. Syst.*, Vol. 4, No. 1, hal. 13–20, Mar. 2021, doi: 10.47738/ijisi.v4i1.73.
- [20] M.A. Imron dan B. Prasetyo, "Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn," *J. Soft Comput. Explor.*, Vol. 1, No. 1, hal. 56–62, Sep. 2020, doi: 10.52465/jossex.v1i1.7.
- [21] H. Al Azies, D. Trishnanti, dan E. Mustikawati P.H., "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," *IPTEK J. Proc. Ser.*, No. 6, hal. 53–57, Nov. 2019, doi: 10.12962/j23546026.y2019i6.6394.
- [22] F. Romadoni, Y. Umaidah, dan B.N. Sari, "Text Mining untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," *J. Sisfokom (Sist. Inf., Komput.)*, Vol. 9, No. 2, hal. 247–253, Jul. 2020, doi: 10.32736/sisfokom.v9i2.903.
- [23] H.P.P. Zuriel and A. Fahrurozi, "Implementasi Algoritma Klasifikasi Support Vector Machine untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan PABB," *J. Ilm. Inform. Komput.*, Vol. 26, No. 2, hal. 149–162, Agu. 2021, doi: 10.35760/ik.2021.v26i2.4289.