

ANALISIS SENTIMEN VAKSINASI DI INDONESIA MENGGUNAKAN OPTIMIZED-SVM

Kisma A^{1,*}, Muhammad Nur Yasir Utomo, S.ST, M.Eng B², Eddy Tungadi, S.T., M.T. C³
^{1,2,3}Teknik Elektro Politeknik Negeri Ujung Pandang, Makassar

ABSTRACT

Coronavirus Disease (COVID-19) is a disease caused by a new type of corona virus, namely Sars-CoV-2, which is a dangerous virus because the rate of spread is very easy and fast. To minimize the impact of COVID-19, the government issued a mandatory vaccine policy. However, the existence of this policy raises pros and cons among the public. Freedom of expression carried out by users of social media (twitter) makes information spread quickly. One method that supports the concept of opinion analysis on texts and supports text processing is sentiment analysis. This study uses the Support Vector Machine (SVM) algorithm for data classification and the Particle Swarm Optimization (PSO) algorithm for parameter optimization. The results of the validation and evaluation using 3 k-fold cross validation showed that PSO was able to optimize the parameters of the SVM model with performance before optimization had 95% Accuracy, F1-measure 94.5%, Precision 93.8%, Recall 96.1% after implementation. parameter optimization increased with the performance of Accuracy 98.3%, F1-measure 98.2%, Precision 98.1%, Recall 98.6% on 2 label data. Meanwhile, for the 3 label data before the parameter optimization was applied, the percentage value of Accuracy 67.70 %, F1-measure 67.85 %, Precision 70.4%, Recall 67.5% and after the parameter optimization was applied increased with the percentage of Accuracy 73.3%, F1-measure 73.9%, Precision 76.9%, Recall 73.2%.

Keywords: COVID-19, Vaccination, Parameter Optimization, Support Vector Machine (SVM), Particle Swarm Optimization (PSO).

ABSTRAK

Coronavirus Disease (COVID-19) merupakan penyakit yang disebabkan oleh virus corona jenis baru, yaitu Sars-CoV-2, yang merupakan salah satu virus yang berbahaya karena tingkat penyebarannya sangat mudah dan cepat. Untuk meminimalisir dampak dari COVID-19, pemerintah mengeluarkan kebijakan wajib vaksin. Namun, dengan adanya kebijakan tersebut menimbulkan Pro dan Kontra di kalangan masyarakat. Kebebasan berpendapat yang dilakukan oleh pengguna media social (twitter) membuat informasi menyebar dengan cepat. Salah satu metode yang mendukung konsep analisis opini pada teks dan mendukung dalam melakukan pemrosesan teks adalah analisis sentimen. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) untuk klasifikasi data dan algoritma Particle Swarm Optimization (PSO) untuk optimasi parameter. Hasil validasi dan evaluasi menggunakan 3 *k-fold cross validation* menunjukkan bahwa PSO mampu mengoptimasikan parameter model SVM dengan performa sebelum diterapkan optimasi memiliki Accuracy 95%, F1-measure 94,5%, Precision 93,8 %, Recall 96,1% setelah diterapkan optimasi parameter meningkat dengan performa Accuracy 98,3 %, F1-measure 98,2%, Precision 98,1%, Recall 98,6% pada data 2 label. Sedangkan untuk data 3 label sebelum diterapkan optimasi parameter didapatkan persentase nilai Accuracy 67,70 %, F1-measure 67,8,5%, Precision 70,4 %, Recall 67,5% dan setelah diterapkan optimasi parameter meningkat dengan persentase nilai Accuracy 73,3 %, F1-measure 73,9%, Precision 76,9 %, Recall 73,2%.

Kata Kunci : COVID-19, Vaccination, Parameter Optimization, Support Vector Machine (SVM), Particle Swarm Optimization (PSO).

1. PENDAHULUAN

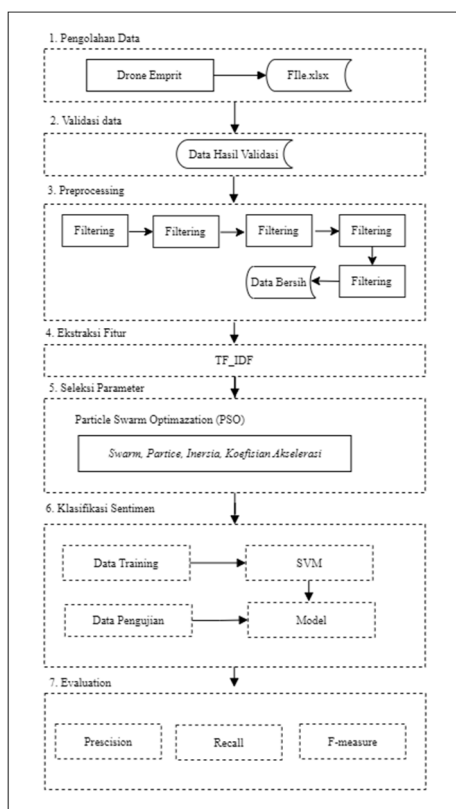
Coronavirus Disease 2019 (COVID-19) secara resmi ditetapkan sebagai pandemi global oleh World Health Organization (WHO) pada tanggal 11 Maret 2020 [1]. COVID-19 merupakan penyakit yang disebabkan oleh virus corona jenis baru, yaitu Sars-CoV-2, dimana virus tersebut termasuk salah satu virus yang sangat berbahaya karena tingkat penyebarannya sangat mudah dan cepat [2]. Di Indonesia sendiri pasien yang terinfeksi oleh COVID-19 telah menyebar ke 34 Provinsi dan 432 Kabupaten/Kota [3]. Berbagai upaya yang dilakukan oleh pemerintah untuk menekan tingkat penyebaran COVID-19, salah satunya dengan melakukan vaksinasi [4]. Vaksin COVID-19 sempat menjadi trending topik di twitter karena ramai dibahas oleh masyarakat Indonesia. Kebebasan berpendapat yang dilakukan oleh pengguna media social (twitter) membuat informasi menyebar dengan cepat sehingga menimbulkan berbagai Pro dan Kontra di kalangan masyarakat. Salah satu cara untuk mendapatkan gambaran dari keseluruhan tanggapan masyarakat adalah dengan menggunakan sosial media, salah satunya adalah twitter. Masyarakat menggunakan twitter untuk menuliskan opininya terhadap suatu objek atau kejadian [5]. Opini masyarakat pada twitter ini berpeluang mampu memberikan gambaran keseluruhan tanggapan masyarakat terhadap kebijakan pemerintah, yang tentunya dapat dijadikan sebagai patokan dalam menentukan strategi ataupun evaluasi bagi pemerintah dalam membuat kebijakan baru di kemudian hari.

* Korespondensi penulis: Kisma, kismakimmaa01@gmail.com

Namun, data twitter belum terstruktur dan memiliki banyak tanda baca serta penggunaan bahasa yang tidak baku. Sehingga diperlukan sebuah metode untuk melakukan pemrosesan teks agar data tersebut dapat di olah dengan baik [6]. Salah satu metode yang mendukung dalam melakukan pemrosesan teks adalah analisis sentimen. Analisis sentimen ini digunakan untuk menyaring opini masyarakat pada media sosial seperti twitter dan mengklasifikasikannya ke dalam kelas positif, netral, dan negatif [6]. Algoritma *Machine Learning* telah banyak digunakan untuk menyelesaikan berbagai macam klasifikasi data [7], seperti Algoritma Support Vector Machine (SVM). Namun, kinerja dari SVM ini memiliki banyak parameter dan sangat bergantung pada pengaturan nilai parameter dan pemilihan kernelnya. Karena Setiap data yang digunakan memiliki nilai parameter yang berbeda-beda. Kesalahan dalam mengatur parameter dari SVM dapat menyebabkan hasil klasifikasi yang buruk. Meskipun SVM bekerja dengan baik dengan nilai *default*, kinerja SVM dapat ditingkatkan secara signifikan dengan menggunakan optimasi parameter [8]. Oleh sebab itu, untuk mengatasi masalah tersebut, maka dilakukanlah optimasi parameter. Terdapat beberapa algoritma yang direkomendasikan untuk melakukan optimasi parameter pada machine learning seperti Genetic Algorithm (GA), Bat Algorithm (BA), Flower Pollination Algorithm (FPA), dan Particle Swarm Optimization (PSO) [5]. PSO merupakan algoritma yang digunakan untuk menentukan parameter optimal.

Penelitian mengenai klasifikasi sentimen dengan menggunakan metode SVM telah dilakukan sebelumnya oleh [9] tentang Analisis Sentimen Penghapusan Ujian Nasional pada Twitter Menggunakan *Support Vector Machine dan Naïve Bayes* berbasis *Particle Swarm Optimization*. Dalam penelitian ini proses klasifikasi menggunakan tools berupa rapid miner. Hasil dari penelitian ini metode SVM dengan PSO yang paling baik untuk mengklasifikasikan sentimen positif dan negatif terkait penghapusan UN dengan akurasi 94,81%. Pada penelitian ini akan menggunakan algoritma SVM untuk memproses atau klasifikasi data dan algoritma Particle Swarm Optimization (PSO) sebagai optimasi parameter menggunakan bahasa pemrograman python. Parameter yang digunakan pada SVM adalah nilai C dan Gamma. Sedangkan parameter yang digunakan pada PSO adalah *Population size* atau *swarm*, Koefisien akselerasi, dan *Inertia Weight*. Penelitian ini diharapkan dapat membantu pemerintah untuk mengetahui tanggapan ataupun kekhawatiran masyarakat terhadap vaksin COVID-19, sehingga pemerintah bisa melakukan evaluasi dan menentukan strategi selanjutnya terkait edukasi maupun sosialisasi tentang vaksin COVID-19 kepada masyarakat.

2. METODE PENELITIAN



Gambar 1. Prosedur Penelitian

Berdasarkan gambar diatas, tahapan penelitian dimulai dari (1) Studi Literatur yang bertujuan untuk mengumpulkan data terlebih dahulu seperti jurnal dan tugas akhir yang berkaitan dengan studi kasus yang diambil, untuk dijadikan sebagai bahan perbandingan untuk melakukan analisis sentimen pada penelitian ini, (2) Pengambilan Data yang dilakukan melalui website Drone Emprit karena data ini telah di labeli dengan 3 indikator yaitu positif, netral, dan negatif sehingga tidak perlu dilakukan pelabelan secara manual. Pengambilan data dilakukan dengan rentang waktu bulan Agustus 2021 sampai bulan Juni 2022, (3) *Preprocessing* yang

terdiri dari beberapa tahapan yaitu : *Filtering* merupakan Teknik *filtering* menggunakan *library re* untuk menghilangkan *string* seperti angka, website, gambar, tanda baca, hastag, @user dan karakter lainnya, *Case Folding* bertujuan untuk melakukan penyeragaman kata dengan mengubah semua huruf mejadi huruf kecil menggunakan *function lowercase ()*, *Stopword Removal* digunakan untuk melakukan filter kata yang bermakna stopword. *Stopword* ini berisi kata penghubung antar kata seperti : dan, ke, atau, yang, sebagai, di, dari, kemana, dimana tetapi dan lainnya. Teknik ini menggunakan *library sastrawi*, *Stemming* digunakan untuk mengubah kata-kata dalam satu dokumen menjadi bentuk kata dasar dan menghapus prefix (kata awalan) dan sufiks (kata akhiran). Teknik ini menggunakan *library sastrawi* dengan fungsi *StemmerFactory*, dan *Tokenizing* bertujuan untuk memisahkan *tweet* dalam bentuk kata-kata. Proses ini dilakukan dengan menggunakan *library NLTK* pada *python*.

Tahapan (4) ekstraksi fitur, Hasil dari preprocessing dilakukan pembagian data training sebanyak 90 % dan data testing sebanyak 10%. Semakin banyak data training maka semakin baik pula model klasifikasi. Begitu pun sebaliknya, semakin banyak data testing semakin banyak terjadi kesalahan. Selanjutnya, dilakukan pembobotan kata dengan menggunakan TF-IDF. Pembobotan kata ini menggunakan *library TfidfVectorizer* untuk mengubah data bentuk teks menjadi data vektor dan memberikan bobot pada setiap kata, serta menyaring data yang akan diproses. Prinsip kerja dari TF-IDF ini yaitu, semakin besar bobot suatu kata, maka kata tersebut dianggap penting, sedangkan kata yang bernilai 0 secara otomatis tidak akan diproses atau ditampilkan. Berikut merupakan tahapan dalam melakukan ekstraksi fitur menggunakan TF-IDF, (5) Optimasi Parameter, yang dilakukan dengan mencari nilai parameter C dan Gamma secara acak, kemudian di evaluasi dengan menghitung nilai *accuracy* dan nilai *fitness* atau *error* pada setiap parameter dan di update secara terus menerus sampai mendapatkan nilai *fitness* atau *error* yang paling rendah. Parameter yang memiliki nilai *fitness* atau *error* yang paling rendah merupakan parameter yang optimal dan akan digunakan dalam proses klasifikasi, (6) Validasi dan Evaluasi, dilakukan pada *python* menggunakan *library sklearn*. Validasi ini bertujuan untuk mengetahui seberapa baik dan akurat model klasifikasi yang telah dibuat . Validasi model klasifikasi dilakukan dengan menggunakan 3, 5, dan 10 *K-Fold Cross validation* dan di evaluasi menggunakan *confusion matrix*. Terdapat indikator yang diukur dalam *confusion matrix* yaitu *accuracy*, *precision*, *recall*, dan *f-measure*.

3. HASIL DAN PEMBAHASAN

A. Hasil Pengambilan dan Validasi Data

Pengambilan data dilakukan dengan rentang waktu mulai dari bulan Agustus 2021 sampai bulan Juni 2022. Total data yang di peroleh dari *Drone Emprit* adalah 35780 baris dan 15 kolom yang di simpan dalam format *xlsx*. Namun, data yang diperoleh masih memiliki kekurangan yakni, adanya beberapa data *Mentions* atau teks yang sama tetapi memiliki label yang berbeda karena pelabelan dilakukan secara otomatis oleh *drone emprit*. Oleh karena itu, sebelum dilanjutkan pada proses *preprocessing*, data tersebut harus divalidasi terlebih dahulu oleh seorang ahli bahasa. Validasi ini bertujuan untuk memastikan teks dan label data yang diperoleh dari *drone emprit* telah sesuai. Berdasarkan hasil validasi, jumlah data yang diperoleh adalah 300 dengan jumlah label positif 100, label netral 100 dan label negatif 100.

B. Hasil Preprocessing

Berikut ini merupakan hasil dari proses *preprocessing* pada data 2 label dan 3 label.

	CleanMentions	label
0	vaksin moderna salah jenis vaksin program vaks...	1
1	vaksin moderna september gedung sasana praja	1
2	takut vaksin moderna	0
3	vaksin pfizer parah efek sinovac	0
4	pfizer serah data uji coba vaksin anak anak	1
...
195	pantau program vaksinasi wiabup vaksin tingkat	1
196	vaksin hilang ayo semangat sehat	1
197	vaksin gamenjamin kebal corona standar	0
198	indonesia juta vaksin september gaspol vaksina...	1
199	hindar hoax putar vaksin indonesia	1

200 rows x 2 columns

Gambar 2 Hasil *Preprocessing*

	CleanMentions	label
0	vaksin moderna salah jenis vaksin program vaks...	2
1	vaksin moderna september gedung sasana praja	2
2	takut vaksin moderna	0
3	vaksin pfizer parah efek sinovac	0
4	pfizer serah data uji coba vaksin anak anak	2
...
295	lacak sebar mudah praktis unduh sertifikat vaksin	1
296	jenis vaksin indonesia izin darurat bpom	1
297	indonesia juta vaksin september gaspol vaksina...	2
298	hindar hoax putar vaksin indonesia	2
299	menteri uang muljani indrawati perintah seleng...	1

300 rows x 2 columns

Gambar 3 Hasil *Preprocessing*

C. Hasil Ekstraksi Fitur

Berikut ini merupakan hasil ekstraksi fitur untuk data 2 label dan 3 label.

F. Hasil Analisa Performa Model

Tabel 1. Perbandingan Nilai *Accuracy*

Dataset	Model	Model (Optimasi)	Nilai Accuracy	Selisih	Keterangan
Data 2 label	SVM	Sebelum	95%	3,3 %	Meningkat
		Setelah	98,3%		
Data 3 Label	SVM	Sebelum	67,7 %	5,6%	Meningkat
		Setelah	73,3%		

Tabel 2. Perbandingan Nilai *F1-measure*

Dataset	Model	Model (Optimasi)	Nilai F1-Measure	Selisih	Keterangan
Data 2 label	SVM	Sebelum	94,5%	3,7 %	Meningkat
		Setelah	98,2%		
Data 3 Label	SVM	Sebelum	67,8 %	6,1 %	Meningkat
		Setelah	73,9%		

Tabel 3. Perbandingan Nilai *Precision*

Dataset	Model	Model (Optimasi)	Nilai Precision	Selisih	Keterangan
Data 2 label	SVM	Sebelum	93,8%	4,3 %	Meningkat
		Setelah	98,1%		
Data 3 Label	SVM	Sebelum	70,4%	6,5%	Meningkat
		Setelah	76,9%		

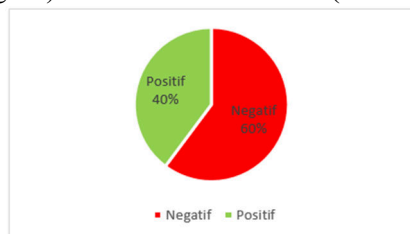
Tabel 4. Perbandingan Nilai *Recall*

Dataset	Model	Model (Optimasi)	Nilai Recall	Selisih	Keterangan
Data 2 label	SVM	Sebelum	96,1%	2,5 %	Meningkat
		Setelah	98,6%		
Data 3 Label	SVM	Sebelum	67,5 %	5,7%	Meningkat
		Setelah	73,2%		

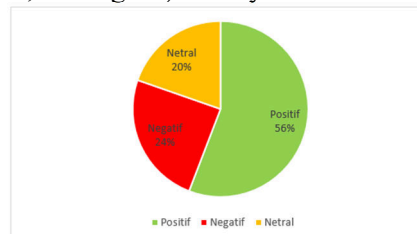
Berdasarkan hasil analisa diatas, performa model sebelum dan setelah optimasi parameter menghasilkan performa nilai *Accuracy* yang lebih baik setelah menerapkan metode optimasi parameter dengan persentase peningkatan *Accuracy* sebesar 3,3 % pada data 2 label sedangkan pada data 3 label meningkat sebesar 5,6 %. Begitu pula dengan performa nilai *F1-measure* dari model klasifikasi lebih baik setelah menerapkan metode optimasi parameter dengan persentase peningkatan *F1-measure* sebesar 3,7 % pada data 2 label sedangkan pada data 3 label persentase *f1-measure* meningkat sebesar 6,1 %. Dan untuk *Precision* dari model klasifikasi juga memiliki *precision* yang lebih baik setelah menerapkan metode optimasi parameter dengan persentase peningkatan *precision* sebesar 4,3 % pada data 2 label sedangkan pada data 3 label persentase *precision* meningkat sebesar 6,5 %. Kemudian nilai *recall* dari model klasifikasi memiliki nilai yang lebih baik setelah menerapkan metode optimasi parameter dengan persentase peningkatan *recall* sebesar 2,5 % pada data 2 label sedangkan pada data 3 label persentase *recall* meningkat sebesar 5,7 %. Angka tersebut membuat nilai performa yang baik sesuai dengan yang diharapkan pada tujuan penelitian.

G. Hasil Implementasi Model Hasil Optimasi

Berikut ini merupakan Implementasi model hasil optimasi menggunakan data baru sebanyak 3736 dengan 2 label (positif, negatif) dan untuk data 3 label (Positif,netral,dan negatif) sebanyak 4830 data.



Gambar 9 Diagram Pie Data 2 Label



Gambar 10 Diagram Pie Data 2 Label

Berdasarkan diagram pie diatas, menunjukkan bahwa hasil *Mentions* atau *tweet* Vaksinasi di Indonesia pada data 2 label sebanyak 40% *tweet* positif dan 60 % *tweet* negatif Sehingga dapat diketahui respon masyarakat lebih banyak beranggapan negatif terhadap kebijakan pemerintah tersebut. Sedangkan hasil *Mentions* atau *tweet* vaksinasi di Indonesia dengan menggunakan data 3 label sebanyak 56% *tweet* positif, 20% *tweet* Netral, dan 24% *tweet* negatif. Sehingga dapat diketahui respon masyarakat lebih banyak beranggapan positif terhadap kebijakan pemerintah tersebut. Kedua hasil prediksi diatas, dapat dijadikan sebagai tolok ukur atau bahan evaluasi bagi pemerintah dalam membuat kebijakan baru berdasarkan tanggapan masyarakat pada media social (twitter).

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode optimasi parameter pada SVM dengan menggunakan algoritma PSO dan terbukti mampu meningkatkan performa *accuracy*, *precision*, *recall*, dan *f-measure* baik menggunakan data dua label (Positif, Negatif) maupun data tiga label (Positif, Netral, Negatif). Pada data 2 label nilai *fitness* atau *error* yang paling rendah yang di dapatkan pada setiap parameter adalah 0.166, sedangkan untuk data 3 label di dapatkan nilai *fitness* atau *error* yang paling rendah adalah 0.0266. Hasil validasi untuk masing-masing data 2 label dan 3 label memiliki performa paling tinggi pada nilai *k-fold* 3.

5. UCAPAN TERIMA KASIH

Ucapan terima kasih penulis kepada orang tua, teman penulis, kedua dosen pembimbing, serta seluruh dosen dan staff terkhusus di Teknik Elektro Program Studi Teknik Komputer dan Jaringan Politeknik Negeri Ujung Pandang.

6. DAFTAR PUSTAKA

- [1] H. A. F. Frizka Fitriana, Ema Utami, “Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes,” *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [2] R. Yasmin, “Covid-19 Menggunakan Metode Naive Bayes Classifier Pada Media Sosial Twitter Covid-19 Menggunakan Metode Naive Bayes,” 2021.
- [3] H. Tuhuteru, “Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine,” *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.
- [4] E. Nufa, “Analisis Klasifikasi Sentimen Tentang Pro Dan Kontra Masyarakat Indonesia Terhadap Vaksin Covid-19 Pada Media,” no. May, p. 2, 2021.
- [5] R. Rosdiana, T. Eddy, S. Zawiyah, and N. Y. U. Muhammad, “Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar,” *Proceeding SNTEI*, no. June 2020, pp. 87–93, 2019.
- [6] L. Q. Zalyhaty, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” vol. 7, p. 6, 2021.
- [7] E. Tungadi, I. Thalib, M. Nur, and Y. Utomo, “Machine Learning Penentuan Penerima Beasiswa Peningkatan Prestasi Akademik (PPA) Menggunakan Metode Jaringan Saraf Tiruan (JST),” no. September, pp. 391–396, 2018.
- [8] I. Syarif, A. Prugel-Bennett, and G. Wills, “SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 4, p. 1502, 2016, doi: 10.12928/telkommika.v14i4.3956.
- [9] Y. Alkhalifi, W. Gata, A. Prasetyo, and I. Budiawan, “Analisis Sentimen Penghapusan Ujian Nasional pada Twitter Menggunakan Support Vector Machine dan Naive Bayes berbasis Particle Swarm Optimization,” *CoreIT*, vol. 6, no. 2, pp. 71–78, 2020.