

Optimasi Parameter *Support Vector Machine* menggunakan Algoritma Genetika untuk Deteksi *Cyberbullying*

Parameter Optimisation of Support Vector Machine using Genetic Algorithm for Cyberbullying Detection

¹Mohd Qorib Alqowiy*, ²Ema Utami

^{1,2}Informatika Program Magister, Universitas Amikom Yogyakarta

^{1,2}Jl Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa
Yogyakarta 55281, Indonesia

*e-mail: qoribalqowiy@students.amikom.ac.id

(received: 1 December 2024, revised: 6 December 2024, accepted: 8 January 2025)

Abstrak

Perkembangan teknologi yang sangat pesat memudahkan beberapa aktifitas sehari-hari salah satunya dalam berkomunikasi. Terdapat beberapa platform media sosial yang memudahkan penggunaannya untuk saling terhubung antara satu dengan yang lain. Namun demikian, tidak jarang perkembangan ini berdampak negatif. *Cyberbullying* adalah salah satu dampak negatif yang dapat terjadi. Sehingga untuk mengatasi masalah tersebut banyak peneliti yang menawarkan solusi terkait dengan deteksi *cyberbullying*. Beberapa solusi yang ditawarkan umumnya menggunakan metode *Support Vector Machine* (SVM) karena memberikan hasil yang lebih akurat dibandingkan algoritma yang lain. Namun, metode ini masih sulit untuk menentukan nilai dari parameter kernel terbaik. Oleh karena itu, beberapa algoritma pencarian disarankan untuk mencari nilai parameter SVM yang optimal. Penelitian ini menggunakan algoritma genetika untuk menemukan nilai terbaik dari nilai C dan gamma. Hasil penelitian menunjukkan peningkatan akurasi ketika algoritma genetika digunakan dengan nilai akurasi mencapai 86%. Hal ini menunjukkan efektifitas algoritma genetika untuk mengoptimasi parameter pada SVM.

Kata kunci: *cyberbullying*, *twitter*, text mining, SVM, algoritma genetika

Abstract

The rapid development of technology has greatly facilitated various daily activities, including communication. Several social media platforms enable users to connect with one another easily. However, this progress often comes with negative impacts. One such impact is cyberbullying, which has become a growing concern. To address this issue, many researchers have proposed solutions for detecting cyberbullying. Among these, the Support Vector Machine (SVM) method is commonly used because it delivers more accurate results compared to other algorithms. However, determining the optimal kernel parameters for SVM remains a challenge. To overcome this, various search algorithms have been suggested to optimize SVM parameters. This study utilizes a genetic algorithm to find the optimal values for the C and gamma parameters. The results demonstrate an accuracy improvement, with the genetic algorithm achieving an accuracy of 86%. This highlights the effectiveness of genetic algorithms in optimizing SVM parameters for cyberbullying detection.

Keywords: *cyberbullying*, *twitter*, text mining, SVM, genetic algorithm

1 Pendahuluan

Kemajuan teknologi pada saat ini mempermudah berbagai macam aktifitas sehari-hari termasuk dalam hal berkomunikasi [1]. Media sosial sebagai sebuah platform komunikasi menjadi bukti dari kemajuan teknologi yang memudahkan siapapun untuk saling terhubung dan bertukar informasi. Twitter, salah satu platform media sosial populer, sering digunakan untuk berkomunikasi atau berkomentar diantara pengguna. Twitter merupakan salah satu platform media sosial yang dapat membagikan pesan teks maupun foto bagi penggunanya.

Seiring meningkatnya pengguna pada media sosial *twitter*, penyalahgunaanya juga tidak dapat dihindari. Salah satu bentuk penyalahgunaan pada media sosial *twitter* adalah *cyberbullying*.

<http://sistemasi.ftik.unisi.ac.id>

Cyberbullying adalah penggunaan teknologi digital untuk menyakiti atau menindas berulang kali, dan *cyberbullying* akan berdampak lebih dalam dan lebih lama dibandingkan dengan intimidasi fisik [2]. *Cyberbullying* telah menjadi isu serius di dunia maya dan menyebabkan dampak psikologis yang mendalam seperti depresi, kecemasan dan bahkan yang paling tragis, bunuh diri [3].

Meningkatnya kasus *cyberbullying* membuat penelitian tentang deteksi *cyberbullying* mendapatkan minat yang tinggi beberapa tahun ini [4]. Salah satu cara yang dapat digunakan untuk mendeteksi sebuah teks atau kalimat mengandung *bullying* atau tidak adalah dengan pendekatan *machine learning*. Penggunaan *machine learning* dapat membantu mendeteksi sebuah teks atau kalimat terindikasi *cyberbullying* atau tidak dengan menerapkan metode yang tepat.

Beberapa metode telah digunakan untuk mendeteksi *cyberbullying*. Penelitian oleh Dalvi, dkk [5] menggunakan *Naïve Bayes* (NB) dan *Support Vector Machines*(SVM) sebagai metode untuk deteksi *cyberbullying*. Pada penelitian ini dataset yang digunakan berasal dari twitter yang kemudian dilakukan *pre-processing*. Hasil penelitian ini menunjukkan bahwa SVM mendapatkan nilai akurasi 71.25 % lebih tinggi dibandingkan dengan NB yang hanya 52.70%. Meskipun SVM mendapatkan nilai akurasi yang lebih tinggi dibandingkan NB, namun hasil yang didapat kurang maksimal. Hal ini bisa terjadi karena kurang tepatnya pemilihan parameter yang dilakukan.

Klasifikasi dilakukan oleh SVM dengan menemukan hyperplane atau garis pemisah terbaik untuk mengkategorikan data ke dalam beberapa kelas. Namun, penggunaan SVM pada data yang tidak linear dan bahkan lebih dari dua kategori menimbulkan masalah. Dalam mengklasifikasikan data non-linier, SVM memiliki metode yang disebut kernel atau karnel trick. Adapun jenis kernel pada SVM yaitu *linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *sigmoid*. Kernel RBF merupakan kernel yang banyak dipilih oleh para peneliti [6]. Salah satu kerumitan dalam SVM yaitu dalam menentukan nilai parameter kernel yang paling optimal [5]. Oleh karena itu, untuk menemukan nilai parameter SVM yang optimal, disarankan dengan menggunakan algoritma algoritma pencarian dengan harapan dapat mengoptimasikan SVM.

Penelitian sebelumnya telah mengusulkan beberapa metode untuk mengoptimasikan SVM, penelitian oleh Al-Khowarizmi, dkk [7] melakukan penelitian untuk mengoptimasi SVM dengan menggunakan fungsi kernel kubik. Penelitian ini mendapatkan nilai akurasi pada SVM menggunakan fungsi kernel linear sebesar 92,3% sedangkan menggunakan fungsi linear kubik sebesar 90%.

Menemukan nilai parameter pada SVM dapat menjadi pilihan untuk mengoptimasikan SVM, algoritma pencarian bisa digunakan untuk mendapatkan nilai parameter yang lebih baik. Pada penelitian Ramasamy, dkk [8] membandingkan tiga argoritma heuristics *Cuckoo Search Optimization* (CSO), *Ant Lion Optimizer* (ALO), dan *Polar Bear Optimization* (PBO) sebagai algoritma pencarian untuk mengoptimasi parameter SVM. Selain itu, algoritma genetika (GA) digunakan dalam mengoptimasikan klasifikasi SVM pada Review Go Jek's Services oleh Windha & Haryoko [9], penelitian ini mendapatkan hasil akurasi yang lebih optimal dengan menggunakan algoritma genetika. Algoritma genetika efektif dalam menangani ruang pencarian yang luas secara efisien dan mengurangi risiko konvergensi pada solusi optimal secara local sehingga dapat menghasilkan solusi yang diharapkan secara optimal.

Penelitian ini akan mengoptimasi parameter SVM dengan menerapkan metode algoritma genetika untuk deteksi *cyberbullying* pada media sosial twitter dengan menguji sejauh mana pengaruh penggunaan kombinasi antara algoritma genetika dengan algoritma SVM terhadap kinerja algoritma SVM dalam deteksi *cyberbullying*.

2 Tinjauan Literatur

Analisis sentimen menggunakan pemrosesan bahasa alami (NLP), analisa teks, dan metode komputasi untuk mengekstrak atau mengklasifikasikan sentimen secara otomatis dari ulasan sentimen. Analisis sentimen dan opini ini telah berkembang ke beragam sektor, seperti sosial, ulasan konsumen, dan sebagainya[2]. Deteksi *Cyberbullying* adalah salah satu bidang yang sering diselesaikan menggunakan analisis sentimen. SVM adalah metode yang populer dalam pembelajaran mesin untuk menyelesaikan masalah klasifikasi dan analisis sentimen. Tujuan utamanya adalah memisahkan dua kelas data menggunakan hyperplane dengan jarak maksimum dari titik data yang paling dekat dengan garis pemisah[10].

Penelitian yang dilakukan oleh Prabowo & Azizah [11] menggunakan SVM sebagai metode klasifikasi pada sentiment analisis untuk deteksi cyberbullying. Pada penelitian ini dataset yang dipakai merupakan komentar dari Instagram dengan jumlah 1500 data. *Features term frequency-inverse document frequency* (TF-IDF) dikombinasikan dengan metode SVM untuk mendapatkan hasil akurasi yang maksimal. Hasil akhir pada penelitian ini dengan model yang ditawarkan mendapatkan nilai akurasi sebesar 93%.

Metode SVM juga dinilai lebih baik oleh peneliti sebelumnya dalam mengatasi masalah *cyberbullying*. Penelitian yang dilakukan oleh Atoum [12] telah membandingkan SVM dengan *Naïve Bayes* (NB). Dataset yang digunakan berasal dari twitter dengan jumlah 5628 tweets. Beberapa percobaan dilakukan dengan menggunakan 2-gram, 3-gram dan 4-gram untuk mengevaluasi pengklasifikasi NB dan Pengklasifikasi NB dan SVM dalam hal akurasi, presisi, recall, Fmeasure, dan ROC. SVM berhasil mencapai hasil rata-rata yang lebih tinggi daripada pengklasifikasi NB di semua n-gram dengan akurasi 92.02%.

Penelitian tentang cyberbullying di twitter juga telah dilakukan pada sentimen bahasa arab [13]. Penelitian ini menggunakan dataset yang dikumpulkan dari twitter dengan dilabeli sebagai *cyberbullying* dan *non-cyberbullying*. SVM dan NB dibandingkan dengan membuat beberapa scenario yang mendapatkan nilai akurasi masing-masing yaitu 95.742% untuk SVM, sedangkan NB mendapatkan nilai akurasi 70.942%. Hal ini dapat membuktikan SVM lebih unggul daripada NB dalam sentiment analisis pada bahasa arab.

Penelitian menggunakan SVM juga telah dilakukan oleh peneliti sebelumnya untuk mendeteksi dan mencegah cyberbullying di Twitter. Penelitian yang dilakukan oleh [10] terkait dengan analisis sentimen untuk mendeteksi *cyberbullying* menggunakan dataset yang diambil dari twitter dengan membagi data kedalam data training dan data testing. Pada penelitian ini, SVM digunakan sebagai metode klasifikasi dengan hasil akurasi mencapai 92,7%. Penelitian ini mengusulkan penelitian lebih lanjut dengan memodifikasi SVM menggunakan kernel lain agar mendapatkan nilai akurasi yang optimal.

Penelitian lainnya juga dilakukan oleh Perera & Fernando [14] menggunakan metode Support Vector Machine (SVM). Penelitian ini memanfaatkan teknik Natural Language Processing (NLP) seperti analisis sentimen, TF-IDF, dan deteksi kata-kata kasar untuk mengidentifikasi teks yang termasuk dalam kategori cyberbullying atau *non-cyberbullying*. Dataset yang digunakan berasal dari Twitter dan melalui proses pelabelan manual setelah dilakukan preprocessing Dataset yang digunakan dibagi menjadi data pelatihan (70%) dan data pengujian (30%), dengan SVM menunjukkan akurasi tertinggi sebesar 75,5% pada dataset yang berisi 500 teks, mengungguli metode seperti *Naïve Bayes* dan *Logistic Regression*. penelitian ini merekomendasikan pengembangan lebih lanjut melalui peningkatan ukuran dataset, serta eksplorasi metode klasifikasi yang lebih kompleks untuk meningkatkan akurasi dan cakupan deteksi.

Penggunaan feature selection pada penelitian tentang deteksi *cyberbullying* telah dilakukan oleh Purnamasari, dkk [15]. Penelitian ini menghasilkan hasil yang lebih baik dengan menggunakan metode SVM dan pemilihan fitur IG. Penggunaan fitur IG menghasilkan nilai akurasi 76.66%, ketepatan 72.22%, *recall* 86.66%, dan *f-measure* 78.78%. Untuk mendapatkan hasil yang lebih baik, penelitian harus dilanjutkan dengan menggunakan atau menggabungkan beberapa pemilihan fitur yang ada untuk mendapatkan hasil yang lebih baik.

Penelitian tentang cyberbullying juga telah diteliti oleh Wijayanti, dkk [16]. Penelitian ini melakukan crawl data pada twitter untuk mendapatkan dataset dengan menggunakan kata kunci terkait dengan cyberbullying. Metode SVM dengan 4 karnel yang dimiliki yaitu Linear, RBF, polynominal dan sigmoid digunakan untuk mengetahui karnel mana yang memiliki akurasi terbaik dalam menganalisa *cyberbullying*. Karnel sigmoid mendapatkan hasil nilai akurasi yang paling tinggi dengan nilai akurasi 83.85%. Untuk menentukan kernel yang terbaik perlu dilakukan penelitian lebih lanjut dengan mencoba beberapa perbandingan penggunaan data training dan data testing sehingga mendapatkan hasil yang lebih beragam.

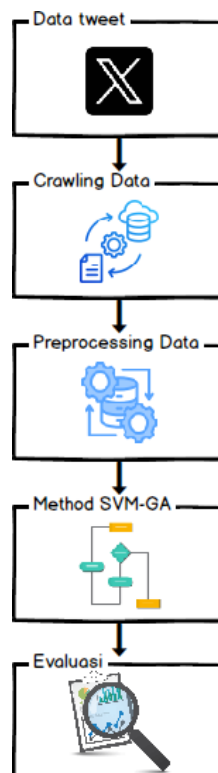
Penelitian oleh Windha & Haryoko [9] mencoba untuk mengoptimalkan algoritma SVM dengan mengatasi tantangan pemilihan parameter yang relevan dalam metode SVM menggunakan algoritma genetika (GA). Algoritma genetika membantu mengoptimalkan parameter-parameter yang ada, sehingga memungkinkan SVM untuk menjalankan klasifikasi dengan lebih efisien. hasil penelitian

menunjukkan bahwa klasifikasi menggunakan SVM-GA memiliki akurasi tertinggi (0.895), sedangkan akurasi klasifikasi menggunakan SVM hanya sebesar 0.621.

Berdasarkan beberapa tinjauan pustaka tersebut peneliti menemukan beberapa kekurangan dan saran untuk penelitian yang akan dilakukan. Pada penelitian ini akan dilakukan deteksi *cyberbullying* dengan melakukan optimasi pada metode SVM menggunakan metode optimasi yaitu Genetik algoritma untuk mencari parameter yang terbaik pada kernel yang digunakan.

3 Metode Penelitian

Metode penelitian yang digunakan oleh peneliti untuk memecahkan masalah terdiri dari beberapa langkah. Langkah-langkah ini meliputi pengumpulan data, praproses data, pemodelan menggunakan metode SVM dan SVM-GA, pengujian dan evaluasi. Langkah-langkah penelitian dapat dilihat pada Gambar 1 berikut:



Gambar 1. Metode penelitian

Pada Gambar 1 dijelaskan bahwa alur penelitian pada penelitian ini diawali dengan pengumpulan data, dimana dataset dikumpulkan dari data tweets yang diambil dari Twitter menggunakan teknik crawling data dengan Google Collaboratory platform dan python sebagai bahasa yang digunakan dengan kata kunci yang terkait *cyberbullying*. Setelah data tweets berhasil dikumpulkan, tahap selanjutnya adalah melakukan preprocessing pada data tweets yang sudah dikumpulkan untuk mendapatkan dataset dengan kualitas yang baik sebelum digunakan pada model yang akan dibangun.

Tahap preprocessing ini akan melewati beberapa tahapan diantaranya:

1. Handle Duplicate

Handle duplicates merupakan tahap yang dilakukan oleh peneliti untuk menghapus baris ganda atau duplikat dari data yang ada di dalam dataset. Berikut contoh *handle duplicate* dapat dilihat pada Tabel 1.

Tabel 1. Contoh hasil handle duplicates

Tweet	Handle Duplicates
-------	-------------------

Kamu sangat bodoh dan jelek!	Kamu sangat bodoh dan jelek!
Mel disini kamu keliatan cakep ☺	Mel disini kamu keliatan cakep ☺
Kamu sangat bodoh dan jelek!	

2. Cleaning

Cleaning merupakan proses menghilangkan karakter yang tidak dibutuhkan pada saat proses mining. Berikut contoh *cleaning* dapat dilihat pada Tabel 2.

Tabel 2. Contoh hasil cleaning

Handle Duplicates	Cleaning
Kamu sangat bodoh dan jelek!	Kamu sangat bodoh dan jelek
Mel disini kamu keliatan cakep ☺	Mel disini kamu keliatan cakep

3. Lower Case

Teks dari data mentah yang didapat dari twitter mengandung huruf besar dan huruf kecil. Tahapan ini akan mengubah huruf menjadi huruf kecil. Berikut contoh *lower case* dapat dilihat pada Tabel 3.

Tabel 3. Contoh hasil lower case

Cleaning	Lower Case
Kamu sangat bodoh dan jelek	kamu sangat bodoh dan jelek
Mel disini kamu keliatan cakep	mel disini kamu keliatan cakep

4. Tokenizing

Pada tahapan ini dilakukan proses pemotongan atau pemisahan dokumen menjadi bagian bagian kata atau token. Berikut contoh *tokenizing* dapat dilihat pada Tabel 4.

Tabel 4. Contoh hasil tokenizing

Lower Case	Tokenizing
kamu sangat bodoh dan jelek	['kamu', 'sangat', 'bodoh', 'dan', 'jelek']
mel disini kamu keliatan cakep	['mel', 'disini', 'kamu', 'keliatan', 'cakep']

5. Stopwords Removal

Stopwords removal digunakan untuk mengambil atau memisahkan kata-kata yang tidak terlalu penting atau yang akan digunakan. Berikut contoh *stopwords removal* dapat dilihat pada Tabel 5.

Tabel 5. Contoh hasil stopwords removal

Tweet	Handle Duplicates
kamu sangat bodoh dan jelek	['kamu', 'sangat', 'bodoh', 'jelek']
mel disini kamu keliatan cakep	['mel', 'disini', 'kamu', 'keliatan', 'cakep']

6. Stemming

Tahapan ini dilakukan untuk mengubah teks menjadi kata kerja infinitif tanpa imbuhan apapun sehingga indek data menjadi kecil. Berikut contoh *stemming* dapat dilihat pada Tabel 6.

Tabel 6. Contoh hasil steaming

Tweet	Handle Duplicates
['kamu', 'sangat', 'bodoh', 'jelek']	Kamu sangat bodoh jelek
['mel', 'disini', 'kamu', 'keliatan', 'cakep']	Mel sini kamu liat cakep

Setelah langkah pre-processing selesai dilakukan, langkah selanjutnya adalah melakukan split atau membagi data kedalam dua bagian yaitu data training dan data testing, dimana 80% data digunakan untuk data training dan 20 % data digunakan untuk data testing. Data yang telah dibagi selanjutnya akan masuk kedalam tahap ekstraksi fitur, dimana metode yang digunakan adalah metode TF-IDF.

Tahap selanjutnya adalah melakukan optimasi parameter pada SVM, algoritma genetika akan digunakan pada penelitian ini. Dimana algoritma ini akan menentukan beberapa parameter penting yang ada pada metode SVM seperti nilai C dan Gamma.

Tahap berikutnya setelah melakukan pengoptimasian adalah melihat seberapa berpengaruh algoritma genetika dalam pengotimasian parameter SVM. Metode SVM akan dibandingkan dengan metode SVM-GA untuk mengetahui tingkat akurasi, presisi, recall dan f1-score dari setiap metode, sehingga akan didapat kesimpulan apakah metode SVM-GA dapat meningkatkan performa dari SVM atau tidak dengan melihat hasil yang didapat dari evaluasi model.

4 Hasil dan Pembahasan

Pada bagian ini, peneliti mencoba sebuah pemodelan system untuk mengoptimasi parameter dari metode SVM dengan memanfaatkan algoritma genetika untuk mengklasifikasikan tweet yang mengandung unsur cyberbullying dan noncyberbullying.

4.1 Dataset

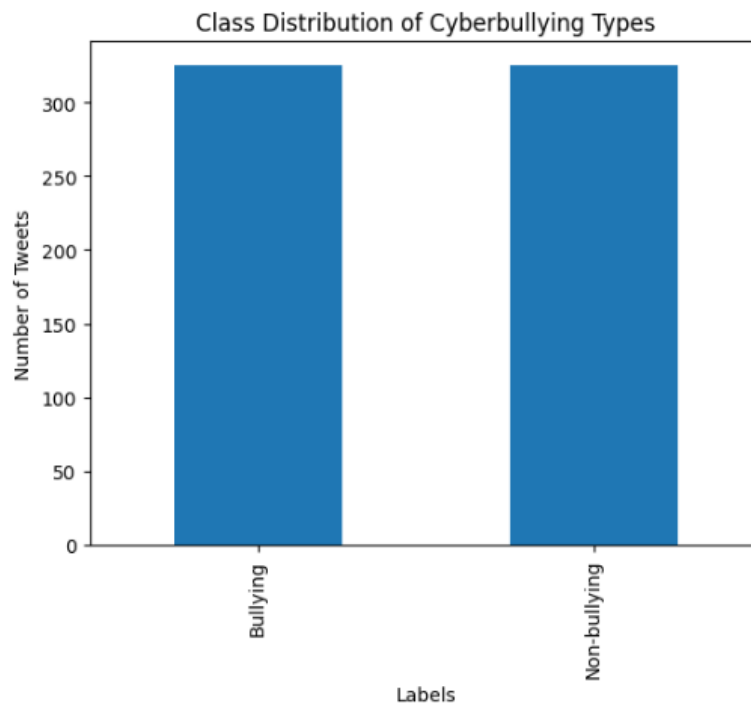
Dataset yang digunakan berasal dari media social twitter yang berisi tweets terkait dengan cyberbullying. Pada Tabel 7 akan disajikan contoh dataset yang digunakan. Dataset yang digunakan adalah data tweets yang menggunakan bahasa Indonesia dengan jumlah data 650.

Tabel 7. Contoh dataset

No	Tweets
1	Apaan sih gak jelas hidupnya
2	Jijik astagah sok bat cakep njs
3	Kaya tulus ya mukanya klo diliat2
4	Jadi diri sendiri, jangan jadi ular apalagi dugong

Setelah mendapatkan dataset yang berisi tweets terkait dengan cyberbullying, dilakukan pelabelan pada dataset secara manual. Dataset yang terdiri dari 650 tweet dilabeli menjadi 2 kategori yaitu cyberbullying dan non-cyberbullying. Pelabelan dilakukan dengan membaca dan memahami setiap tweet secara mendalam oleh tim peneliti yang memahami konteks dan budaya bahasa Indonesia. Untuk mengurangi bias, tim peneliti menetapkan kriteria yang jelas tentang apa yang dianggap sebagai cyberbullying, seperti tweet yang mengandung hinaan, ancaman, atau serangan verbal. Kriteria ini didiskusikan dan disepakati di awal agar proses pelabelan konsisten.

Setiap kategori memiliki jumlah data yang sama yaitu 325 data cyberbullying dan 325 data non-cyberbullying atau dengan persentase perbandingan 50% banding 50%. Hasil visualisasi pembagian data ditunjukkan pada Gambar 2.



Gambar 2. Hasil visualisasi pembagian dataset

Dataset yang seimbang dengan perbandingan 50% banding 50% antara kategori cyberbullying dan non-cyberbullying dapat mempengaruhi performa model. Keberimbangan ini terbukti penting dalam mencegah bias klasifikasi, seperti yang diungkapkan oleh Prabowo & Azizah [11], di mana dataset yang terdistribusi secara merata membantu meningkatkan keandalan dan stabilitas model pada berbagai metrik evaluasi. Setelah pelabelan selesai, sebagian data diperiksa ulang sebagai validasi untuk memastikan tidak ada kesalahan atau bias sistematis. Hal ini memastikan dataset yang dihasilkan memiliki kualitas tinggi, sehingga mendukung performa model yang dibangun.

4.2 Preprocessing

Pada tahap preprocessing, tahap pertama yang dilakukan adalah pemeriksaan duplikasi dataset. Hasil dari crawling data pada twitter dipastikan tidak terjadi duplikasi data untuk menghasilkan dataset yang lebih baik. Setelah diperiksa, tidak ada data yang sama sehingga dataset yang dipakai tetap 650 tweets. Tahap selanjutnya dilakukan cleaning data agar karakter yang tidak penting dalam dataset dibersihkan atau dihilangkan.

Setelah data dibersihkan, proses selanjutnya adalah membuat semua huruf yang ada pada dataset menjadi huruf kecil atau yang disebut dengan proses lower case. Ketika data sudah bersih, tahap selanjutnya memotong kalimat menjadi kata-kata. Tahapan ini disebut dengan tokenisasi. Kemudian potongan kalimat (kata) yang terbentuk, diproses untuk menentukan kata mana yang dianggap dianggap berguna atau penting dengan menggunakan stop word removal.

Langkah terakhir pada tahap preprocessing adalah stemming yang bertujuan untuk membentuk kata dasar. Pada tahap ini tahap ini, peneliti menggunakan pustaka literatur yang telah dipasang sebelumnya. Setelah semua tahap preprocessing dilalui, seluruh dataset tersebut akan langsung tersimpan dengan ekstensi .csv yang kemudian akan digunakan sebagai dataset pada proses pemodelan SVM dan SVM-GA.

4.3 Metode

Metode yang digunakan untuk mengoptimasi parameter SVM adalah algoritma genetika. Dataset yang telah dilakukan preprocessing sebelumnya dipecah menjadi dua bagian yang kemudian digunakan sebagai data training dan data testing. Data training yang digunakan sebesar 80% dan data

testing sebesar 20%. Namun sebelum masuk ke tahap pemodelan, peneliti melakukan TF-IDF untuk menghitung bobot kata. Menurut penelitian oleh Ramasamy, dkk[8], kombinasi TF-IDF dengan algoritma SVM dapat meningkatkan kemampuan model dalam menangani data berbasis teks.

Untuk melakukan pembobotan, peneliti memasang library sklearn. Setelah proses TF-IDF, dilakukan skenario pertama menggunakan SVM default dengan kernel RBF untuk mengetahui nilai akurasi yang dihasilkan ketika menggunakan model SVM tanpa GA. Selanjutnya dilakukan skenario kedua menggunakan SVM-GA dengan Kernel RBF. Dimana parameter yang akan digunakan oleh GA yaitu mencari nilai terbaik dari C dan Gamma.

4.4 Testing dan Evaluasi

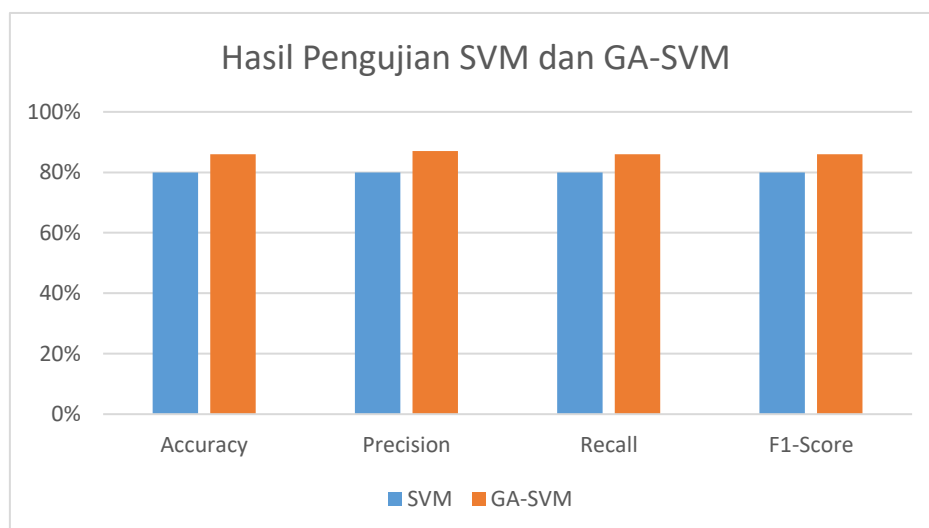
Penelitian ini menguji model yang dibangun yaitu model SVM dan SVM-GA. Pengujian yang dilakukan mendapatkan nilai akurasi, presisi, recall, dan nilai f1-score pada masing-masing model yang diuji. Berikut perbandingan hasil pengujian antara SVM dan SVM-GA.

Tabel 8. Hasil pengujian SVM dan SVM-GA

Model	Akurasi	Presisi	Recall	F1-Score
SVM	80%	80%	80%	80%
SVM-GA	86%	87%	86%	86%

Berdasarkan hasil uji coba pada Tabel 8 SVM-GA menghasilkan nilai akurasi lebih tinggi dari pada SVM. Nilai akurasi yang dihasilkan oleh SVM-GA adalah 87% sedangkan nilai akurasi SVM 80%, hal ini menunjukkan adanya peningkatan pada model SVM yang menggunakan algoritma genetika sebesar 6% dari yang awalnya 80% menjadi 86% setelah menggunakan algoritma genetika untuk mencari nilai terbaik dari nilai c dan nilai gamma. Sementara itu, jika dilihat dari nilai precision, recall dan f1-Score. SVM-GA juga menunjukkan peningkatan dimana model SVM-GA menghasilkan nilai 87% untuk precision dan 86% untuk accuracy dan recall. Sementara itu model SVM menghasilkan nilai 80% untuk precision, recall dan f1-Score.

Peningkatan ini dapat terjadi karena kemampuan algoritma genetika dalam mengoptimalkan parameter *C* dan *gamma* pada kernel RBF, yang berperan penting dalam menentukan performa SVM. Penelitian oleh Harafani [6] mencatat bahwa algoritma genetika dapat mengeksplorasi ruang parameter yang besar dan kompleks secara efisien, sehingga menghasilkan solusi optimal yang lebih baik dibandingkan metode tradisional. Selain itu, penelitian Windha Mega dan Haryoko [9] juga mencatat bahwa optimasi parameter SVM menggunakan algoritma genetika menghasilkan akurasi tertinggi hingga 89,5% pada studi kasus lain, yang menunjukkan konsistensi manfaat pendekatan ini dalam berbagai konteks. Perbandingan hasil pengujian SVM dan SVM-GA dapat dilihat dalam bentuk grafik pada Gambar 3.



Gambar 3. Hasil perbandingan SVM dan SVM-GA

Gambar 3 menunjukkan perbandingan hasil evaluasi antara model SVM dan SVM-GA dalam bentuk grafik batang yang mengilustrasikan nilai akurasi, presisi, recall, dan F1-Score pada kedua model. Grafik ini menggambarkan bahwa SVM-GA memiliki performa yang lebih unggul dibandingkan SVM standar dalam semua metrik evaluasi. Grafik ini memberikan visualisasi terhadap peningkatan model oleh algoritma genetika dalam mengoptimasi parameter SVM, yang secara signifikan meningkatkan kinerja model pada dataset yang digunakan.

5 Kesimpulan

Hasil pengujian yang dilakukan peneliti dengan membuat dua scenario pengujian yaitu menggunakan model SVM dan model SVM-GA menunjukkan peningkatan akurasi pada model SVM ketika menggunakan algoritma genetika untuk mencari nilai terbaik dari nilai c dan γ . SVM-GA menunjukkan peningkatan akurasi sebesar 6%. Dengan demikian dapat disimpulkan penentuan nilai c dan γ dapat mengoptimasi model SVM menjadi lebih baik. Untuk penelitian selanjutnya disarankan dapat lebih mengeksplor dengan menggunakan algoritma optimasi yang lain dan menggunakan dataset yang lebih besar.

Referensi

- [1] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying using Deep Neural Network," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 604–607, 2019, doi: 10.1109/ICACCS.2019.8728378.
- [2] M. Xin, J. Shen, and P. Hao, "Cyberbullying Detection and Classification with Improved IG and BiLSTM," *Proc. - 2022 Int. Conf. Electron. Devices, Comput. Sci. ICEDCS 2022*, pp. 259–262, 2022, doi: 10.1109/ICEDCS57360.2022.00065.
- [3] G. S. O’Keeffe et al., "Clinical Report - The Impact of Social Media on Children, Adolescents, and Families," *Pediatrics*, vol. 127, no. 4, pp. 800–804, 2011, doi: 10.1542/peds.2011-0054.
- [4] T. Agrawal and V. D. Chakravarthy, "Cyberbullying Detection and Hate Speech Identification using Machine Learning Techniques," *Proc. - 2022 2nd Int. Conf. Interdiscip. Cyber Phys. Syst. ICPS 2022*, pp. 182–187, 2022, doi: 10.1109/ICPS55917.2022.00041.
- [5] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, "Detecting a Twitter Cyberbullying using Machine Learning," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iccics, pp. 297–301, 2020, doi: 10.1109/ICICCS48265.2020.9120893.
- [6] H. Harafani, "Support Vector Machine Parameter Optimization to Improve Liver Disease Estimation with Genetic Algorithm," *Sinkron*, vol. 4, no. 2, p. 106, 2020, doi: 10.33395/sinkron.v4i2.10524.
- [7] Al-Khowarizmi, I. P. Sari, and H. Maulana, "Optimization of Support Vector Machine with Cubic Kernel Function to Detect Cyberbullying in Social Networks," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 22, no. 2, pp. 329–339, 2024, doi: 10.12928/TELKOMNIKA.v22i2.25437.
- [8] L. K. Ramasamy, S. Kadry, and S. Lim, "Selection of Optimal Hyper-Parameter Values of Support Vector Machine for Sentiment Analysis Tasks using Nature-Inspired Optimization Methods," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 290–298, 2021, doi: 10.11591/eei.v10i1.2098.
- [9] P. D. Windha Mega and Haryoko, "Optimization of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek’s Services," *2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019*, vol. 6, pp. 301–304, 2019, doi: 10.1109/ICITISEE48480.2019.9003894.
- [10] Al-Khowarizmi, I. P. Sari, and H. Maulana, "Detecting Cyberbullying on Social Media using Support Vector Machine: A Case Study on Twitter," *Int. J. Saf. Secur. Eng.*, vol. 13, no. 4, pp. 709–714, 2023, doi: 10.18280/ijssse.130413.
- [11] W. A. Prabowo and F. Azizah, "Sentiment Analysis for Detecting Cyberbullying using TF-IDF and SVM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 11–12, 2020, doi: 10.29207/resti.v4i6.2753.
- [12] J. O. Atoum, "Cyberbullying Detection Through Sentiment Analysis," *Proc. - 2020 Int. Conf. Comput. Sci. Comput. Intell. CSCCI 2020*, pp. 292–297, 2020, doi: 10.1109/CSCCI48480.2020.9003894.

- 10.1109/CSCI51800.2020.00056.
- [13] A. M. Alduailaj and A. Belghith, “Detecting Arabic Cyberbullying Tweets using Machine Learning,” *Mach. Learn. Knowl. Extr.*, vol. 5, no. 1, pp. 29–42, 2023, doi: 10.3390/make5010003.
- [14] A. Perera and P. Fernando, “Cyberbullying Detection System on Social Media using Supervised Machine Learning,” *Procedia Comput. Sci.*, vol. 239, no. 2023, pp. 506–516, 2024, doi: 10.1016/j.procs.2024.06.200.
- [15] N. M. G. D. Purnamasari, M. Ali Fauzi, Indriati, and L. S. Dewi, “Cyberbullying Identification in Twitter using Support Vector Machine and Information Gain based Feature Selection,” *Indonesia. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 3, pp. 1494–1500, 2020, doi: 10.11591/ijeecs.v18.i3.pp1494-1500.
- [16] S. S. Wijayanti, E. Utami, and A. Yaqin, “Comparison of Kernels on Support Vector Machine (SVM) Methods for Analysis of Cyberbullying,” 2022.