Deteksi Konten Negatif pada Media Sosial *Twitter* dengan menggunakan Metode Support Vector Machine

[1]Andrian Rizky Restuyadi, [2]Irvan Nugraha Marzuki, [3]Rifqi Adli Septian
[1,2,3] Teknik Informatika, Universitas Sebelas April
[1,2,3] Jl. Angkrek Situ No. 19, Sumedang Utara, Sumedang, Jawa Barat, Indonesia
[1]a2.2000008@mhs.stmik-sumedang.ac.id, [2]a2.2000054@mhs.stmik-sumedang.ac.id,
[3]a2.2000093@mhs.stmik-sumedang.ac.id

ABSTRACT
Media sosial *Twitter* merupakan salah satu media yang banyak digunakan oleh masyarakat Indonesia untuk berbagi informasi dan opini. Namun, media sosial ini juga rentan terhadap penyebaran konten yang merugikan, seperti konten rasis, konten yang mengandung ujaran kebencian (sara), dan konten pornografi. Konten-konten ini dapat memberikan dampak negatif pada pengguna dan masyarakat. Oleh karena itu, diperlukan sebuah sistem yang dapat mendeteksi konten-konten tersebut secara otomatis dan akurat. Penelitian ini bertujuan untuk mengembangkan sistem deteksi konten rasis, sara, dan pornografi di *Twitter* menggunakan metode Support Vector Machine (SVM). SVM adalah metode klasifikasi yang menggunakan hyperplane untuk memisahkan antara kelas konten negatif dan positif. Dalam penelitian ini, dilakukan klasifikasi teks yang melibatkan seleksi fitur menggunakan metode Information Gain untuk mengidentifikasi fitur-fitur yang relevan dalam proses klasifikasi antara kelas konten negatif (rasis, sara, pornografi) dan konten positif.

Kata kunci: *konten negatif, twitter, support vector machine*

1. Introduction

Media sosial adalah sebuah wadah berbasis daring yang memungkinkan penggunanya berinteraksi dengan orang lain tanpa adanya batasan waktu, bahkan wilayah. Indonesia merupakan salah satu negara dengan angka pengguna media sosial tertinggi di dunia. Kementerian Komunikasi dan Informatika (Kemenkominfo) menyatakan 95% dari sekitar 63 juta pengguna internet adalah pengguna media sosial [1]. *Twitter*, sebagai salah satu media sosial yang sering digunakan oleh masyarakat Indonesia, berperan sebagai alat yang memungkinkan mereka untuk berbagi informasi dan pendapat. Country Industry Head Twitter Indonesia mengklaim bahwa Indonesia merupakan negara dengan pertumbuhan pengguna aktif harian Twitter-nya paling besar [2].

Metode Machine learning adalah pembelajaran yang didasarkan pada sekumpulan sampel data berlabel. Kumpulan sampel digunakan untuk meringkas karakteristik distribusi skala perilaku pada setiap jenis aplikasi untuk membentuk model perilaku dari data. Pembelajaran yang diawasi kemudian dikelompokkan ke dalam masalah klasifikasi dan regresi. Masalah klasifikasi terjadi ketika variabel keluaran bersifat kategoris, seperti merah atau biru, atau penyakit dan tidak ada penyakit. Sementara itu, masalah regresi muncul ketika variabel output adalah nilai riil, seperti dolar atau bobot. Supervised learning memiliki beberapa algoritma populer seperti Back propagation, linear regression, random forest, support vector machine, naive bayesian, metode Rocchio, decision tree, k-nearest neighbor, neural network, logistic regression, dan neural network [3].

Setiap klasifikasi memiliki kelebihan dan kekurangannya masing-masing. Kelebihan Decision Tree adalah bersifat fleksibel untuk dapat meningkatkan kualitas keputusan yang dihasilkan, sedangkan kelemahan dari algoritma ini adalah akan terjadi tumpang tindih jika menggunakan data dengan jumlah kelas dan kriteria yang banyak. Meskipun keuntungan dari metode Naïve Bayes adalah perhitungannya

sederhana untuk membuat proses lebih cepat dan lebih efisien, mengingat fakta bahwa setiap variabel independen, hal ini dapat mengurangi keakuratan hasil. Terakhir k-nearest neighbor, kelebihan dari metode ini adalah dapat diterapkan secara efektif pada data yang besar dengan hasil yang akurat, namun kekurangannya adalah membutuhkan biaya komputasi yang tinggi karena harus menghitung jarak satu sama lain pada setiap individu. [4].

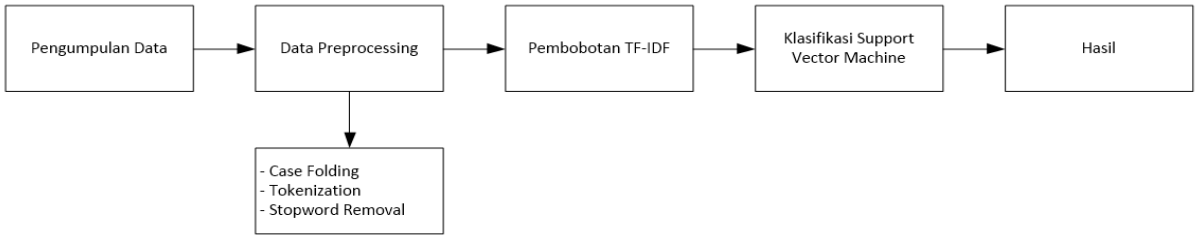
Support Vector Machine (SVM) adalah salah satu metode dalam pembelajaran terbimbing (supervised learning) yang umumnya digunakan untuk melakukan klasifikasi. Metode ini mengubah vektor fitur masukan ke dalam dimensi non-linier yang lebih tinggi [5]. Namun, SVM memiliki beberapa kelemahan ketika dihadapkan pada jumlah data yang besar. SVM tidak efisien dalam menyelesaikan masalah optimasi tanpa kendala dan menggunakan fungsi kernel untuk memisahkan bidang non-linier. Hal ini menyebabkan SVM memerlukan penggunaan memori yang sangat besar, dan seringkali mengalami masalah kehabisan memori sebelum proses pencarian solusi dimulai. Ketika jumlah data melebihi batas, SVM tidak dapat memberikan solusi klasifikasi yang memadai [6]. Selain itu, SVM juga sensitif terhadap pengaturan parameter dan data pelatihan. SVM tidak mampu secara otomatis memilih parameter yang optimal, sehingga penggunaan parameter yang tidak sesuai dapat menghasilkan performa yang suboptimal [7].

Dalam menangani masalah pemilihan parameter kernel, pendekatan yang diusulkan adalah menentukan nilai-nilai parameter kernel dan parameter cost (C) untuk melakukan optimasi. Tujuannya adalah memilih parameter terbaik untuk mengoptimalkan data latihan dan melakukan prediksi pada data uji, serta menghitung tingkat keakuratan prediksi [8]. Namun, dalam menghadapi data besar, diperlukan metode tambahan untuk memilih parameter yang sesuai dalam metode Support Vector Machine. Penelitian ini melakukan beberapa eksperimen untuk mencapai akurasi optimal, dan metode yang diusulkan adalah menerapkan algoritma optimasi pada parameter dalam Support Vector Machine [9].

Tujuan dari penelitian ini adalah untuk mengembangkan sistem deteksi konten negatif di Twitter menggunakan metode Support Vector Machine (SVM) dan seleksi fitur Information Gain. Sistem ini bertujuan untuk secara otomatis dan akurat mendeteksi konten-konten negatif, seperti konten rasis, konten yang mengandung ujaran kebencian (sara), dan konten pornografi. Dengan adanya sistem deteksi ini, diharapkan dapat mencegah penyebaran konten negatif yang merugikan pengguna dan masyarakat di media sosial Twitter, diharapkan pengguna sosial media lebih bijak lagi dalam berkomentar dan dapat menjaga etika dalam berkomunikasi [10].

2. Research Method

Penelitian ini dilakukan melalui studi literatur 4 jurnal dengan beberapa tahapan penelitian yang terdiri dari: (1) Pengumpulan Data, (2) Data Preprosessing, (3) Pembobotan Probabilitas TF-IDF, dan (4) Klasifikasi SVM dan Hasilnya. Diagram alur tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Desain Alur Sistem

.

2.1 Pengumpulan Data

Pada tahapan ini, akan dilakukan pengumpulan data melalui 4 jurnal ilmiah untuk memahami metode yang terkait dengan Support Vector Machine dalam menyelesaikan permasalahannya. Dalam

pengembangan sistem Klasifikasi konten negatif di twitter ini user akan melakukan crawling data dari twitter untuk mendapatkan dataset berupa konten tweet yang mengandung konten negatif dan positif dengan implementasi text mining untuk klasifikasi dataset tweet yang mana diubah menjadi file dengan format CSV yang dibagi menjadi 2 kelas klasifikasi, yaitu positif dan negatif. Kumpulan tweet yang berisi data tweet sebagai data uji [5].

2.2 Data Preprosessing

Dalam proses preprocessing, langkah-langkah yang dilakukan pada tweet-termasuk penghapusan tautan, penghapusan karakter khusus, mengubah huruf kapital menjadi huruf kecil, menghapus kata-kata stopword, dan stemming [2].

- **Case Folding:**

  Ubah semua huruf dalam tweet menjadi huruf kecil. Hal ini akan membantu menghilangkan perbedaan kapitalisasi yang tidak relevan.

  Contoh:

  Input: "I HATE this weather!"

- **Tokenization:**

  Pecah tweet menjadi token individu untuk memisahkan kata-kata dan tanda baca. Ini membantu dalam analisis lebih lanjut dan pemrosesan kata-kata secara terpisah.

  Contoh:

  Input: "I hate this weather!"

  Output setelah tokenization: ["I", "hate", "this", "weather", "!"]Output setelah case folding: "i hate this weather!"

- **Stopword Removal:**

  Hapus stopwords dari tweet. Stopwords dalam konteks ini dapat mencakup kata-kata umum seperti "a", "an", "the", "is", "in", dan sebagainya. Menghapus stopwords membantu mengurangi noise dan memfokuskan pada kata-kata yang lebih informatif.

  Contoh:

  Input: "I hate this weather!"

  Output setelah stopword removal: ["hate", "weather", "!"]

2.3 Pembobotan TF-IDF

Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode yang digunakan dalam pemrosesan teks untuk memberikan bobot pada kata-kata dalam dokumen berdasarkan pentingnya kata tersebut dalam koleksi dokumen yang lebih luas.

- **Term Frequency (TF):**

TF mengukur seberapa sering sebuah kata muncul dalam suatu dokumen. Meningkatnya frekuensi kata dalam dokumen akan menunjukkan bahwa kata tersebut penting dalam konteks dokumen tersebut. Nilai TF biasanya dihitung dengan menggunakan rumus:

TF = (Jumlah kemunculan kata dalam dokumen) / (Total kata dalam dokumen)

Contoh: Jika kata "hate" muncul 10 kali dalam sebuah dokumen yang berisi 100 kata, maka TF untuk kata "hate" adalah 10/100 = 0.1.

- **Inverse Document Frequency (IDF):**

IDF mengukur seberapa umum atau jarang sebuah kata dalam seluruh koleksi dokumen. Kata yang jarang muncul dalam koleksi dokumen memiliki nilai IDF yang tinggi, sementara kata yang sering muncul memiliki nilai IDF yang rendah. Rumus umum untuk menghitung IDF adalah:

IDF = log((Total dokumen dalam koleksi) / (Jumlah dokumen yang mengandung kata))

Contoh: Jika total dokumen dalam koleksi adalah 1000 dan kata "hate" muncul dalam 100 dokumen, maka IDF untuk kata "hate" adalah log(1000/100) = log(10) = 1.

Dengan menggunakan pembobotan TF-IDF, kata-kata yang memiliki bobot tinggi menunjukkan tingkat kepentingan yang lebih besar dalam konteks dokumen tersebut. Dalam pemrosesan teks, pembobotan TF-IDF digunakan untuk menemukan kata-kata kunci, membandingkan dokumen, melakukan clustering, atau membangun model klasifikasi yang lebih baik.[10]

2.4 Klasifikasi Support Vector Machine

Algoritma Support Vector Machine menggunakan kalsifikasi dan regresi. Penelitian ini menggunakan algoritma SVM dengan membagi data menjadi data test dan data train. Data test dilakukan dengan permodelan SVM dan mendapatkan hasil dengan accuracy, recall, precision dan f-1 score sebagai bahan perbandingan. Dengan dilakukannyan training pada klasifikasi SVM, maka akan menghasilkan sebuah nilai atau pola yang akan digunakan pada proses testing untuk proses testing SVM, Dalam evaluasi klasifikasi menggunakan algoritma Support Vector Machine (SVM), terdapat beberapa metrik evaluasi yang umum digunakan, yaitu accuracy, recall, precision, dan F1-score. Berikut adalah rumus-rumus untuk masing-masing metrik tersebut:

- **Accuracy (Akurasi):**

Accuracy mengukur sejauh mana model SVM dapat mengklasifikasikan data dengan benar secara keseluruhan. Rumusnya adalah:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

di mana TP (True Positive) adalah jumlah sampel positif yang diklasifikasikan dengan benar, TN (True Negative) adalah jumlah sampel negatif yang diklasifikasikan dengan benar, FP (False Positive) adalah jumlah sampel negatif yang salah diklasifikasikan sebagai positif, dan FN (False Negative) adalah jumlah sampel positif yang salah diklasifikasikan sebagai negatif.

- **Recall (Recall)**:

Recall, juga dikenal sebagai sensitivity atau true positive rate, mengukur sejauh mana model SVM dapat mengidentifikasi dan menemukan sampel positif dengan benar. Rumusnya adalah:

Recall = TP / (TP + FN)

Recall memberikan informasi tentang jumlah sampel positif yang berhasil ditemukan oleh model SVM dibandingkan dengan keseluruhan jumlah sampel positif yang sebenarnya.

- **Precision (Presisi):**

Precision mengukur sejauh mana model SVM mengklasifikasikan sampel sebagai positif dengan benar. Rumusnya adalah:

Precision = TP / (TP + FP)

Precision memberikan informasi tentang jumlah sampel positif yang benar diklasifikasikan oleh model SVM dibandingkan dengan total jumlah sampel yang diklasifikasikan sebagai positif.

- **F1-score:**

F1-score adalah ukuran gabungan yang memperhitungkan kedua precision dan recall. F1-score merupakan harmonic mean dari precision dan recall, dan berguna ketika perlu mempertimbangkan keseimbangan antara precision dan recall. Rumusnya adalah:

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

F1-score memberikan indikasi tentang kinerja keseluruhan model SVM dalam mengklasifikasikan data dengan mempertimbangkan keseimbangan antara precision dan recall.

Metrik evaluasi ini membantu dalam mengevaluasi kinerja model SVM dalam memprediksi kelas pada data yang belum pernah dilihat sebelumnya [10].

2.5 Hasil Klasifikasi

Dalam penelitian ini, terdapat hasil dari studi literatur dari 4 jurnal terdahulu yang membahas solusi dari klasifikasi SVM untuk menyelesaikan permasalahan konten negatif. Oleh karena itu, tujuan dibuatnya penelitian jurnal ini adalah untuk mengkaji, memahami, dan memastikan bahwa algoritma klasifikasi SVM tetap menjadi pilihan yang efektif dan dapat diandalkan dalam deteksi konten negatif pada media sosial Twitter. Dalam rangka mencapai tujuan tersebut, penelitian ini melakukan analisis terhadap studi literatur terdahulu yang telah mengadopsi SVM dalam konteks deteksi konten negatif. Dengan demikian, penelitian ini berkontribusi dalam menggabungkan pemahaman dan temuan dari penelitian-penelitian sebelumnya dengan konteks dan karakteristik khusus dari media sosial Twitter.

3. Result and Analysis

Hasil Pembahan jurnal ini menyajikan ringkasan dan analisis dari berbagai materi dari 4 jurnal yang relevan tentang Klasifikasi konten negatif pada Twitter menggunakan metode Support Vector Machine (SVM). Penelitian ini bertujuan untuk mengumpulkan dan menyusun informasi terkini tentang topik ini dari berbagai sumber dalam upaya untuk menghadirkan pemahaman yang komprehensif

3.1 Hasil Klasifikasi

Jurnal yang diterpilih

No	Penulis	Hasil Penelitian
1	(Dwi P., 2018)	Hasil identifikasi penelitian ini dibutuhkan untuk mengetahui apakah informasi yang diminta oleh pengguna sudah sesuai dengan informasi yang diberikan oleh sistem.Jumlah data yang digunakan adalah sebanyak 300 tweet, dimana 150 tweetbullydan 150 tweetbukan bully. Data divalidasi oleh Ibu Liana Shinta Dewi, M.A yang merupakan Dosen Bahasa Indonesia Universitas Brawijaya. Pada pengujian skenario perbandingan

		yang digunakan adalah 240 data training dan 60 data testing.didapatkan bahwa hasil yang didapatkan adalah konstan pada semua nilai lambda yang diujikan yaitu accuracy70% precision68,75%, recall73,33% dan f-measure70,96%.Hal ini terjadi karena nilai lambdahanya digunakan untuk melakukan perhitungan matriks hessian.
2	(Lalu M., 2019)	Pada penelitian ini, telah berhasil dibangun model untuk melakukan klasifikasi pada text. Model dibangun dengan menggunakan algoritma support vector machinedengan metode nonlinier. Model yang telah dibuat kemudian diuji dengan cara menghitung nilai presisi, recall, dan nilai akruasi. Hasil evaluasi menujukkan bahwa model yang dibangun cukup baik dalam melakukan klasifikasi dengan masing-masing diperoleh nilai presisi sebesar 83%, nilai recall sebesar 80%, dan didapatkan nilai uji presisi sebesar 80%
3	(Oryza., 2021)	Hasil pengujian dengan menggunakan beberapa fungsi kernel dan penggunaan data latih sebanyak 700 data serta data uji sebanyak 300 data, penggunaan kernel RBF menghasilkan nilai accuracy yang paling tinggi di antara kernel linear dan sigmoid. Kernel RBF memiliki nilai accuracy sebesar 93%, nilai precision sebesar 84%, nilai recall sebesar 86%, dan nilai F-measure sebesar 83%. Sistem klasifikasi ujaran kebencian diharapkan dapat membantu menganalisa perilaku masyarakat di dunia maya, yang dilihat berdasarkan hasil keluaran sistem yang selanjutnya dapat dianalisis lebih lanjut.

Dari beberapa jurnal literatur tentang klasifikasi konten negatif dengan metode Support Vector Machine (SVM), telah ditemukan bahwa SVM memberikan solusi yang efektif untuk menangani masalah tersebut. Penelitian-penelitian sebelumnya telah membuktikan bahwa SVM mampu memisahkan dengan baik konten negatif dari konten positif atau netral, dengan tingkat akurasi yang tinggi dan performa klasifikasi yang baik. Metode SVM ini telah berhasil mengatasi tantangan dalam mendeteksi konten negatif di media sosial, dan hasil-hasil penelitian tersebut memberikan dasar yang solid untuk pengembangan solusi yang lebih baik dan lebih andal dalam mengatasi masalah klasifikasi konten negatif.

4. Conclusion

Klasifikasi konten negatif pada media sosial, seperti Twitter, memiliki implikasi penting dalam meningkatkan pengalaman pengguna. Dengan mengidentifikasi dan menghapus konten negatif, platform media sosial dapat menciptakan lingkungan yang lebih positif dan aman bagi pengguna. Implementasi metode SVM yang diusulkan dalam penelitian ini dapat membantu platform media sosial untuk memonitor dan mengontrol konten yang tidak diinginkan atau berbahaya. dengan Preprocessing yang tepat pada tweet, termasuk penghapusan tautan, penghapusan karakter khusus, pengubahan huruf kapital menjadi huruf kecil, penghapusan kata-kata stopword, dan stemming, membantu dalam

mempersiapkan data untuk analisis. Selain itu, penggunaan metode pembobotan TF-IDF memberikan representasi fitur yang relevan dan memainkan peran penting dalam meningkatkan kinerja deteksi konten negatif.

References

[1] Firmansyah, E., & Rahman, A. B. A. (2023). Pengukuran Kesiapan Kota Cerdas Berdasarkan SNI ISO 37122: 2019. Infoman's: Jurnal Ilmu-ilmu Informatika dan Manajemen, 17(2).

[1] Kanthi, Y. A., Gumilang, K., & Aminah, S. (2024). Evaluasi Kepuasan Pengguna BRImo Menggunakan EUCS. Teknika, 13(1), 155-163.

[2] Anahyu, Y. D. (2024). Analisis Kepuasan Pengguna Akhir Aplikasi Mytelkomsel Menggunakan Metode End User Computing Satisfaction (EUCS). JURNAL INOVTEK POLBENG-SERI INFORMATIKA, 9(1).

[3] Firmansyah, E., Helmiawan, M. A., Fadil, I., Mahardika, F., Budiana, D., & Marliana, R. R. (2022, September). Integrated Academic Information System Based on The Perception of Ease And Benefits. In 2022 10th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE.

[4] Azzumar, M. F. (2023). Analisis Kepuasan Pengguna Terhadap Aplikasi Mobile Tiket. Com Menggunakan Metode End User Computing Satisfaction (EUCS) yang Dikembangkan (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).

[5] Zulfikar, W. B., Irfan, M., Ghufron, M., Jumadi, J., & Firmansyah, E. (2020). Marketplace affiliates potential analysis using cosine similarity and vision-based page segmentation. Bulletin of Electrical Engineering and Informatics, 9(6), 2492-2498.

[6] Rospita, S., Firmansyah, E., & Helmiawan, M. A. (2024). Implementasi Google Family Link Sebagai Solusi Pengawasan Penggunaan Gadget Anak di Desa Sundamekar. JIMT: Jurnal Informatika, Multimedia dan Teknik, 1(1), 83-88.

[7] Wahana, A., Firmansyah, E., Al Rosyid, H. I., Fuadi, R. S., & Maylawati, D. S. A. (2021). Fuzzy Tahani Method in the Recommendation System for Selecting Mountain Tourism Destinations in West Java.

[8] Helmiawan, M. A., Fadil, I., Sofiyan, Y., & Firmansyah, E. (2021, September). Security model using intrusion detection system on cloud computing security management. In 2021 9th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-5). IEEE.

[9] Firmansyah, E., Herdiana, D., Yuniarto, D., & Junaedi, D. I. (2021, September). The K-Nearest Neighbor Algorithm for the Classification of Internet Users in Rural Campus. In 2021 9th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE.

[10] N. Abdulloh, "Aplikasi deteksi cyberbullying pada media sosial twitter," 2019.

[11] A. F. Hidayatullah, "Deteksi Cyberbullying pada Cuitan Media Sosial Twitter," journal.uii.ac.id, vol. 1, p. 1, 2019.

[12] J. Homepage, A. Roihan, P. Abas Sunarya and A. S. Rafika, "Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," IJCIT (Indonesian Journal on Computer and Information Technology), vol. 5, no. 1, pp. 75-82, 2019.

[13] A. Prayoga Permana, K. Ainiyah and K. Fahmi Hayati Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," 2021.

[14] O. V. P. F. R. P. Resa Triyana, "Deteksi Cyberbullying Pada Tweet Berbahasa Inggris Dengan Metode Support Vector Machine," prosiding.unipma.ac.id, vol. 1, 2020.

[15] S. W. P. Epa Suryanto, "Perbandingan Reduced Support Vector Machine dan Smooth Support Vector Machine untuk Klasifikasi," ejurnal.its.ac.id, vol. 4, no. 2337-3520 (2301-928X Print), p. 1, 2015.

[16] A. N. Z. A. &. H. S. Styawati, "Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly," Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), vol. 5, no. 904 - 910, p. 2, 2021.

[17] K. E. T. L. Rooy Thaniket, "Prediksi kelulusan mahasiswa tepat waktu menggunakan algoritma support vector machine," JURNAL FATEKSA: Jurnal Teknologi dan Rekayasa, vol. 5, p. 2, 2020.

[18] Ispandi, "Penerapan Algoritma Genetika untuk Optimasi Parameter pada," Journal of Intelligent Systems, vol. 1, no. ISSN 2356-3982, p. 2, 2015.

[19] A. R. I. .. Y. Miftahuddin, "Implementasi SVM Untuk Deteksi Komentar Negatif Berbahasa Indonesia di Twitter," eproceeding.itenas.ac.id, vol. x, no. x, p. x, 2022.