

PENINGKATAN MODEL PREDIKSI KONTEN CYBERBULLYING PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Reksa Budiarti¹, Ade Irma Purnamasari², Agus Bahtiar³, Edi Wahyudin⁴

¹²Teknik Informatika, STMIK IKMI Cirebon

³Sistem Informasi, STMIK IKMI Cirebon

⁴Komputerisasi Akuntansi, STMIK IKMI Cirebon

Email : ¹reksabudiarti@gmail.com, ²irma2974@yahoo.com, ³agusbahtiar038@gmail.com, ⁴ediwahyudin@gmail.com

^{1,2,3}STMIK IKMI Cirebon

Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon

ABSTRAK

Platform media sosial seperti Instagram telah menjadi sangat populer di kalangan masyarakat di seluruh dunia. Namun, cyberbullying dapat membahayakan kesehatan mental pengguna di platform ini. Dengan menggunakan algoritma Support Vector Machine (SVM), penelitian ini bertujuan untuk meningkatkan model prediksi konten cyberbullying di Instagram. Dengan meningkatnya jumlah data di media sosial, mengidentifikasi konten negatif seperti cyberbullying secara otomatis menjadi sangat penting untuk membuat internet lebih aman. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) yang merupakan salah satu metode pengajaran yang diawasi yang paling efektif untuk klasifikasi data. Dataset yang digunakan terdiri dari komentar Instagram yang diklasifikasikan sebagai bullying atau non-bullying. Untuk meningkatkan kualitas fitur teks yang diolah, proses pengolahan data melibatkan tahap preprocessing seperti cleaning, tokenisasi, stopword removal, stemming, dan TFIDF. Selanjutnya, model SVM diimplementasikan dan dioptimalkan untuk meningkatkan akurasi prediksi dengan beberapa parameter. Hasil penelitian menunjukkan bahwa algoritma SVM dapat memberikan hasil prediksi yang akurat dengan tingkat akurasi sebesar 79,76% dalam mengidentifikasi konten cyberbullying pada Instagram. Jumlah kasus yang termasuk dalam kategori prediksi bullying dan benar-benar termasuk kategori konten bullying adalah sebanyak 482 kasus. Sementara itu, yang benar-benar termasuk konten non-bullying terdapat 155 kasus. Selain itu, yang diprediksi benar sebagai kategori konten non-bullying berjumlah 243 kasus, dan yang termasuk dalam kategori prediksi non-bullying tetapi sebenarnya bullying terdapat 29 kasus. Ini menunjukkan bahwa SVM adalah metode yang efektif untuk menganalisis teks berbasis media sosial dalam mendeteksi tindakan bullying. Hasil penelitian ini dapat meningkatkan kesadaran masyarakat terhadap masalah cyberbullying dan dampak negatifnya.

Keywords: Support Vector Machine (SVM), cyberbullying, Prediksi konten, Instagram, Sosial Media.

ABSTRACT

Social media platforms such as Instagram have become very popular among people around the world. However, cyberbullying can harm the mental health of users on this platform. Using the Support Vector Machine (SVM) algorithm, this study aims to improve the prediction model of cyberbullying content on Instagram. With the increasing amount of data on social media, automatically identifying negative content such as cyberbullying has become very important to make the internet safer. This study uses the Support Vector Machine (SVM) algorithm, which is one of the most effective supervised learning methods for data classification. The dataset used consists of Instagram comments classified as bullying or non-bullying. To improve the quality of the processed text features, the data processing process involves preprocessing stages such as cleaning, tokenization, stopword removal, stemming, and TF-IDF. Furthermore, the SVM model is implemented and optimized to improve the prediction accuracy with several parameters. The results of the study show that the SVM algorithm can provide accurate prediction results with an accuracy rate of 79.76% in identifying cyberbullying content on Instagram. The number of cases included in the bullying prediction category and actually included in the bullying content category is 482 cases. Meanwhile, there were 155 cases that were truly non-bullying content. In addition, there were 243 cases that were correctly predicted as non-bullying content and 29 cases that were predicted as non-bullying but were actually bullying. This shows that SVM is an effective method for analyzing social media-based texts in detecting bullying. The results of this study can increase public awareness of the problem of cyberbullying and its negative impacts.

Keywords: Support Vector Machine (SVM), cyberbullying, Content prediction, Instagram, Social Media.

1. PENDAHULUAN

Media sosial telah menjadi bagian penting dari kehidupan sehari-hari orang modern, dan Instagram adalah salah satu platform yang paling banyak digunakan. Namun, cyberbullying atau perundungan online juga meningkat seiring dengan meningkatnya penggunaan media sosial [1]. Cyberbullying adalah tindakan buruk yang dilakukan secara online, yang dapat berupa komentar yang menunjukkan kebencian, penghinaan, atau ancaman. Cyberbullying dapat berdampak besar pada pengguna, terutama pada remaja, dan dapat menyebabkan gangguan psikologis seperti kecemasan dan depresi serta penurunan kualitas hidup [2]. Akibatnya, untuk mengurangi efek negatif dari cyberbullying, diperlukan sistem yang dapat mendeteksi dan mengklasifikasikan konten secara otomatis. Studi sebelumnya menunjukkan bahwa algoritma Support Vector Machine (SVM) sangat efektif dalam mengklasifikasi teks, terutama untuk mendeteksi cyberbullying. Untuk meningkatkan keakuratan prediksi, metode ini sering dikombinasikan dengan metode prapemrosesan teks seperti TF-IDF dan Bag of Words [3]. Namun, prediksi yang akurat sering terhambat oleh variasi bahasa, konteks, dan slang yang digunakan dalam komentar Instagram. Teknik deep learning telah dicoba, tetapi biasanya membutuhkan sumber daya komputasi yang besar dan sulit diimplementasikan pada skala besar. Menurut beberapa penelitian, pelaku cyberbullying menggunakan media sosial terutama Instagram [2]. Dalam situasi seperti ini, mengidentifikasi dan memperkirakan konten cyberbullying menjadi penting untuk mengurangi efek negatif dari media sosial. Dengan mengoptimalkan algoritma Support Vector Machine, penelitian ini bertujuan untuk meningkatkan model prediksi cyberbullying. Ini penting karena hasil dari model prediksi dapat digunakan secara efektif dengan mempertimbangkan berbagai jenis variasi teks yang biasa ditemukan dalam komentar Instagram [3].

Studi sebelumnya telah menggunakan algoritma Support Vector Machine (SVM) untuk mengkategorikan teks yang berkaitan dengan cyberbullying. Namun, temuan penelitian menunjukkan bahwa keanekaragaman bahasa, gaya bahasa, dan konteks yang digunakan di media sosial adalah kendala terbesar dalam mendeteksi cyberbullying. Di Instagram, deteksi teks cyberbullying menjadi lebih sulit karena kontennya sering mengandung bahasa informal, slang, hingga singkatan. Selain itu, adanya elemen visual yang memungkinkan pengguna mengirimkan pesan dalam bentuk gambar atau video juga menjadi masalah. [3] menggunakan kombinasi Teknik Bag of Words dan TF-IDF bersama SVM untuk meningkatkan keakuratan prediksinya. Namun, temuan mereka terbatas pada teks formal. Sebuah penelitian tambahan oleh [2] menunjukkan bahwa metode klasifikasi saat ini tidak dapat mencakup semua aspek percakapan di media sosial. Akibatnya, ada kemungkinan besar bahwa sistem klasifikasi salah mengidentifikasi konten yang tidak bersifat bullying sebagai konten cyberbullying atau sebaliknya. Ini menunjukkan bahwa deteksi cyberbullying perlu ditingkatkan di platform media sosial, khususnya Instagram. Selain itu, metode klasifikasi saat ini harus diperbarui untuk menyesuaikan dengan bahasa dan gaya komunikasi di media sosial. Tanpa peningkatan yang signifikan dalam Teknik deteksi, cyberbullying akan terus mengancam pengguna media sosial, terutama kelompok usia rentan seperti remaja.

Metode klasifikasi berbasis teks seperti Support Vector Machine (SVM) dan teknik prapemrosesan teks biasanya menjadi pusat penelitian sebelumnya tentang deteksi cyberbullying. [3] menggabungkan teknik Bag Of Words dan TF-IDF dengan SVM untuk memperkirakan konten berbahaya, terutama komentar berbahaya di media sosial, dengan lebih akurat. Selain itu, penelitian yang dilakukan oleh [2] menunjukkan bagaimana metode SVM dapat diterapkan pada media Instagram. Namun, karena SVM tidak dapat menangani konteks dan variabel bahasa yang beragam, hasilnya masih belum ideal. Studi [1] menemukan bahwa metode yang menggunakan semantic similarity dan SVM dapat menjadi alternatif yang bagus untuk mendeteksi konten cyberbullying dengan lebih akurat di media sosial. Namun, metode ini masih membutuhkan pengujian lebih lanjut.

Dengan menggunakan algoritma *Support Vector Machine* (SVM) yang dioptimalkan, penelitian ini bertujuan untuk mengembangkan dan meningkatkan model prediksi konten cyberbullying di platform Instagram. Ada kemungkinan bahwa model yang lebih akurat akan membantu mencegah dan mengurangi efek buruk dari pelecehan online terhadap pengguna Instagram, terutama bagi remaja yang rentan terhadap tekanan sosial. Selain itu, penelitian ini diharapkan dapat membantu pengembang platform media sosial secara otomatis memantau dan menyaring konten yang berpotensi berbahaya. Diharapkan bahwa temuan penelitian ini akan melengkapi kekurangan studi sebelumnya yang belum mampu menangani berbagai gaya bahasa dan komunikasi yang ada di media sosial. Akibatnya, penelitian ini tidak

hanya bermanfaat bagi komunitas ilmiah tetapi juga memiliki nilai praktis untuk meningkatkan kualitas interaksi di media sosial.

Untuk meningkatkan akurasi prediksi konten cyberbullying, penelitian ini akan menggunakan kombinasi Teknik prapemrosesan teks seperti TF-IDF dan algoritma Support Vector Machine. TF-IDF dipilih karena kemampuan untuk menemukan kata-kata yang memiliki bobot besar dalam kalimat atau paragraf, yang penting untuk menemukan kata-kata yang memiliki bobot besar dalam kalimat atau paragraf, yang penting untuk menemukan kata beracun. Diharapkan, algoritma klasifikasi SVM dapat mengklasifikasi teks bullying dan non-bullying dengan lebih akurat karena memiliki keunggulan dalam memisahkan data berdasarkan margin optimal. Dataset yang digunakan akan mencakup berbagai komentar Instagram yang telah dikategorikan dan dianotasi sebelumnya. Diharapkan metode ini akan memberikan hasil yang lebih akurat daripada metode sebelumnya karena penggunaan dataset yang relevan dan representatif. Perangkat lunak yang mendukung implementasi SVM akan digunakan untuk melakukan seluruh proses analisis dan pengujian.

Dengan model prediksi yang akurat dan relevan, penelitian ini diharapkan dapat membantu mengurangi tingkat cyberbullying di Indonesia, khususnya di Instagram. Dengan demikian, platform media sosial dapat menggunakan teknologi ini untuk mendeteksi konten yang melanggar kebijakan mereka dengan lebih cepat dan efisien. Selain itu, model prediksi yang dikembangkan dalam penelitian ini dapat berguna untuk lembaga pemerintah, sekolah, dan organisasi non pemerintah dalam memantau dan menangani kasus cyberbullying. Hal ini juga akan membantu meningkatkan kesadaran publik tentang pentingnya menjaga keamanan internet dan bertanggung jawab dalam penggunaan media sosial. Implementasi model ini dapat membantu menciptakan lingkungan yang lebih aman dan mengurangi tekanan psikologis yang dialami oleh korban cyberbullying. Penelitian ini juga diharapkan dapat mendorong penelitian lebih lanjut tentang penggunaan pembelajaran mesin dalam bidang keamanan digital, khususnya yang berkaitan dengan bahasa dan budaya lokal. Pada akhirnya, penelitian ini akan berkontribusi pada pengembangan teknologi yang lebih sesuai dengan kebutuhan pengguna media sosial Indonesia.

2. TINJAUAN PUSTAKA

2.1 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah salah satu Teknik dalam machine learning yang berdasarkan teori struktural pembelajaran statistik. SVM mempunyai performa yang lebih baik apabila dibandingkan dengan metode machine learning lain. Proses utama dari SVM adalah menemukan batas yang memisahkan tiap kelas (hyperlane) [4]. *Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin berbasis statistik yang bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam kelas tertentu. SVM menggunakan vector support untuk menemukan pemisahan terbaik dengan margin maksimal untuk meningkatkan akurasi prediksi. SVM memiliki banyak keunggulan, terutama karena mereka dapat menangani data berukuran besar dan tahan terhadap overfitting, terutama ketika digunakan dengan teknik kernel trick seperti RBF dan polynomial. SVM dapat mendeteksi pola berbasis fitur seperti TF-IDF dalam klasifikasi teks, termasuk deteksi cyberbullying. Namun, pemilihan parameter dan ketidakseimbangan data dapat memengaruhi kinerjanya, jadi teknik balancing seperti SMOTE sering digunakan. SVM masih digunakan untuk analisis teks dan deteksi cyberbullying di media sosial, meskipun memiliki keterbatasan.

2.2 *Cyberbullying*

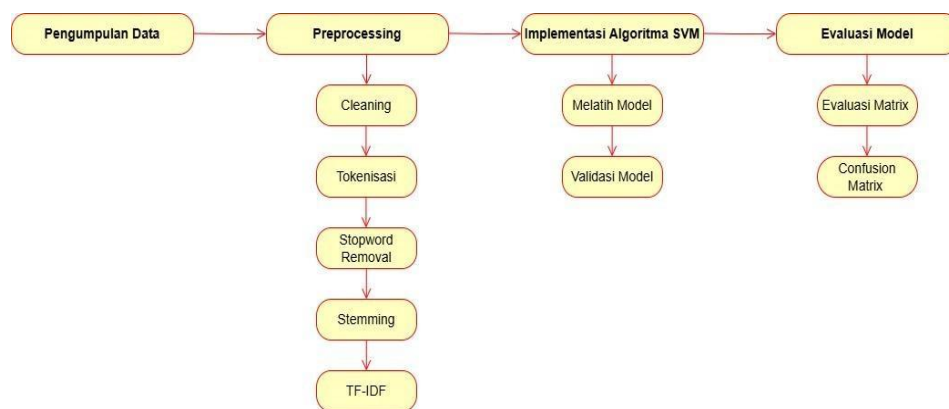
Cyberbullying adalah tindakan negatif yang dilakukan oleh seseorang atau kelompok tertentu dengan cara mengirimkan pesan teks, foto, gambar meme, dan video ke akun media sosial seseorang dengan tujuan untuk menyindir, menghina, melecehkan, mendiskriminasi bahkan mempersekusi individu. Berdasarkan hasil data statistik, sebagian besar pelaku cyberbullying didominasi adalah remaja [5].

2.3 Instagram

Instagram merupakan salah satu platform media sosial yang populer yang dapat dimanfaatkan oleh perpustakaan untuk memenuhi kebutuhan informasi pengguna. Instagram pertama kali diluncurkan pada Oktober 2010 oleh Kevin Systrom dan Mike Krieger, dua insinyur perangkat lunak yang belajar di Universitas Stanford. Instagram berasal adalah dari dua kata, yaitu “insta” dan “gram”. Kata “insta” berasal dari kata instan, yang berarti pengguna aplikasi ini dapat membagikan foto maupun video secara instan. Sedangkan “gram” berasal dari kata telegram, yang berarti pengguna aplikasi ini dapat membagikan foto maupun video dengan cepat [6].

3. METODE PENELITIAN

Pada gambar 1 menunjukan tahapan penelitian dimulai dengan melalui tahapan pengumpulan data, preprocessing, seperti cleaning, tokenisasi, stopwords removal, stemming, dan representasi teks menggunakan TF-IDF. Algoritma Support Vector Machine (SVM) dilatih dengan dataset yang telah diproses, lalu dievaluasi menggunakan data uji untuk mengukur akurasi, presisi, dan recall. Hasil evaluasi digunakan untuk menarik kesimpulan dan memberikan rekomendasi.



Gambar 1 Tahapan Teknik Analisis Data

Berikut adalah rincian dari masing-masing tahap:

1. Pengumpulan Data

Proses sistematis untuk mengumpulkan, merekam, dan mengorganisasi informasi yang relevan dengan tujuan penelitian dikenal sebagai pengumpulan data. Data dalam machine learning dapat berasal dari berbagai sumber, seperti data primer yang dikumpulkan melalui observasi, wawancara, dan kuesioner, atau data sekunder dari artikel ilmiah, jurnal, atau dataset yang tersedia di repositori publik seperti Kaggle. Selain itu, pengaksesan data secara otomatis sering dilakukan dengan scraping data dari situs web atau menggunakan API dari platform seperti Instagram. Tujuan dari proses ini adalah untuk memberikan data berkualitas tinggi yang akan digunakan untuk melatih, menguji, dan mengevaluasi model, memastikan bahwa hasil analisis benar dan relevan. Sebelum mengumpulkan data, penting untuk melakukan preprocessing agar data bersih, terorganisir, dan siap digunakan.

2. Preprocessing adalah tahap awal pengolahan data, khususnya data teks, yang bertujuan untuk mempersiapkan data agar dapat digunakan dengan lebih baik oleh algoritma pengajaran mesin.

a. Cleaning

Cleaning adalah Untuk membuat data lebih rapi dan siap untuk dianalisis atau diproses lebih lanjut dalam model pembelajaran mesin, proses pembersihan dilakukan pada data teks.

b. Tokenisasi

Tokenisasi adalah proses membagi teks menjadi unit yang lebih kecil yang disebut token. Tanda dapat berupa kata-kata, frasa, karakter, atau unit-unit lainnya tergantung pada tujuan analisis atau konteks pemrosesan teks yang dilakukan. Tokenisasi penting dalam pemrosesan bahasa alami dan analisis teks karena dapat memberikan representasi yang lebih terstruktur dan terperinci dari teks yang akan diproses [7].

c. Stopword Removal

Stopword Removal merupakan proses untuk membersihkan teks dari kata-kata umum yang tidak penting. Kata-kata itu berfungsi sebagai penghubung atau pengganti kata lain, tetapi tidak memiliki makna tersendiri [8].

d. Stemming

Stemming adalah proses mengubah kata turunan atau kata dengan imbuhan (awalan, sisipan, atau akhiran) menjadi bentuk dasarnya (root word) atau kata akar. Tujuan stemming adalah untuk menyederhanakan teks sehingga kata-kata dengan makna yang sama dapat diproses secara konsisten dalam analisis teks atau pembelajaran mesin.

e. TF-IDF

TF-IDF adalah metode statistik yang banyak digunakan untuk menunjukkan pentingnya sebuah kata dalam suatu dokumen dibandingkan dengan keseluruhan korpus dokumen. Metode TF-IDF akan menghitung nilai pada setiap kata yang telah diubah menjadi token dalam tahap tokenization sebelumnya [9].

3. Implementasi Algoritma SVM

a. Melatih Model

Melatih model adalah proses di mana algoritma pembelajaran mesin mempelajari pola dari data latih untuk membangun model prediksi atau klasifikasi. Dalam kasus Support Vector Machine (SVM), melatih model berarti menemukan hyperplane terbaik yang dapat memisahkan data ke dalam kelas yang berbeda dengan margin maksimum.

b. Validasi Model

Proses validasi model dilakukan untuk memeriksa kinerja model pembelajaran mesin dengan menggunakan data uji atau validasi yang berbeda dari data latih. Tujuan dari validasi model adalah untuk memastikan bahwa model yang dilatih mampu melakukan generalisasi dengan baik terhadap data baru daripada hanya menghafal data latih (overfitting).

4. Evaluasi Model

a. Evaluasi Matrix

Evaluasi Matrix adalah metode untuk mengukur performa model machine learning setelah dilatih menggunakan data latih. Dalam proses ini, metrik evaluasi digunakan untuk memahami seberapa baik model bekerja, terutama dalam tugas klasifikasi atau prediksi. Metrik ini membantu mengevaluasi akurasi, presisi, recall, F1 score, dan kinerja lainnya berdasarkan hasil prediksi terhadap data uji atau validasi.

b. Confusion Matrix

Tabel matriks Confusion Matrix digunakan untuk menilai kinerja model klasifikasi dalam pembelajaran mesin. Matriks ini menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas dalam dataset uji, memberikan gambaran lengkap tentang bagaimana model menangani data.

4. PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan teknik scrapping dengan menggunakan tools Data miner, menghasilkan komentar bullying dan non-bullying. Hasil dari pengumpulan data yaitu sebesar 909 data yang sudah diberi label meliputi 511 bullying dan 398 non-bullying. Berikut adalah sebagian dataset dari hasil pengumpulan data menggunakan tools Data Miner. Pada tabel 1 menunjukan sebagian komentar yang dikategorikan sebagai bullying termasuk pernyataan bernada negatif, menghina, atau merendahkan, seperti penghinaan fisik, moral, atau karakter, seperti "jijik", "tubuh triplek", dan "muka polos kelakuan skandal abis", yang semuanya mencerminkan perilaku cyberbullying. Dan pada tabel 2 menunjukan sebagian komentar non-bullying yang mencakup pernyataan yang bersifat positif atau netral yang tidak mengandung unsur penghinaan atau penghinaan. Komentar seperti "gemes banget", "cantiknya kelewatan", dan "cakep banget ya Tuhan" menunjukkan pujian atau kekaguman.

Tabel 1 Komentar Bullying

Komentar	Kategori
pentingnya ngasih makan anak pakai uang halal ya gini	Bullying
muka polos kelakuan skandal abis	Bullying
nyesel banget pernah idolain cewe spek wc umum ini	Bullying
jijik	Bullying
tubuh triplek	Bullying
.....

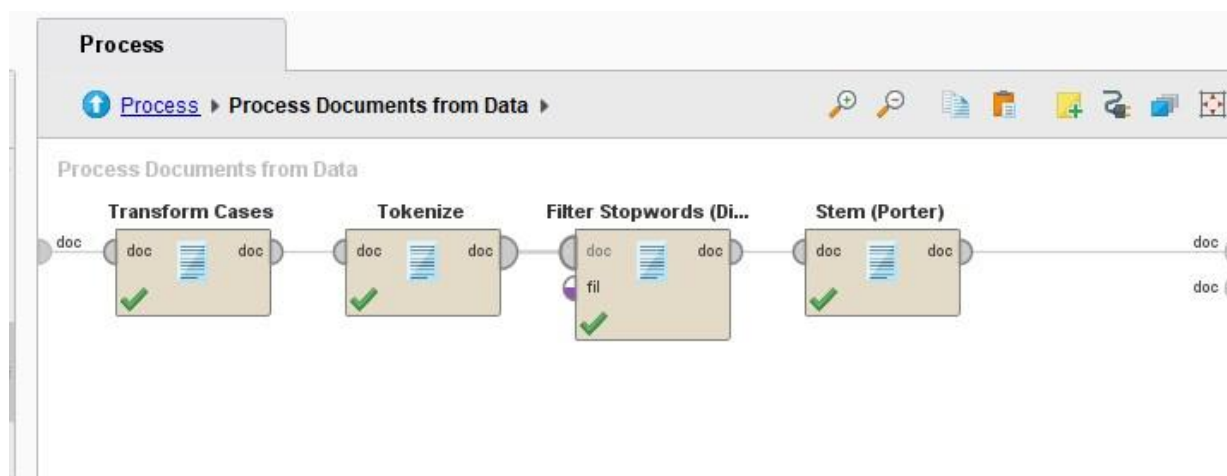
Tabel 2 Komentar Non-Bullying

Komentar	Kategori
gemes banget	Non-Bullying
cip agak mundur dikit cantiknya kelewatan	Non-Bullying
cakep banget ya tuhan	Non-Bullying
barbie mahal	Non-Bullying
gila makin cantik	Non-Bullying
.....

4.2 Preprocessing

Preprocessing merupakan tahap pertama dari text mining yang mencakup semua rutinitas dan mempersiapkan data yang akan digunakan pada operasi pada text mining. Preprocessing digunakan untuk mempersingkat dan memperpendek kalimat yang telah diucapkan dalam group dengan cara mengambil kata-kata yang dianggap penting saja untuk mempermudah dan mempercepat proses identifikasi kasus cyberbullying. Preprocessing terdapat beberapa tahapan untuk mendapatkan kalimat atau kata yang berisi informasi yang diinginkan agar identifikasi dapat lebih akurat [10]. Pada gambar 2 proses preprocessing terdiri dari transform cases untuk mengubah teks menjadi huruf kecil, tokenize untuk memotong teks menjadi bagian-bagian lebih kecil, seperti kata-kata atau kalimat, filter stopwords untuk menghapus kata-

kata umum seperti “dan”, “di”, dan “yang”, dan stem untuk mengubah kata-kata menjadi bentuk dasarnya seperti “lari”, dan “pelari”.



Gambar 2 Proses Preprocessing

Dalam preprocessing di atas menggunakan empat operator diantaranya transform cases, tokenize, filter stopwords, dan stemming. Transform Cases untuk menghindari perbedaan kapitalisasi, semua huruf diubah menjadi huruf yang lebih kecil. Tokenize memecah teks menjadi token, seperti kata per kata. Untuk memfokuskan pada kata-kata penting, filter stopwords menghapus kata-kata umum yang tidak signifikan, seperti “dan” atau “di”. Terakhir, Porter menggunakan stem untuk mengubah kata ke bentuk aslinya (stemming). Ini dilakukan untuk menyederhanakan perbedaan kata dengan arti serupa, seperti “belajar” dan “belajar”, menjadi “belajar”. Tujuan dari proses ini adalah untuk membersihkan teks sehingga menjadi lebih konsisten dan siap untuk diolah dalam analisis atau model pembelajaran mesin. Pada tabel 3 menunjukkan hasil preprocessing teks, di mana setiap kata khusus yang diambil dari komentar ditampilkan bersama dengan atribut dan statistik frekuensi. Dalam kolom “Kata” terdapat kata-kata hasil pemrosesan, seperti “aamiin” dan “abang”, yang telah melalui proses seperti tokenisasi, penghapusan stopwords, dan stemming. Dalam kolom “Total Occurrences”, jumlah total kata yang muncul di seluruh data, dan “Dokumentasi” menunjukkan jumlah dokumen yang mengandung kata tersebut.

Tabel 3 Hasil Preprocessing

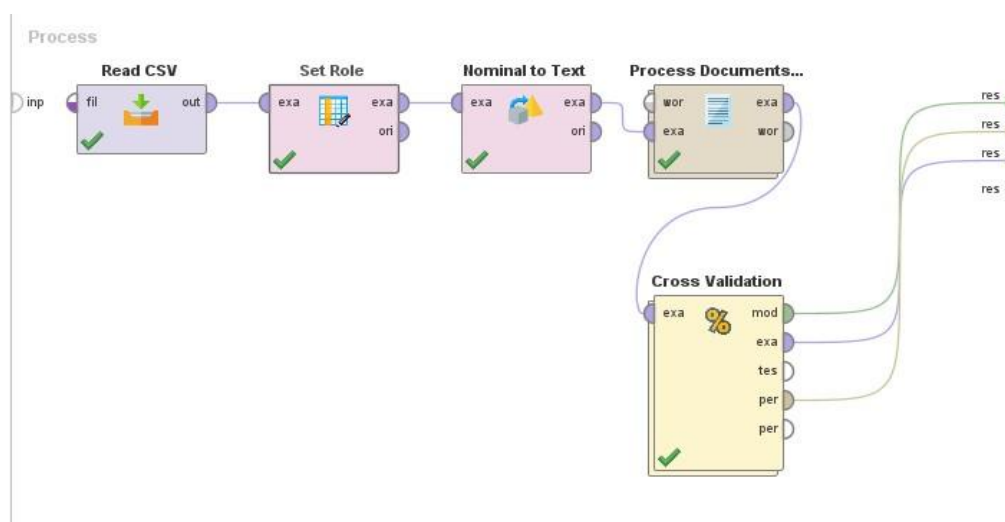
Word	Attribute Name	Total Occurrences	Documents
aamiin	aamiin	1	1
abang	abang	1	1
abis	abis	2	2
adek	adek	1	1
adem	adem	2	1
.....
adikku	adikku	1	1

4.3 Implementasi Algoritma SVM

a. Pelatihan Model

Tujuan dari proses ini adalah untuk melatih dan mengevaluasi model klasifikasi menggunakan teknik validasi silang dengan data teks. Data diproses melalui tahap

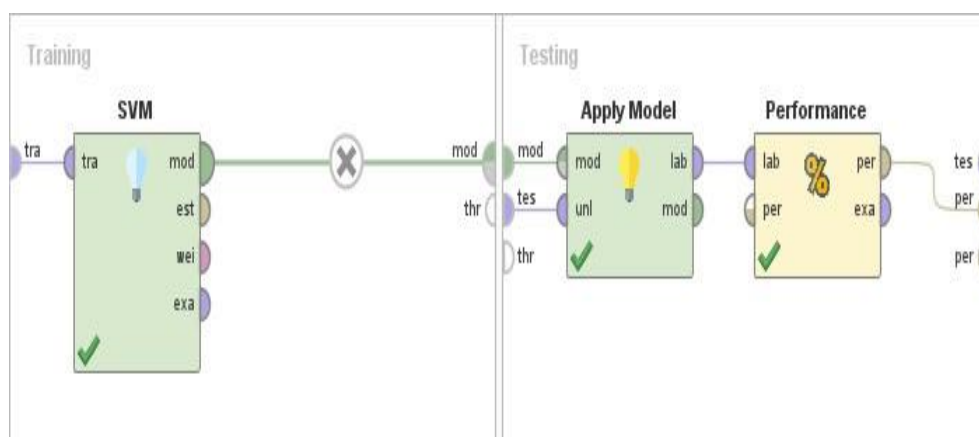
preprocessing agar siap diolah, dan kemudian model dievaluasi menggunakan validasi silang untuk memastikan kinerja yang stabil dan mengurangi overfitting. Pada gambar 3 pelatihan model menunjukkan yang dimulai dengan membaca data dari file CSV dan kemudian menetapkan peran untuk setiap kolom dataset sebagai fitur atau label (Set Role). Pada tahap Nominal to Text, data kategori kemudian diubah menjadi format teks sebelum proses preprocessing teks pada tahap Dokumen Prosedur, yang mencakup tokenisasi, penghapusan stopwords, dan stemming. Setelah itu, model dievaluasi menggunakan metode Cross Validation, yang membagi data menjadi beberapa bagian, untuk memastikan bahwa pelatihan dan pengujian model berjalan dengan benar dan menghasilkan hasil evaluasi yang dapat diandalkan.



Gambar 3 Pelatihan Model

b. Validasi Model

Proses validasi model SVM membagi data menjadi dua bagian, pelatihan dan pengujian. Model dilatih menggunakan SVM, kemudian diuji dengan data uji, dan kinerjanya dinilai dengan metrik performa. Proses ini memastikan bahwa model SVM dapat memprediksi data baru dengan akurasi yang dapat diukur. Pada gambar 4 validasi model menunjukkan proses validasi model yang menggunakan pelatihan dan pengujian melalui metode Support Vector Machine (SVM). Pada bagian pelatihan, model terlatih dibuat dengan menggunakan data pelatihan (training data) untuk menghasilkan model terlatih (mod). Pada bagian uji, model terlatih diterapkan pada data uji untuk memprediksi label pada data yang tidak diketahui. Pada langkah terakhir, kinerja, digunakan untuk mengevaluasi hasil prediksi dengan membandingkan label yang diprediksi dengan label yang sebenarnya menggunakan metrik seperti akurasi, presisi, atau recall.



Gambar 4 Validasi model

4.4 Evaluasi Model

a. Evaluasi Matrix

Hasil dari evaluasi matrix menunjukkan bahwa model lebih baik dalam memprediksi kelas bullying (recall 94,32%), tetapi kurang baik dalam mendeteksi nonbullying (recall 61,06%). Precision juga menunjukkan bahwa akurasi prediksi kedua kelas berbeda. Pada tabel 4 evaluasi menunjukkan hasil evaluasi menunjukkan akurasi model sebesar $79.76\% \pm 3.59\%$, yang menunjukkan bahwa model dapat memprediksi dengan cukup baik. Model memiliki presisi 75.67% untuk komentar Bullying dan 89.34% untuk Non-Bullying, dan tingkat recall-nya adalah 94.32% untuk Bullying, yang menunjukkan bahwa hampir semua komentar Bullying terdeteksi.

Tabel 4 Evaluasi Matrix

Accuracy: 79.76% \pm 3.59% (micro average: 79.76%)			
	True Bullying	True Non-Bullying	Class Precision
Pred. Bullying	482	155	75.67%
Pred. Non-Bullying	29	243	89.34%
Class Recall	94.32%	61.06%	

b. Confusion Matrix

Kinerja model klasifikasi dengan akurasi 79.76%, yang menunjukkan bahwa model memprediksi dengan benar sebagian besar data. Model memiliki True Positive (482), yang menunjukkan kemampuan yang baik untuk mendeteksi bullying, dengan recall 94.32%. Namun, False Positive (155) memiliki kelemahan, karena data nonbullying sering ditafsirkan sebagai bullying, sehingga akurasi untuk kelas bullying hanya 75.67% dan recall untuk kelas non-bullying hanya 61.06%. Meskipun model mendeteksi bullying berhasil, ada perbaikan yang diperlukan untuk mengurangi kesalahan prediksi dan meningkatkan kinerja di kelas yang tidak mengalami bullying. Pada tabel 5 confusion matrix yang menunjukkan bagaimana model memprediksi bullying dan data non-bullying. Model mencatat 482 komentar bullying dengan benar sebagai Positif Asli (TP) dan 243 komentar bullying dengan benar sebagai Positif Negatif (TN). Namun, 155 komentar bullying yang salah dianggap bullying (False Positive, FP), dan 29 komentar bullying yang salah dianggap bullying (False Negative, FN). Dari keseluruhan 909 komentar yang ada.

Tabel 5 Confusion Matrix

Predicted / Actual	True Bullying	True Non-Bullying	Total
Pred. Bullying	482 (TP)	155 (FP)	637
Pred. Non-Bullying	29 (FN)	243 (TN)	272
Total	511	398	909

5.1 KESIMPULAN

1. Dengan menggunakan alat scraping Data Miner, penelitian ini berhasil mengumpulkan dataset dari platform media sosial Instagram, yang terdiri dari 909 data, termasuk 511 kasus bullying dan 398 kasus non-bullying yang dikategorikan secara manual. Untuk melatih algoritma Support Vector Machine (SVM), proses preprocessing TF-IDF, yang mencakup proses pembersihan, tokenisasi, penghapusan stopword, stemming, dan pembobotan, memastikan bahwa data berada dalam format yang ideal. Dengan bantuan langkah-langkah ini, model dapat menganalisis data dengan kualitas terbaik. Ini menghasilkan prediksi yang lebih akurat dan efisien dalam mendeteksi konten yang diduga cyberbullying.
2. Dengan penggunaan algoritma Support Vector Machine (SVM) untuk mendeteksi konten cyberbullying di media sosial, khususnya dalam bahasa Indonesia, cukup menguntungkan. Dengan tingkat akurasi 79,76% dan margin error $\pm 3,59\%$, algoritma SVM berhasil mengklasifikasikan data. Tingkat recall yang tinggi sebesar 94,32% menunjukkan bahwa algoritma SVM mampu mendeteksi sebagian besar kasus cyberbullying dengan efektif. Meskipun demikian, masih ada 29 kasus yang salah yang dianggap sebagai non-bullying. Ini menunjukkan bahwa ada ruang untuk perbaikan dalam mengurangi kesalahan prediksi. Secara keseluruhan, SVM menunjukkan hasil yang baik dalam mendeteksi cyberbullying. Namun, perlu ada upaya tambahan untuk meningkatkan ketepatan dan mengurangi kesalahan klasifikasi.

5.2 SARAN

1. Menggabungkan algoritma Support Vector Machine (SVM) dengan teknik pembelajaran mendalam untuk meningkatkan akurasi dalam mendeteksi konten cyberbullying yang lebih kompleks, termasuk analisis multimodal yang melibatkan gambar atau video, akan memungkinkan pengembangan penelitian ini. Selain itu, penggunaan dataset yang lebih besar dan beragam akan membantu model menjadi lebih fleksibel terhadap berbagai gaya bahasa dan komunikasi. Selain itu, sangat penting untuk bekerja sama dengan situs media sosial seperti Instagram agar model ini dapat diterapkan secara langsung ke sistem moderasi otomatis.
2. Hasil penelitian ini dapat digunakan sebagai dasar untuk program penyuluhan tentang bahaya cyberbullying. Mereka juga dapat membantu pemerintah membuat kebijakan yang memanfaatkan teknologi berbasis kecerdasan buatan untuk mengawasi konten daring. Selain itu, disarankan untuk melakukan evaluasi jangka panjang terhadap model yang dikembangkan untuk memastikan bahwa itu berguna dan relevan di tengah perubahan dalam cara komunikasi di media sosial.

DAFTAR PUSTAKA

- [1] J. T. Informatika, F. Informatika, U. Telkom, T. B. Batu, and J. Barat-indonesia, "Identifikasi Cyberbullying pada Kolom Komentar Instagram dengan Metode Support Vector Machine dan Semantic Similarity," vol. 4, no. 1, pp. 1–8, 2020.
- [2] D. Nugraha, P. Astuti, P. S. Informatika, U. N. Mandiri, P. S. Informatika, and U. N. Mandiri, "Media Instagram Menggunakan Metode," vol. 8, no. 2, pp. 153–164, 2023.
- [3] D. Krisnandi, R. N. Ambarwati, A. Y. Asih, A. Ardiansyah, and H. F. Pardede, "Analisis Komentar Cyberbullying Terhadap Kata Yang Mengandung Toksisitas Dan Agresi Menggunakan Bag of Words dan TF-IDF Dengan Klasifikasi SVM," *J. Linguist. Komputasional*, vol. 6, no. 2, pp. 36–41, 2023, doi: 10.26418/jlk.v6i2.85.

- [4] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, “Perbandingan Metode Lexiconbased dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter,” *Fountain Informatics J.*, vol. 4, no. 2, p. 41, 2019, doi: 10.21111/fij.v4i2.3573.
- [5] D. Riswanto and R. Marsinun, “Perilaku Cyberbullying Remaja di Media Sosial,” *Analitika*, vol. 12, no. 2, pp. 98–111, 2020, doi: 10.31289/analitika.v12i2.3704.
- [6] S. Sinaga, M. A. Muqsith, and F. Ayuningtyas, “Instagram sebagai Media Informasi Digital Perpustakaan Universitas Pembangunan Nasional ‘Veteran’ Jakarta,” *Ekspresi Dan Persepsi J. Ilmu Komun.*, vol. 7, no. 1, pp. 232–253, 2024, doi: 10.33822/jep.v7i1.5444.
- [7] R. Nurlaely, S. D. Sartika, Kamdan, and I. L. Kharisma, “Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM),” *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 376–384, 2023, [Online]. Available: <http://ejurnal.umri.ac.id/index.php/coscitech/indexhttps://doi.org/10.37859/coscitech.v4i2.5161>
- [8] F. Nufairi, N. Pratiwi, and F. Herlando, “Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 339–348, 2024, doi: 10.29100/jupi.v9i1.4929.
- [9] B. Indra Kusuma and A. Nugroho, “Cyberbullying Detection on Twitter Uses the Support Vector Machine Method,” *J. Tek. Inform.*, vol. 5, no. 1, pp. 11–17, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.809>
- [10] Imam Riadi, Sunardi, and P. Widiandana, “Investigating Cyberbullying on WhatsApp Using Digital Forensics Research Workshop,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 730–735, 2020, doi: 10.29207/resti.v4i4.2161.