

Deteksi *Cyberbullying* Dalam Grup Telegram Menggunakan *Support Vector Machine* (SVM)

Thomas Filikano¹, Andronikus Gunawan², Andri Wijaya^{3*}

¹ Sistem Informasi, Universitas Katolik Musi Charitas, Palembang

² Sistem Informasi, Universitas Katolik Musi Charitas, Palembang

³ Sistem Informasi, Universitas Katolik Musi Charitas, Palembang

¹thomasfilikano555@gmail.com, ²andronikusgunawan26@gmail.com, ³andri_wijaya@ukmc.ac.id*

INTISARI

Cyberbullying di platform media sosial, khususnya di grup Telegram, merupakan permasalahan serius yang perlu ditangani. Penelitian ini bertujuan untuk mengembangkan pendekatan deteksi cyberbullying menggunakan metode Support Vector Machine (SVM). Sebanyak 1400 pesan dengan label cyberbullying dan non-cyberbullying digunakan sebagai data pelatihan untuk melatih model SVM. Pengujian dilakukan pada data uji sebanyak 600 pesan guna mengevaluasi kinerja model. Hasil penelitian menunjukkan bahwa penggunaan kernel linear pada SVM menghasilkan akurasi tertinggi sebesar 89%. Performa sistem klasifikasi pesan cyberbullying juga mencapai nilai precision, recall, dan F1-score masing-masing sebesar 89%, 87%, dan 88%. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam menganalisis perilaku masyarakat di dunia maya terkait pesan-pesan cyberbullying di grup Telegram.

Kata kunci: cyberbullying, deteksi cyberbullying, support vector machine, kernel linear, grup Telegram

ABSTRACT

Cyberbullying on social media platforms, particularly in Telegram groups, is a serious issue that needs attention. This research aims to develop a cyberbullying detection approach using the Support Vector Machine (SVM) method. A total of 1400 messages labeled as cyberbullying and non-cyberbullying were used as training data to train the SVM model. Testing was conducted on 600 test messages to evaluate the model's performance. The research results indicate that using a linear kernel in SVM yields the highest accuracy of 89%. The performance of the cyberbullying message classification system also achieves precision, recall, and F1-score values of 89%, 87%, and 88%, respectively. This study is expected to make a significant contribution to analyzing online community behavior related to cyberbullying messages in Telegram groups.

Keywords: cyberbullying, cyberbullying detection, support vector machine, linear kernel, Telegram group

1. PENDAHULUAN

Dalam era digital yang semakin berkembang, platform media sosial dan aplikasi pesan telah menjadi sarana utama bagi komunikasi *online*. Salah satu platform yang populer adalah Telegram, yang menyentuh angka 700 juta pengguna aktif di seluruh dunia (Simon Kemp, 2022). Meskipun Telegram menyediakan ruang berbagi informasi dan

berkomunikasi, fenomena *cyberbullying* atau pelecehan daring masih merupakan masalah yang serius.

Cyberbullying adalah fenomena yang semakin meningkat dalam lingkungan daring, dengan grup *chat* di aplikasi seperti Telegram menjadi salah satu tempat utama di mana tindakan *cyberbullying* dapat terjadi. Penanganan *cyberbullying* menjadi semakin mendesak karena dampak negatifnya terhadap kesejahteraan mental individu yang terkena dampak. Oleh karena itu, mendeteksi *cyberbullying* secara otomatis dalam grup Telegram telah menjadi tantangan yang signifikan. Dalam penelitian ini, penulis mengusulkan sebuah pendekatan menggunakan *data mining*, khususnya dengan mengimplementasikan metode *Support Vector Machine* (SVM), untuk mendeteksi adanya tindakan *cyberbullying* dalam grup Telegram.

Cyberbullying dalam konteks media sosial: *Cyberbullying*, atau perundungan dunia maya, merujuk pada tindakan perundungan yang menggunakan teknologi digital (UNICEF, 2020). Ini termasuk mengirim pesan menghina, mengancam, atau melecehkan, serta menyebarkan informasi pribadi dengan tujuan merugikan individu. Grup Telegram telah menjadi wadah untuk *cyberbullying* karena karakteristiknya yang memungkinkan anggota grup untuk berinteraksi dalam lingkungan yang relatif tertutup.

Deteksi *cyberbullying* dalam media sosial: Beberapa penelitian mendalam telah ditujukan untuk mengidentifikasi *cyberbullying* dalam berbagai platform media sosial, termasuk Twitter, Facebook, dan Instagram. Dalam literatur ini, metode-metode beragam telah diterapkan, termasuk analisis teks, pemrosesan bahasa alami, dan penerapan teknik pembelajaran mesin seperti SVM, yang terbukti efektif dalam mendeteksi *cyberbullying*.

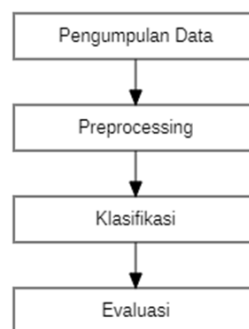
Data mining dalam deteksi *cyberbullying*: *Data mining* adalah suatu proses dimana kita secara aktif mengeksplorasi data yang ada dalam repositori dengan jumlah besar, dengan tujuan menemukan korelasi, pola, dan tren baru yang memiliki signifikansi. Untuk mencapai hal ini, digunakan teknologi pengenalan pola, serta penerapan teknik statistik dan matematika (Larose, 2005). Pada penelitian sebelumnya, *data mining* telah digunakan dalam konteks deteksi *cyberbullying*, terutama dengan mengidentifikasi kata-kata kunci dan pola perilaku tertentu yang mengindikasikan tindakan *cyberbullying*.

Library Telethon untuk pengumpulan data: Dalam penelitian ini, penulis menggunakan *library Telethon* untuk mengumpulkan data dari grup Telegram, memungkinkan akses otomatis ke pesan dan metadata dalam grup. Dengan akses ini, penulis dapat mengumpulkan sejumlah besar data yang akan digunakan sebagai fitur dalam analisis. Manfaat dari penggunaan *library* ini adalah mempermudah pembuatan program Python yang berinteraksi dengan Telegram (LonamiWebs, 2017).

Dengan mengintegrasikan pendekatan *data mining*, metode SVM, dan *library Telethon*, penelitian ini bertujuan untuk eksplorasi yang lebih mendalam terkait perilaku dan pola komunikasi yang mengindikasikan tindakan *cyberbullying* di grup Telegram. Upaya ini diharapkan dapat memberikan kontribusi positif dalam pemahaman fenomena *cyberbullying* di platform Telegram dan mendukung langkah-langkah penanggulangan masalah ini.

2. METODOLOGI

Penelitian ini dimulai dengan pengambilan data berupa pesan dari platform media sosial Telegram. Setelah itu, pesan-pesan tersebut akan diberi label sesuai dengan kategorinya. Proses identifikasi pesan melibatkan empat langkah penyelesaian, yakni pengumpulan data, *preprocessing*, klasifikasi, dan evaluasi, seperti yang tergambar pada gambar 1, dan akan dijelaskan secara lebih rinci pada sub bab berikutnya.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Dalam penelitian ini, Telegram digunakan sebagai sumber data, dan pengumpulan data dilakukan melalui penggunaan API Telegram. Penulis mendaftar sebagai pengembang untuk mendapatkan izin akses, yang mencakup API ID dan API *Hash*. Data yang diambil melibatkan pesan berbahasa Indonesia, dan jumlah total pesan yang

dikumpulkan untuk membangun model dalam penelitian ini adalah sebanyak 2.000. Format dari data yang diperoleh dapat dilihat pada Tabel 1.

Tabel 1. Format Data

ID	Sent Date	Message Content
1	2023-11-03 03:10:44+00:00	Bego banget ni orang
2	2023-11-06 19:46:34+00:00	Kenapa si orang2 jawa kek mereka pada tolol2
3	2023-11-08 06:21:37+00:00	Ya najis amat gendong " cowo berjuang sendiri lah tolol
...
1999	2023-11-12 09:48:21+00:00	Dari tadi ketemu player jawa isinya tolol semua
2000	2023-11-12 19:15:24+00:00	GTW DIRI AJG LONTE

B. Preprocessing

Dalam tahapan ini, diuraikan langkah-langkah pembersihan data berupa pesan teks dengan tujuan menghasilkan data yang bersih dan terstruktur. Beberapa langkah pra-pemrosesan yang diterapkan melibatkan penghapusan *URL*, karakter *NON-ASCII*, angka, simbol, dan tanda baca, penerapan *case folding*, penghapusan *stopwords*, penghapusan kata-kata yang hanya terdiri dari satu huruf, serta perubahan kata-kata singkatan seperti "bgt" menjadi "banget", "ajg" menjadi "anjing", dan "klo" menjadi "kalau". Selain itu, tahapan ini juga melibatkan tokenisasi untuk memecah teks menjadi satuan kata dan penerapan *stemming*.

Case folding digunakan untuk mengkonversi seluruh teks dalam dokumen menjadi bentuk yang seragam, yaitu huruf kecil. Berikut adalah contoh penerapan *case folding* pada Tabel 2.

Tabel 2. Case Folding

Teks Asli	Hasil
Kenapa si orang2 jawa kek mereka pada tolol2	kenapa si orang jawa kek mereka pada tolol
Ya najis amat gendong " cowo berjuang sendiri lah tolol	ya najis amat gendong cowo berjuang sendiri lah tolol
Dari tadi ketemu player jawa isinya tolol semua	dari tadi ketemu player jawa isinya tolol semua

Tokenisasi dilakukan untuk memecah sekumpulan karakter dalam teks menjadi satuan kata, dengan tujuan untuk mendapatkan fitur dari setiap dokumen setelah diberikan bobot nilai. Berikut adalah contoh dari tahapan tokenisasi pada Tabel 3.

Table 3. Tokenisasi

Teks	Token
kenapa si orang jawa kek mereka pada tolol	kenapa
	si
	orang
	jawa
	kek
	mereka
	pada
	tolol

Dengan serangkaian pra-pemrosesan ini, data siap untuk digunakan dalam tahap pemodelan klasifikasi kalimat *cyberbullying* menggunakan metode SVM.

C. Klasifikasi

Bagian ini merincikan proses klasifikasi pesan di dalam grup Telegram menggunakan metode SVM, sebuah algoritma pembelajaran mesin yang berguna untuk mengelompokkan data atau memprediksi nilai berdasarkan pola yang ditemukan dalam *dataset* (Müller, 2017). Pilihan penggunaan SVM didasarkan pada sejumlah penelitian sebelumnya yang menunjukkan akurasi yang sangat baik dalam mengklasifikasikan data berbentuk teks (Oryza Habibie Rahman et al, 2021; Wanda Athira Luqyana et al, 2021; Fatmawati et al, 2018; Agus Setiyono et al, 2019).

D. Evaluasi

Evaluasi model dilakukan untuk mengukur kinerja klasifikasi pesan *cyberbullying* menggunakan metode SVM. Metrik evaluasi yang digunakan melibatkan akurasi, presisi, *recall*, dan *F1-score*. Akurasi memberikan gambaran tentang sejauh mana model dapat mengklasifikasikan dengan benar, sementara presisi mengukur tingkat ketepatan model dalam mengidentifikasi pesan *cyberbullying*. *Recall* mengukur seberapa baik model dapat menangkap seluruh pesan *cyberbullying* yang ada, dan *F1-score* memberikan gabungan dari presisi dan *recall*.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil penelitian berdasarkan skenario dan langkah-langkah yang telah dijabarkan sebelumnya. Data dibagi menjadi dua kelas, yaitu data uji (*testing*) dan data latih (*training*), dengan porsi masing-masing sebesar 30% untuk data uji dan 70% untuk data latih.

Selanjutnya, Gambar 2 menyajikan hasil dari klasifikasi yang mencakup metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Metrik-metrik ini memberikan gambaran komprehensif tentang kinerja model dalam mengklasifikasikan pesan *cyberbullying*.

	precision	recall	f1-score	support
0	0.90	0.79	0.84	68
1	0.88	0.95	0.91	112
accuracy			0.89	180
macro avg	0.89	0.87	0.88	180
weighted avg	0.89	0.89	0.89	180

Gambar 2. Akurasi, presisi, *recall*, dan *F1-score*

Gambar tersebut memberikan informasi tentang sejauh mana model dapat mengklasifikasikan pesan *cyberbullying* dengan benar. Dengan akurasi sebesar 0.89 setelah pengujian pada data uji, terlihat bahwa model mampu memberikan persentase klasifikasi yang benar. Selain akurasi, presisi juga menjadi metrik penting yang mengukur ketepatan model dalam mengidentifikasi pesan *cyberbullying*. *Recall* menggambarkan sejauh mana model dapat menangkap seluruh pesan *cyberbullying* yang ada, sedangkan *F1-score* memberikan nilai gabungan dari presisi dan *recall*.

Selain itu, visualisasi hasil dapat diperkaya dengan menggunakan *Word Cloud* untuk memperlihatkan kata-kata yang sering muncul dalam pesan *cyberbullying*. *Word Cloud* merupakan representasi visual dari suatu dokumen yang dibuat dengan cara memetakan kata-kata yang paling sering muncul dalam dokumen tersebut. Representasi ini dilakukan dalam bentuk grafis dua dimensi, di mana ukuran huruf menunjukkan seberapa sering kata tersebut muncul dalam dokumen (Quim Castellà et al, 2014). Hasil visualisasi dapat dilihat pada Gambar 3 berikut.



Gambar 3. Word Cloud

Gambar tersebut menyajikan visualisasi dengan *Word Cloud* yang memperkuat analisis hasil klasifikasi. *Word Cloud* membantu mengidentifikasi kata-kata kunci yang sering muncul dalam pesan *cyberbullying*, memberikan pemahaman lebih lanjut tentang karakteristik teks yang dianggap sebagai *cyberbullying*. Dalam gambar tersebut, terlihat bahwa kata "anjing" memiliki ukuran tulisan yang lebih besar, menunjukkan bahwa kata tersebut sering muncul dalam konteks pesan *cyberbullying*.

4. KESIMPULAN

Berdasarkan hasil pengujian menggunakan metode SVM dengan *kernel linear* pada *dataset* latih sebanyak 1400 data dan *dataset* uji sebanyak 600 data, ditemukan bahwa *kernel linear* memberikan akurasi tertinggi, mencapai 89%. Kinerja sistem klasifikasi pesan *cyberbullying* menunjukkan performa yang memuaskan, dengan nilai *precision*, *recall*, dan *F1-score* untuk klasifikasi pesan *cyberbullying* masing-masing sebesar 89%, 87%, dan 88%. Penerapan *kernel linear* dalam SVM diharapkan memberikan kontribusi yang berarti dalam eksplorasi perilaku masyarakat di dunia maya. Hasil dari sistem ini dapat menjadi landasan untuk analisis lebih mendalam terkait pesan-pesan *cyberbullying* di platform *online*.

5. DAFTAR PUSTAKA

- Castella, Q., Sutton, C., 2014. Word Storms: Multiples of Word Clouds for Visual Comparison of Documents. Seoul, International Conference on World Wide Web, Vol. 1.
- Fatmawati, F., Affandes, M., 2018. Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) Pada Akun Facebook Group iRaise Helpdesk. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 3(1), 24-30.

- Kemp, S., 2022. *Top Digital Trends for 2023. DataReportal – Global Digital Insights*. Website: <https://datareportal.com/reports/looking-ahead-to-what-2023-holds>, diakses tanggal 11 November.
- Larose, D. T., 2005. *Discovering knowledge in data : an introduction to data mining*. Wiley.
- LonamiWebs., 2017. *GitHub*. Website: <https://github.com/LonamiWebs/Telethon>, diakses tanggal 11 November.
- Luqyana, W. A., Cholissodin, I., Perdana, R. S., 2018. Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704-4713.
- Müller, A. C., Guido, S., 2017. *Introduction to machine learning with Python : a guide for data scientists*. O'reilly.
- Rahman, O. H., Abdillah, G., Komarudin, A., 2021. Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17-23.
- Setiyono, A., Pardede, H. F., 2019. Klasifikasi Sms Spam Menggunakan Support Vector Machine. *Jurnal Pilar Nusa Mandiri*, 15(2), 275-280.
- UNICEF., 2020. *Cyberbullying: Apa itu dan bagaimana menghentikannya*. Website: <https://www.unicef.org/indonesia/id/child-protection/apa-itu-cyberbullying>, diakses tanggal 11 November.