Deteksi *Cyberbullying* Pada *Tweet* Berbahasa Inggris Dengan Metode *Support Vector Machine*

Cyberbullying Detection on English-Language Tweets With The Support Vector Machine Method

Resa Triyana*[1], Oddy Virgantara Putra[2], Faisal Reza Pradhana[3]
[1,2,3] *Universitas Darussalam Gontor, Ponorogo, Indonesia*
e-mail: *[1]**resa.triyana@mhs.unida.gontor.ac.id**, [2]oddy@unida.gontor.ac.id,
[3]faisalrezapradhana@unida.gontor.ac.id

Abstrak - Cyberbullying dianggap sebagai salah satu kejahatan dunia maya yang paling umum Dalam bentuk ancaman atau pelecehan elektronik, Juga dikenal sebagai intimidasi online. Dalam jeratan hukum terhadap pelaku cyberbullying, kurangnya pemahaman pengguna media sosial, khususnya di Amerika Serikat, yang mana merupakan posisi pertama dalam penggunaan Twitter membuat banyak kasus cyberbullying tidak ditanggapi dengan serius. Klasifikasi dilakukan dengan menggunakan metode Support Vector Machine dimana metode ini bertujuan mencari hyperplane pemisah antara kelas negatif dan positif. Proses dimulai dengan melakukan Preprocessing data dengan tahapan case folding, tokenization, stopwords Removal dan stemming, dan pembobotan data. Setelah data selesai dikonfigurasi, kami menggunakan metode Support Vector Machine dan menunjukkan hasil yang didapatkan dalam pengujian ini adalah nilai akurasi 90,59%, recall 92,74%, precision 95,42% dan f-measure 94,06%.

Kata kunci – Cyberbullying, Klasifikasi, Support Vector Machine, Twitter.

Abstract - Cyberbullying is considered one of the most common cybercrimes in the form of electronic threats or harassment, also known as online bullying. In legal cases against cyberbullying perpetrators, the lack of understanding of social media users, especially in the United States, which is the first position in the use of Twitter, makes many cases of cyberbullying not taken seriously. Classification is carried out using the Support Vector Machine method where this method aims to find a separating hyperplane between negative and positive classes. The process begins by preprocessing the data with the stages of case folding, tokenization, stopwords removal and stemming, and data weighting. After the data is configured, we use the Support Vector Machine method and show the results obtained in this test are 90.59% accuracy, 92.74% recall, 95.42% precision and 94.06% f-measure.

Keywords – Cyberbullying, Classification, Support Vector Machine, Twitter.

I. PENDAHULUAN

Di zaman yang telah berteknologi sekarang ini, popularitas bermedia sosial pun tak lepas dari para penggunanya. Media sosial adalah tempat dimana penggunanya melakukan interaksi dengan pengguna lain secara daring tanpa mengenal waktu dan tempat. Banyak pengguna dari media sosial bertukar opini secara daring apalagi dengan situasi yang seperti sekarang ini. Yang mengharuskan masyarakat untuk melakukan segala aktivitas secara daring, dari bekerja, memperluas relasi, hingga proses ngajar-mengajar. Menurut Kementerian Perhubungan dan Informatika (Kemenkominfo), 95% dari sekitar 63 juta pengguna internet adalah pengguna media sosial. Twitter merupakan aplikasi yang sering digunakan oleh masyarakat Indonesia. jika ada yang sedang tren, Twitter bisa menjadi acuannya. Kepala industri nasional Twitter Indonesia mengklaim bahwa Indonesia adalah salah satu negara dengan pertumbuhan tercepat pengguna aktif harian Twitter.

Seperti yang kita ketahui bahwasannya twitter merupakan sebuah aplikasi yang dapat diakses oleh siapa saja dan bersifat umum. Di Twitter, pengguna yang tidak terdaftar dapat

hanya membaca tweet dari pengguna lain, akan tetapi pengguna yang terdaftar dapat melakukan *tweet*, berbagi *tweet*, *re-tweet*, dan lain sebagainya melalui situs web atau aplikasi *smartphone*. Karena dapat diakses oleh siapa saja dan bersifat umum, maka dalam postingannya juga bercampur tentang sesuatu yang bersifat formal dan yang bersifat non-formal. Begitu pula kata-kata yang digunakan dalam postingan para pengguna twitter itu sendiri. Akan tetapi, kebayakan pengguna media sosial twitter cenderung menggunakan twitter untuk memposting postingan yang bersifat *cyberbullying* yang biasanya ditujukan kepada para *public figure* masyarakat. Popularitas media sosial yang semakin meningkat tidak lepas dari fenomena cyberbullying[1].

Disamping itu, seperti yang dilansir dari Badan Pengembangan dan Pembinaan Bahasa Kementrian Pendidikan dan Kebudayaan (Kemendikbud), bahasa inggris adalah bahasa yang masih menjadi prioritas utama ASEAN. Selain itu, menurut goodnewsfromindonesia.id per bulan Juli 2021 yang dirilis pada bulan September 2021, Amerika Serikat (AS) menjadi negara dengan pengguna Twitter terbanyak nomer 1 didunia.

Oleh karena hal tersebut, peneliti akan mengidentifikasi kata-kata yang bersifat cyberbullying pada postingan media sosial Twitter berbahasa inggris dengan metode Support Vector Machine (SVM) yang diikuti dengan teknik preprocessing data dan pembobotan kata yang menyebabkan banyaknya kasus *cyberbullying*. Untuk kemudian peneliti akan melakukan penelitian seberapa besar nilai *accuracy*, *precision*, *recall* dan *f-measure* yang diharapkan kedepannya akan mengurangi adanya tindakan *cyberbullying* yang kerap terjadi saat ini.

II. LANDASAN TEORI

2.1 Twitter

Twitter adalah sosial media yang memungkinkan penggunanya membaca pesan berbasis teks hingga 280 karakter. Pesan dalam Twitter dikenal dengan sebutan *tweet*[2]. Twitter pertama kali diluncurkan oleh mahasiswa Universitas New York Jack Dorsey selama diskusi di sebuah acara yang diselenggarakan oleh perusahaan podcast bernama Odeo. Jack Dorsey mengusulkan ide bagaimana berkomunikasi menggunakan pesan singkat. Kemudian proyek ini dikembangkan oleh Jack Dorsey, yang mana Flickr dan kode singkat SMS Amerika yang jumlahnya hanya lima digit menjadi inspirasinya yang dimulai pada tanggal 21 Maret 2006. Awalnya Twitter hanya digunakan oleh para karyawan Odeo, lalu kemudian pada akhirnya pada tanggal 5 Juli 2006, Twitter dikemukakan ke publik. Twitter memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus[3].

2.2 Sentimen Analisis

Merupakan salah satu teknik yang digunakan untuk mengekstrak data teks untuk dapat mengetahui informasi tentang sentimen bernilai negatif, positif atau netral. Biasanya teknik ini digunakan oleh pebisnis untuk memahami kemauan pelanggan, mendeteksi sentimen pada *review brand*, dan mengukur reputasi brand tersebut.

2.3 Cyberbullying

Merupakan salah satu tindakan kekerasan yang dilakukan oleh seseorang terhadap korbannya di internet, dimana korban dihina, diejek, dipermalukan dan didominasi oleh pelaku, dengan kata lain *cyberbullying* adalah salah satu bentuk perundungan melalui dunia maya atau social media. Banyak kasus *cyberbullying* yang terjadi hingga menyebabkan kematian pada korbannya.

2.4 *Text* Mining

Text Mining (juga dikenal sebagai analisis teks) menggunakan *Natural Language Processing* (NLP) untuk menormalkan teks bebas (tidak terstruktur) dalam dokumen dan database untuk analisis dan pembelajaran. Teknologi kecerdasan buatan atau *Artificial Intelligence* (AI) yang bertransformasi menjadi data terstruktur. Algoritma ini cocok untuk mengendalikan mesin.

2.5 Support *Vector Machine*

Support Vector Machine (SVM) ialah salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi. Model ini memproyeksikan vektor fitur masukan ke dalam ruang dimensi non-linier yang lebih tinggi[4]. SVM memiliki konsep yang lebih matang dan lebih jelas secara sistematis dibandingkan dengan metode-metode klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non-linear. *Support Vector Machine* (SVM) telah terbukti menjadi sangat efektif dalam kategorisasi teks tradisional[5]. SVM digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antara kelas. *Hyperplane* adalah sebuah fungsi yang dapat digunakan untuk sebagai pemisah antar kelas.

2.6 Penelitian Terdahulu

Beberapa penelitian mengenai deteksi dan klasifikasi *cyberbullying* yang telah dilakukan. Penelitian yang berkaitan dengan penelitian saat ini dipaparkan untuk memberikan gambaran dan pemahaman lebih mendalam terhadap penelitian. Beberapa tahun sebelumnya penelitian mengenai dokumen teks telah dilakukan oleh banyak peneliti termasuk dengan berbagai model algoritma yang berbeda-beda. Studi Pustaka berikut akan mengidentifikasi metode yang pernah digunakan oleh peneliti sebelumnya. Berikut penyajian pustaka terdahulu dapat dirangkum dalam Tabel 1.

Tabel 1. Penelitian Terdahulu

No	Judul	Tahun, Penulis	Metode	Hasil/Kesimpulan
1	Identifikasi *Tweet Cyberbullying* pada Aplikasi Twitter menggunakan Metode *Support Vector Machine* (SVM) *dan Information Gain* (IG) sebagai Seleksi Fitur	2018, Ni Made Gita Dwi Purnamasari, M. Ali Fauzi, Indriati dan Liana Shinta Dewi	*Support Vector Machine* dan *Information Gain*	Hasil yang didapatkan dengan metode SVM adalah accuracy 75%, precision 70,27%, recall 86,66% dan f-measure 77,61% pada percobaan nilai iterMax = 20, λ = 0,5, γ = 0,001, ε = 0,000001, dan C = 1. Nilai threshold terbaik seleksi fitur information gain adalah 90%, dengan nilai accuracy 76,66%, precision 72,22%, recall 86,66% dan f-measure 78,78%
2	Identifikasi *Cyberbullying* pada Kolom Komentar Instagram dengan metode S*upport Vector Machine* dan *Semantic Similarity*	2020, Lintani Afina Hajar Raudhoti, Anisa Hardiani dan Ade Romadhony	*Support Vector Machine* dan *Semantic Similarity*	Dalam penelitian ini, para peneliti menggunakan informasi semantik yang diturunkan dari penyisipan kata yang telah dilatih sebelumnya untuk mengumpulkan kata-kata yang serupa yang muncul dalam data pelatihan untuk menggantikan yang tidak diketahui kata-kata dalam data pengujian. Hasil percobaan penelitian ini menunjukkan bahwa penggunaan informasi kesamaan semantik meningkat akurasi klasifikasi sebesar 7%, dari 67% menjadi 74%

3	Deteksi *Cyberbullying* berdasarkan Unsur Perbuatan Pidana yang Dilanggar dengan *Naive Bayes* dan *Support Vector Machine*	2021, Tommy Nugraha Manoppo dan Dhomas Hatta Fudholi	*Naïve Bayes* dan *Support Vector Machine*	Penerapan dua model klasifikasi Naïve Bayes dan Support Vector Machine setelah re-sampling dan over-sampling dengan menggunakan metode SMOTE dapat memprediksi dengan benar kemungkinan cyberbullying berdasarkan rata-rata pelanggaran elemen dengan performance measurement adalah lebih dari 90%

III. METODE

Penelitian menggunakan data dari beberapa dataset yang ada di website seperti, GitHub dan Kaggle. Pengambilan dataset ini dilakukan dengan cara mengunduh file secara langsung dari website tersebut diatas. Dataset yang telah didapatkan kemudian di olah kembali dan mendapatkan hasil dataset yang valid dan yang dibutuhkan untuk penelitian ini. Dataset tersebut berupa *sample tweet* pada Twitter. Model algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM).

3.1 Preprocessing Data

Dokumen teks sebelum dilakukan proses klasifikasi perlu dilakukan tahapan *preprocessing.* Tujuannya untuk menghilangkan beberapa karakter dan kata yang tidak diperlukan. Tahapan *praprocessing* dalam klasifikasi bertujuan untuk meningkatkan akurasi klasifikasi data[6]. *Preprocessing* data meliputi *case folding, tokenization serta stopword removal* dan *stemming.*

a) *Case Folding*

Setiap kata pada setiap *tweet* disamaratakan format hurufnya. Penyamarataan format huruf ini dilakukan menggunakan teknik *case folding*. Teknik ini digunakan untuk menyamaratakan format huruf serta membuang komponen yang tidak diperlukan dalam setiap *tweet*. Contoh penggunaannya adalah untuk mengganti huruf kapital dengan huruf kecil agar komputer dapat dengan mudah membacanya.

b) *Tokenization*

Tokenization adalah proses pemecahan teks berupa kalimat, paragraf, atau dokumen menjadi token/bagian tertentu. Tokenisasi adalah proses untuk memotong dokumen menjadi pecahan kecil yang dapat berupa bab, sub-bab, paragraf, kalimat, dan kata (token)[7]. Pada proses ini akan menghilangkan whitespace. Dilalukannya tokenization adalah untuk memotong string input berasarkan tiap kata. Contohnya dengan menghapus imbuhan dalam setiap kata.

c) *Stopword Removal* dan *Stemming*

Selanjutnya tahap ini dilakukan untuk menghilangkan atau menghapus kata yang tidak diperlukan dan yang tidak terlalu penting dalah suatu kalimat. Mengeluarkan dari karakter html bertujuan untuk menghapus tautan URL dan juga karakter html yang sering ditemukan di tweet. Penghapusan tanda baca digunakan untuk menghapus karakter khusus yang sering ditemukan dalam tweet seperti hastag (#), @user, retweet (RT)[8]. Tabel 2 menunjukkan contoh output tweet yang telah melewati tahap preprocessing.

Tabel 2. Setelah tahapan *Preprocessing* dilakukan

Data Input	Data Output
1. !!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. & as a man you should always take the trash out...	1. mayasolovely as a woman you shouldn t complain about cleaning up your house amp as a man you should always take the trash out
2. !!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya 	2. shenikarobe s the shit you hear about me might be true or it might be faker than the bitch who told it to ya 57361

Proses eksekusi menggunakan *Google Collaboratory*, dengan perangkat keras (hardware) laptop berkapasitas RAM 4 GB dan processor intel inside Core i7. Selanjutnya data yang sudah bersih dilakukan analisis menggunakan model *Support Vector Machine*.

3.2 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, yang bertujuan memberi label sentimen pada tweet[9].

3.3 *TFIDF*

Pembobotan *TFIDF* merupakan perhitungan statistik untuk mengambil fitur kata yang penting pada suatu dokumen. Tahapan *preprocessing* menghasilkan data yang siap untuk diolah pada proses pembobotan setiap kata (term)[1].

IV. HASIL DAN PEMBAHASAN

Model yang telah diimplementasikan tentu mendapatkan hasil klasifikasi antara positif dan negatif. Dalam penelitian ini menggunakan algoritma *Support Vector Machine* untuk klasifikasi. Pengukuran performa model menggunakan *accuracy*, *recall*, *precision* dan *F1-score*. Hal ini dikarenakan pada penelitian ini hanya mengidentifikasi *tweet* yang mengandung *cyberbullying*, sehingga *retrieve* hasil identifikasi dibutuhkan untuk mengetahui apakagh informasi yang diminta oleh pengguna sudah sesuai dengan informasi yang diberikan sistem[10].

a) *Support Vector Machine*

Data masukan yang digunakan dalam penelitian ini adalah *tweet* dari postingan Twitter yang telah ditentukan. Data *input* yaitu 24783 data *tweet* yang akan digunakan, Dimana terdiri dari, 22304 data *tweet* sebagai data latih, dan 2479 data *tweet* menjadi data uji[7]. Pada bagian ini diperlihatkan hasil pengujian yang telah dilakukan beserta analisis terhadap nilai hasil yang telah didapatkan dari proses pengujian[3]. Model *Support Vector Machine* merupakan kategori model yang menggunakan regresi atau data klasifikasi dengan metode *supervised learning*. Sehingga data dilatih dahulu dan baru akan diuji setelah dilatih. Nilai hasil evaluasi menggunakan model SVM ditunjukkan pada Tabel 3.

Tabel 3. Evaluasi Model SVM

Measures	SVM
Accuracy	90,59
Recall	92,74
Precision	95,42
F1-Score	94,06

Pengujian perhitungan dilakukan menggunakan pengujian confussion matrix. Pengujian confussion matrix dilakukan untuk memeriksa apakah aplikasi telah berjalan dengan baik dan benar sesuai dengan tujuan yang diharapkan[11]. *Support Vector Machine* memiliki

hasil akurasi 90,59%. Model *Support Vector Machine* meraih nilai *recall* 92,74%, *precision* 95,42% dan nilai f-measure 94.06%.

V. KESIMPULAN

Dalam penelitian ini, 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 sekitar 24783 data *tweet*, Dimana terdiri dari 22304 data *tweet* sebagai data latih, dan 2479 data *tweet* sebagai data uji. Setelah data selesai dikonfigurasi, kami menggunakan metode *Support Vector Machine* dan menunjukkan hasil yang didapatkan dalam pengujian ini adalah nilai akurasi 90,59%, *recall* 92,74%, *precision* 95,42% dan *f-measure* 94,06%. Hasil penelitian ini menunjukkan bahwa metode *Support Vector Machine* memberikan kinerja yang baik dalam mendeteksi *cyberbullying*. Selanjutnya untuk mengembangkan penelitian dapat dilakukan dengan permodelan lain dan menggunakan konsep *deep learning*.

DAFTAR PUSTAKA

[1] L. Afina, H. Raudhoti, A. Herdiani, Romadhony, and Ade, "Identifikasi Cyberbullying pada Kolom Komentar Instagram dengan Metode Support Vector Machine dan Semantic Similarity (Cyberbullying Identification on Instagram Comment Using Support Vector Machine and Semantic Similarity)," vol. 4, no. 1, pp. 1–8, 2020, [Online]. Available: http://jcosine.if.unram.ac.id/.

[2] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," *HP Lab. Tech. Rep.*, no. 89, 2011.

[3] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.

[4] A. Geet *et al.*, "Classification of Hate Speech Using Deep Neural Networks To cite this version : HAL Id : hal-03101938 Classification of Hate Speech Using Deep Neural Networks," 2021.

[5] S. Almutiry and M. Abdel Fattah, "Arabic CyberBullying Detection Using Arabic Sentiment Analysis," *Egypt. J. Lang. Eng.*, vol. 8, no. 1, pp. 39–50, 2021, doi: 10.21608/ejle.2021.50240.1017.

[6] T. Kurniawan, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine," *IT J.*, vol. 23, p. 1, 2017.

[7] R. M. Kamal and E. Rainarli, "Analisis Sentimen Cyberbullying Pada Komentar Facebook Dengan Metode Klasifikasi Support Vector Machine," *Univ. Komput. Indones.*, 2019.

[8] J. Patihullah and E. Winarko, "Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 43, 2019, doi: 10.22146/ijccs.40125.

[9] S. J. Lewis, "Thumbs up," *Am. J. Orthod. Oral Surg.*, vol. 31, no. 9, pp. 481–482, 1945, doi: 10.1016/0096-6347(45)90048-2.

[10] N. M. G. D. Purnamasari, M. A. Fauzi, Indriarti, and L. S. Dewi, "Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5326–5332, 2018.

[11] A. S. Hutagalung, A. B. P. Negara, and E. E. Pratama, "Aplikasi Pendeteksi Cyberbullying Terhadap Komentar Postingan Media Sosial Instagram dengan Metode Naïve Bayes ClassifierBerbasis Website," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. XI, no. 3, pp. 364–371, 2021, doi: 10.26418/justin.v9i3.44843.