

CYBERBULLYING DETECTION ON TWITTER USES THE SUPPORT VECTOR MACHINE METHOD

Bayu Indra Kusuma^{*1}, Aryo Nugroho²

^{1,2}Information Systems, Faculty of Computer Science, Universitas Narotama, Indonesia
Email: ¹bayu.indra@narotama.ac.id, ²aryo.nugroho@narotama.ac.id

(Article received: January 05, 2023; Revision: February 03, 2023; published: January 31, 2024)

Abstract

Social media is a platform that provides facilities for users to engage in various social activities. However, the increasing popularity of social media in the modern era also cannot be separated from the occurrence of several negative impacts, one of which is cyberbullying. Cyberbullying is an action that is done online that can harm the mental and emotional condition of an individual. To reduce this problem, this research aims to investigate the performance of the C-SVC and Nu-SVC algorithms from the Support Vector Machine method in classifying cyberbullying sentences. The data used is comments data from the @puanmaharani_ri account on Twitter, which was collected from September 25, 2020, to September 29, 2022, totaling 5,000 data. After the data is collected, it is labeled and preprocessed, and then the data will be weighted using the TF-IDF method. The result of the TF-IDF will be displayed in the form of a word cloud. Next, the Support Vector Machine method will classify cyberbullying sentences using several percentages split combinations such as 60%, 70%, 80%, and 90%. The test results show that the C-SVC method has the highest accuracy of 79.6% at a 70% percentage split, while Nu-SVC has the highest accuracy of 78.9% at a 60% percentage split. From these results, it can be concluded that the Support Vector Machine method with the C-SVC algorithm provides better results than Nu-SVC in classifying cyberbullying sentences.

Keywords: C-SVC, cyberbullying, data mining, Nu-SVC, support vector machine.

DETEKSI CYBERBULLYING PADA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Abstrak

Media sosial adalah platform yang menyediakan fasilitas bagi pengguna untuk melakukan berbagai aktivitas sosial. Namun, penggunaan media sosial yang semakin populer di era modern ini tidak terlepas dari beberapa dampak negatif yang terjadi, salah satunya adalah cyberbullying. Cyberbullying merupakan tindakan yang dilakukan secara online yang dapat merugikan kondisi jiwa dan kesehatan mental seseorang. Untuk mengurangi masalah ini, penelitian ini bertujuan untuk mengetahui performa algoritma C-SVC dan Nu-SVC dari metode Support Vector Machine dalam melakukan klasifikasi kalimat cyberbullying. Data yang digunakan adalah data komentar dari postingan akun @puanmaharani_ri di Twitter, yang dikumpulkan pada rentang waktu 25 september 2020 hingga 29 September 2022 sebanyak 5000 data. Setelah data dikumpulkan, dilakukan pelabelan dan preprocessing, kemudian data akan dilakukan pembobotan kata menggunakan metode TF-IDF. Hasil dari TF-IDF akan ditampilkan dalam bentuk word cloud. Selanjutnya, metode Support Vector Machine akan digunakan untuk melakukan klasifikasi kalimat cyberbullying dengan menggunakan beberapa kombinasi percentage split seperti 60%, 70%, 80%, dan 90%. Hasil pengujian menunjukkan bahwa metode C-SVC menunjukkan akurasi tertinggi sebesar 79.6% pada percentage split 70%, sedangkan Nu-SVC menunjukkan akurasi tertinggi sebesar 78.9% pada percentage split 60%. Dari hasil tersebut, dapat disimpulkan bahwa metode Support Vector Machine dengan algoritma C-SVC memberikan hasil yang lebih baik daripada Nu-SVC dalam melakukan klasifikasi kalimat cyberbullying.

Kata kunci: C-SVC, cyberbullying, data mining, Nu-SVC, support vector machine.

1. PENDAHULUAN

Media sosial merupakan salah satu platform yang menyediakan fasilitas untuk melakukan

berbagai aktivitas sosial bagi penggunanya. Pengguna media sosial dapat melakukan komunikasi dan menyampaikan informasi melalui tulisan, foto,

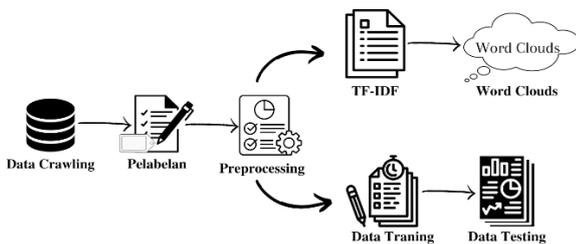
atau video. Meskipun media sosial memiliki banyak keuntungan, namun penggunaannya dapat menimbulkan berbagai permasalahan, salah satunya adalah perundungan di dunia maya atau yang sering disebut *cyberbullying*.

Cyberbullying merupakan salah satu dampak dari kebebasan berinteraksi di media sosial [1], yang dilakukan dengan menggunakan media teks dan visual sebagai sarana komunikasi. Penelitian terkait *cyberbullying* di media sosial Twitter menjadi perhatian khusus dalam studi *data mining*. Karena dampak dari *cyberbullying* dapat menimbulkan perasaan kekecewaan, sedih, depresi, dan merasa tidak berharga sehingga korban akan menjauh dari lingkungannya karena merasa kurang percaya diri [2]. Selain itu pada penelitian yang dilakukan oleh [3] menyatakan bahwa korban *cyberbullying* memiliki risiko lebih tinggi untuk melakukan bunuh diri dibandingkan dengan mereka yang mengalami *bullying* di dunia nyata.

Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen terhadap Puan Maharani pada media sosial Twitter, dengan membedakan apakah teks komentar tersebut termasuk dalam kategori *cyberbullying* atau tidak. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi tingkat akurasi dari kedua metode pada *Support Vector Machine*, yaitu Nu-SVC dan C-SVC, dalam mendeteksi kalimat *cyberbullying* pada media sosial Twitter.

2. METODE PENELITIAN

Pada bab metode penelitian dilakukan beberapa tahapan untuk menyelesaikan penelitian “Deteksi *Cyberbullying* Pada Twitter Menggunakan Metode *Support Vector Machine*” yang diilustrasikan pada gambar dibawah ini.



Gambar 1. Metode Penelitian

Berdasarkan gambar 1, penelitian ini dimulai dari proses pengambilan data dengan cara *crawling* komentar pada media sosial Twitter. Selanjutnya data yang diperoleh masuk pada tahap pelabelan yang dilakukan secara manual. Berikutnya dari hasil pelabelan dilakukan *preprocessing* antara lain yaitu tahap *casefolding*, *cleansing*, *tokenization*, *stopword removal*. Selanjutnya, setiap kata dari dokumen akan dilakukan pembobotan kata menggunakan TF-IDF dan dapat di visualisasikan menggunakan *Word*

Cloud. Dataset yang sudah melalui tahap *preprocessing* digunakan untuk tahap *training* dan *testing* dengan menggunakan metode *Support Vector Machine* dengan algoritme C-SVC dan Nu-SVC untuk menentukan model yang paling akurat dalam mendeteksi kalimat *cyberbullying* pada Twitter. Dalam penelitian ini, alat yang digunakan dalam menjalankan program adalah Google Colaboratory.

2.1. Data Crawling

Pada penelitian ini data diambil menggunakan metode *crawling* memakai *library* Tweepy yang memanfaatkan API Twitter. Metode ini memungkinkan untuk mengambil data berupa komentar Twitter pada postingan yang akan digunakan sebagai sumber data. Data yang dikumpulkan adalah data komentar dari postingan akun @puanmaharani_ri. Pengumpulan data dilakukan pada rentang waktu 25 september 2022 sampai 29 september 2022.

2.2. Pelabelan

Pelabelan data dilakukan secara manual, dengan memberikan label *cyberbullying* atau *non-cyberbullying* pada setiap komentar atau kalimat. Label *cyberbullying* mengindikasikan bahwa kalimat tersebut negatif dan mengandung unsur perundungan, sedangkan label *non-cyberbullying* menunjukkan bahwa kalimat tersebut positif dan tidak mengandung unsur perundungan.

2.3. Preprocessing

Data komentar yang didapatkan dari hasil *crawling* pada media sosial Twitter masih belum bisa digunakan. Data komentar tersebut belum bisa digunakan karena masih dalam bentuk kalimat yang tidak terstruktur. Oleh karena itu diperlukan adanya perubahan data menjadi data terstruktur yang memiliki nilai numerik [4]. Tahap ini disebut *Preprocessing* dan meliputi beberapa tahapan diantaranya *casefolding*, *cleansing*, *tokenization*, dan *stopword removal*.

2.3.1. Casefolding

Dalam menulis kalimat, biasanya terdapat bentuk huruf yang berbeda terkait keseragaman huruf kapital atau bukan, tahap ini adalah proses keseragaman huruf [5]. *Case folding* merupakan proses penghilangan perbedaan huruf besar dan kecil pada suatu kalimat, sehingga semua karakter dianggap sama. Misalnya, kalimat "Saya Sedang Makan Sate" akan diubah menjadi "saya sedang makan sate" setelah proses *case folding*.

2.3.2. Cleansing

Pada tahap ini adalah proses yang dilakukan untuk membersihkan atribut yang tidak diperlukan dalam input data seperti simbol dan tanda baca [6].

Tujuannya adalah untuk meningkatkan kualitas data sehingga lebih berguna dalam analisis dan pengambilan keputusan.

2.3.3. Tokenization

Pada tahap ini dilakukan proses untuk pemisahan kata dalam setiap dokumen menjadi sebuah token [7]. Setiap bagian yang dipecah disebut token. Contohnya, kalimat "saya sedang makan sate" akan dipecah menjadi beberapa token seperti "saya", "sedang", "makan", dan "sate". Token ini kemudian digunakan dalam proses TF-IDF untuk melakukan pembobotan kata. Dengan *tokenization*, data dapat diolah dengan lebih mudah dan efisien.

2.3.4. Stopword Removal

Stopword Removal merupakan suatu proses untuk menghapus kata-kata yang dianggap tidak memiliki makna seperti contoh kata 'di', 'yang', 'ke', dan semua kata yang terdapat dalam kamus *stopword* [8]. *Stopword removal* berguna untuk mempercepat pemrosesan teks dan mengurangi ukuran dokumen, sehingga memudahkan pencarian dan analisis teks.

2.4. TF-IDF

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

2.5. Word Cloud

Word cloud merupakan alat yang digunakan dalam memvisualisasikan data teks. *Word cloud* merupakan representasi dari kata-kata yang paling sering muncul di tweet, dengan kata yang sering digunakan tampak lebih besar dalam *word cloud* [10]. *Word cloud* biasanya digunakan untuk Menampilkan data teks yang cukup besar sehingga memudahkan pembaca untuk mengetahui kata-kata yang paling sering muncul.

2.6. Support Vector Machine

Support Vector Machine merupakan suatu metode yang dapat menyelesaikan masalah linear menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi [11]. *Support Vector Machine* mencoba mengoptimalkan batas pemisah antara dua kelas data dengan memaksimalkan jarak antara data dan batas pemisah tersebut. *Support Vector Machine* dapat digunakan untuk klasifikasi biner maupun multi-kelas. Metode ini biasanya digunakan untuk mengatasi masalah klasifikasi yang memiliki data linier yang tidak terpisah secara jelas. Selain itu, *Support Vector Machine* juga dapat digunakan untuk mengatasi masalah klasifikasi non-

linier dengan menggunakan kernel untuk mengubah data ke dalam bentuk yang dapat dipisahkan secara linier.

Metode *Support Vector Machine* memiliki beberapa algoritma yang dapat digunakan, di antaranya adalah C-SVC dan Nu-SVC. C-SVC adalah algoritma *Support Vector Machine* yang digunakan untuk klasifikasi dengan menggunakan parameter C sebagai parameter regularisasinya [12]. Sedangkan Nu-SVC adalah algoritma *Support Vector Machine* yang digunakan untuk melakukan klasifikasi dengan menggunakan parameter nu sebagai parameter regularisasinya [12]. Nu-SVC digunakan untuk mengatur jumlah *support vectors* yang diinginkan. Kedua implementasi ini dapat digunakan untuk mengoptimalkan akurasi dari hasil klasifikasi yang dilakukan oleh *Support Vector Machine*.

3. HASIL DAN PEMBAHASAN

Penelitian yang telah dilakukan pada tahap ini menggunakan metode *Support Vector Machine* dengan menggunakan algoritma C-SVC dan Nu-SVC. Data yang digunakan merupakan data yang telah melalui tahap *preprocessing*. Pada proses *training data* yang digunakan merupakan data *Use Training Set* yang bertujuan untuk mengetahui performa dari model yang digunakan. Lalu yang terakhir melakukan pengujian dengan beberapa kombinasi *percentage split* untuk mengetahui performa dari model dalam melakukan klasifikasi terhadap komentar *cyberbullying* pada Twitter.

3.1. Hasil Crawling

Tabel 1 menampilkan beberapa contoh dari hasil *crawling* data menggunakan library Tweepy yang ditulis dalam bahasa pemrograman Python. Total jumlah data yang didapatkan melalui *crawling* adalah 5000 data. Library Tweepy merupakan salah satu library Python yang digunakan untuk mengakses API Twitter dan melakukan *crawling* data dari media sosial tersebut.

Tabel 1. Hasil Crawling

User	Comment
AsepSul46335267	@puanmaharani_ri Wong cilik yang mba...
pengushrminari	@puanmaharani_ri Aku wong cilik tapi pake sepatu bu
Kontrol893	@puanmaharani_ri Ahahaha. Kontlo. Niatan mau merakyat malah gak ngerti mulai dari mana
Hajir11336579	@puanmaharani_ri Mantaab mbak puan..maju trus..smoga jadi presiden.
abon_kang	@puanmaharani_ri Wes tah puan, rekam jejakmu ora genah blas ðŸ˜Ž

3.2. Hasil Pelabelan

Pada tabel 2 menunjukkan hasil pelabelan yang telah dilakukan dengan pelabelan manual. Label yang

digunakan ada dua label, yaitu *cyberbullying* dan *non-cyberbullying*.

Tabel 2. Hasil Pelabelan

User	Comment	Label
AsepSul46335267	@puanmaharani _ri Wong cilik yang mba...	NON-CYBERBULLYI NG
pengushrnminari	@puanmaharani _ri Aku wong cilik tapi pake sepatu bu	NON-CYBERBULLYI NG
Kontol893	@puanmaharani _ri Aahaha. Kontlo. Niatan mau merakyat malah gak ngerti mulai dari mana	CYBERBULLYI NG
Hajir11336579	@puanmaharani _ri Mantaab mbak puan..maju trus..smoga jadi presiden.	NON-CYBERBULLYI NG
abon_kang	@puanmaharani _ri Wes tah puan, rekam jejakmu ora genah blas ðŸ˜Ž	CYBERBULLYI NG

3.3. Hasil Preprocessing

Tabel 3 menampilkan hasil dari tahap *preprocessing* yang telah dilakukan. Tahap *preprocessing* ini meliputi beberapa langkah, yaitu *casefolding*, *cleansing*, *tokenization*, dan *stopword removal*. Langkah-langkah ini dilakukan untuk mengubah data mentah menjadi data yang siap digunakan dalam analisis. Hasil dari tahap ini akan digunakan sebagai input untuk tahap selanjutnya, yaitu tahap *training*.

Tabel 3. Hasil Preprocessing

Comment	Label
['cilik', 'mba']	NON-CYBERBULLYING
['cilik', 'pake', 'sepatu', 'bu']	NON-CYBERBULLYING
['ahaha', 'kontlo', 'niatan', 'merakyat', 'gak', 'ngerti']	CYBERBULLYING
['mantaab', 'mbak', 'puan', 'maju', 'trus', 'smoga', 'presiden']	NON-CYBERBULLYING
['wes', 'tah', 'puan', 'rekam', 'jejakmu', 'ora', 'genah', 'blas']	CYBERBULLYING

3.4. Hasil Visualisasi Word Cloud

Hasil visualisasi *word cloud* menunjukkan kata-kata *cyberbullying* dan *non-cyberbullying* yang sering digunakan dalam berkomentar pada media sosial Twitter. Ukuran kata pada visualisasi *word cloud* memiliki makna seberapa sering kata tersebut digunakan dalam berkomentar, yang artinya pengguna Twitter sering menggunakan kata tersebut sebagai topik percakapan dalam memberikan komentar.



Gambar 2. Visualisasi Word Cloud Non Cyberbullying

Berdasarkan gambar 2 hasil dari visualisasi *word cloud* yang menunjukkan kalimat *non-cyberbullying* dan bernada positif serta merupakan dukungan dari pengguna media sosial Twitter, kata-kata yang ditampilkan dengan ukuran yang besar dapat dijadikan sebagai contoh kalimat yang mencerminkan karakteristik tersebut, misalnya "mantaab mbak puan, maju terus". Kalimat ini tidak mengandung *cyberbullying* dan memiliki nada yang positif serta merupakan dukungan dari pengguna media sosial Twitter.



Gambar 3. Visualisasi Word Cloud Cyberbullying

Visualisasi *word cloud* dari kalimat *cyberbullying* pada gambar 3 menunjukkan bahwa terdapat kalimat yang bernada negatif dan berisi *cyberbullying* terhadap Puan Maharani. Dari kata-kata yang berukuran besar, dapat dibuat suatu kalimat sebagai contoh, seperti "niatan pura-pura merakyat". Visualisasi *Word cloud* ini dapat menjadi bahan evaluasi bagi Puan Maharani dalam melakukan kampanye.

3.5. Hasil Training Data

Setelah *crawling* dan *preprocessing*, data yang telah diperoleh digunakan dalam proses pelatihan menggunakan metode *support vector machine* dengan algoritma C-SVC dan Nu-SVC. Data *training set* yang digunakan sebanyak 5000 data. Pada tahap ini, metode ini dilatih untuk dapat mengklasifikasikan

kalimat menjadi *cyberbullying* atau *non-cyberbullying*. Hasil dari proses pelatihan ini dapat dilihat pada tabel di bawah. Pada tabel ini ditampilkan informasi mengenai akurasi dari masing-masing metode yang digunakan dalam proses pelatihan.

Tabel 4. Hasil Akurasi C-SVC Data Training

Metode	Akurasi
C-SVC	93.2%

Tabel 4 menunjukkan hasil dari proses pelatihan menggunakan metode *Support Vector Machine* dengan algoritma C-SVC. Hasil akurasi yang diperoleh menunjukkan bahwa ketika menggunakan data *training set*, model menghasilkan akurasi sebesar 93.2%. Ini menunjukkan bahwa model tersebut cukup efektif dalam memprediksi kalimat *cyberbullying*.

Tabel 5. Hasil Akurasi Nu-SVC Data Training

nu	Akurasi
0.005	73.94%
0.01	78.64%
0.05	99.12%
0.1	99.08%
0.15	99.06%
0.2	99.04%
0.25	98.92%
0.3	98.92%
0.35	98.80%
0.4	98.42%

Tabel 5 menunjukkan hasil akurasi dari Nu-SVC yang menggunakan data *training set* dengan beberapa parameter. Hasil dari penerapan Nu-SVC pada data *training set* menunjukkan bahwa parameter nu 0.05 memiliki akurasi terbaik, yaitu 99.12%. Nilai akurasi yang tinggi ini menunjukkan bahwa parameter tersebut dapat melakukan klasifikasi dengan baik. Dengan demikian, pada tahap data testing, parameter nu 0.05 akan digunakan untuk mengetahui kemampuannya dalam melakukan klasifikasi pada data baru. Selain parameter nu 0.05, terdapat beberapa parameter lain yang juga memiliki akurasi yang tinggi, seperti nu 0.1 dan nu 0.15. Namun, karena nilai akurasinya lebih rendah daripada parameter nu 0.05, maka parameter tersebut tidak dipilih untuk digunakan dalam tahap *testing data*.

3.6. Hasil Testing Data

Pengujian data dilakukan dengan membagi data menjadi beberapa bagian dan menganalisis masing-masing bagian untuk mengetahui seberapa akurat hasilnya. Hasil dari pengujian data ditunjukkan dalam tabel di bawah ini.

Tabel 6. Hasil Akurasi C-SVC *Percentage Split*

Percentage Split	Akurasi
60%	79.4%
70%	79.6%
80%	78.8%
90%	79%

Pada Tabel 6 menampilkan hasil dari beberapa kombinasi *percentage split* yang telah dilakukan diantaranya adalah *Percentage Split* 60% dengan akurasi 79.4%, 70% dengan akurasi 79.6%, 80% dengan akurasi 78.8%, dan 90% dengan akurasi 79%.

Tabel 7. Hasil Akurasi Nu-SVC *Percentage Split*

Percentage Split	Akurasi
60%	78.9%
70%	78.5%
80%	78.2%
90%	78%

Tabel 7 menunjukkan hasil akurasi Nu-SVC menggunakan parameter nu 0.05 dengan berbagai *percentage split*. Hasil terbaik didapatkan pada *percentage split* 60% dengan akurasi sebesar 78.9%, sedangkan hasil terendah terjadi pada *percentage split* 90% dengan akurasi 78%.

4. DISKUSI

Penelitian sebelumnya telah menunjukkan bahwa metode *Support Vector Machine* dengan *Poly Kernel* memiliki akurasi yang paling optimal dalam klasifikasi *cyberbullying* pada pesan, dengan nilai 97.11% [13]. Namun, penelitian lain yang dilakukan oleh [14] menunjukkan bahwa metode *Support Vector Machine* secara keseluruhan memiliki akurasi yang lebih tinggi daripada metode *Naive Bayes* dalam melakukan klasifikasi kalimat *cyberbullying*. Data yang digunakan dalam penelitian ini diperoleh dari berbagai sumber seperti Kaggle dan Github. Hasil akurasi yang didapatkan dari penelitian ini menunjukkan bahwa metode *Support Vector Machine* memiliki nilai 71.25%, sedangkan *Naive Bayes* memiliki nilai 52.70%.

Pada penelitian lainnya metode *Support Vector Machine* juga dinilai memiliki akurasi yang lebih baik dibandingkan *Naive Bayes*, Hasil dari penelitian tersebut menunjukkan bahwa *Support Vector Machine* mendapatkan akurasi sebesar 92.02% sedangkan *Naive Bayes* mendapatkan akurasi sebesar 81.01% [15].

Penelitian terkait lainnya yang dilakukan oleh [16] yang membandingkan metode *K-Nearest Neighbor*, *Naive Bayes Classifier* dan *Support Vector Machine* dengan menggunakan data yang berasal dari teks obrolan pada grup Whatsapp "Pascasarjana 2017". Hasil pengujian menunjukkan bahwa *K-Nearest Neighbor*, *Naive Bayes Classifier*, dan *Support Vector Machine* memiliki akurasi masing-masing sebesar 81.32%, 78.95%, dan 81.58%.

Hasil penelitian yang dilakukan oleh [17] menunjukkan bahwa metode *support vector machine* memiliki tingkat akurasi sebesar 79,412% dalam mendeteksi kalimat *cyberbullying* pada 1053 data komentar yang diambil dari akun selebgram di Indonesia. Data tersebut diolah menggunakan bahasa pemrograman R dengan menggunakan Rstudio.

Penelitian sebelumnya menunjukkan bahwa metode *Support Vector Machine* memiliki hasil

akurasi yang tinggi dalam melakukan klasifikasi *data mining*. Namun, belum ada penelitian yang membandingkan penggunaan C-SVC dan Nu-SVC dalam mengidentifikasi kalimat *cyberbullying* pada media sosial Twitter. Hal ini menjadi penting untuk diketahui karena dapat membantu dalam meningkatkan efektivitas deteksi kalimat *cyberbullying*. Penggunaan C-SVC dan Nu-SVC dapat memberikan informasi yang lebih akurat dan dapat menghindari kemungkinan kesalahan dalam deteksi kalimat *cyberbullying*. Oleh karena itu, pada penelitian ini dilakukan dengan tujuan untuk mengetahui seberapa efektif kedua metode tersebut dalam melakukan deteksi kalimat *cyberbullying* pada media sosial Twitter.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diketahui bahwa pada tahap *training*, C-SVC mendapatkan akurasi 93.2%, sedangkan Nu-SVC mendapatkan akurasi tertinggi 99.12% dengan parameter nu 0.05 dan akurasi terendah 73.94% dengan parameter nu 0.005. Pada tahap pengujian dengan beberapa kombinasi *percentage split*, C-SVC memiliki akurasi terbaik di angka 79.6% dengan *percentage split* 70%. Sedangkan Nu-SVC dengan parameter nu 0.05 mendapatkan akurasi tertinggi di angka 78.9% dengan *percentage split* 60%. Dari hasil ini, dapat disimpulkan bahwa metode *Support Vector Machine* yang menggunakan algoritma C-SVC memiliki akurasi yang lebih baik daripada Nu-SVC dalam melakukan klasifikasi kalimat *cyberbullying* pada Twitter.

Pada penelitian ini, data yang digunakan hanya berasal dari satu sumber media sosial, yaitu Twitter dengan dataset sebesar 5000 data. Pada penelitian selanjutnya, dapat dipertimbangkan penggunaan dataset yang lebih besar dari berbagai sumber data yang berbeda untuk mendapatkan hasil yang lebih akurat dan dapat dijadikan sebagai rujukan yang lebih baik.

DAFTAR PUSTAKA

- [1] D. Primasti and S. I. Dewi, "Pengaruh media sosial terhadap penyimpangan perilaku remaja (*cyberbullying*)," *Reformasi*, vol. 7, no. 2, 2018.
- [2] R. Syah and I. Hermawati, "Upaya Pencegahan Kasus *Cyberbullying* bagi Remaja Pengguna Media Sosial di Indonesia," *J. Penelit. Kesejaht. Sos.*, vol. 17, no. 2, Art. no. 2, Jun. 2018, doi: 10.31105/jpks.v17i2.1473.
- [3] Z. Peng *et al.*, "Associations between Chinese adolescents subjected to traditional and cyber bullying and suicidal ideation, self-harm and suicide attempts," *BMC Psychiatry*, vol. 19, no. 1, p. 324, Dec. 2019, doi: 10.1186/s12888-019-2319-9.
- [4] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, Sep. 2019, doi: 10.30864/eksplora.v9i1.254.
- [5] M. Syarifuddin, "Analisis sentimen opini publik terhadap efek PSBB pada twitter dengan algoritma decision tree, knn, dan naïve bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020.
- [6] A. Amalia, D. Gunawan, Y. Fithri, and I. Aulia, "Automated Bahasa Indonesia essay evaluation with latent semantic analysis," presented at the Journal of Physics: Conference Series, 2019, vol. 1235, no. 1, p. 012100.
- [7] Z. P. Putra and A. Nugroho, "Pebandingan Performa Naïve Bayes dan KNN pada Klasifikasi Teks Sentimen Jasa Ekspedisi," *JOINTECS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 3, pp. 145–152, 2021.
- [8] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *INSYST J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019.
- [9] A. Thakkar and K. Chaudhari, "Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks," *Appl. Soft Comput.*, vol. 96, p. 106684, Nov. 2020, doi: 10.1016/j.asoc.2020.106684.
- [10] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, Jan. 2018, pp. 1259–1262. doi: 10.1109/ICISC.2018.8399007.
- [11] A. M. Puspitasari, D. E. Ratnawati, and A. W. Widodo, "Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector Machine," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput. E-ISSN*, vol. 2548, p. 964X, 2018.
- [12] V. R. Oktavia, "Aplikasi Deteksi Kejadian di Jalan Raya berdasarkan Data Twitter Menggunakan Metode Support Vector Machine," 2018.
- [13] S. I. Noviantho and L. Ashianti, "Cyberbullying classification using text mining," in *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, 2017, pp. 241–246.
- [14] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, "Detecting A Twitter Cyberbullying

- Using Machine Learning,” in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, May 2020, pp. 297–301. doi: 10.1109/ICICCS48265.2020.9120893.
- [15] J. O. Atoum, “Cyberbullying Detection Through Sentiment Analysis,” in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, Dec. 2020, pp. 292–297. doi: 10.1109/CSCI51800.2020.00056.
- [16] I. Saputra and D. Rosiyadi, “Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp,” *Fakt. Exacta*, vol. 12, no. 2, p. 101, Jul. 2019, doi: 10.30998/faktorexacta.v12i2.4181.
- [17] M. Andriansyah *et al.*, “Cyberbullying comment classification on Indonesian Selebgram using support vector machine method,” in *2017 Second International Conference on Informatics and Computing (ICIC)*, Jayapura, Nov. 2017, pp. 1–5. doi: 10.1109/IAC.2017.8280617..