

ANALISIS SENTIMEN CYBERBULLYING PADA SOSIAL MEDIA INSTAGRAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Damar Nugraha¹, Puji Astuti^{2,*}

¹Program Studi Informatika; Universitas Nusa Mandiri; Alamat, telp/fax dari Institusi; e-mail: damarnugraha96@gmail.com

^{2,*}Program Studi Informatika; Universitas Nusa Mandiri; Alamat, telp/fax dari Institusi; e-mail: puji.pat@nusamandiri.ac.id

* Korespondensi: e-mail: puji.pat@nusamandiri.ac.id

Diterima: 15 Agustus 2023; Review: 11 November 2023; Disetujui: 21 November 2023

Cara sitasi: Nugraha D, Astuti P. 2023. Analisis Sentimen *Cyberbullying Pada Sosial Media Instagram* Menggunakan Metode *Support Vector Machine*. Information System for Educators and Professionals. Vol 8(2): 153-164.

Abstrak: Media sosial termasuk suatu *platform* digital yang mengizinkan pemakainya untuk berinteraksi, bertukar informasi, serta membuat konten yang dapat diakses dan dibagikan oleh pengguna lainnya melalui internet. Pengguna media sosial dapat berhubungan dengan orang-orang di seluruh dunia dan membangun jaringan sosial online. Banyak media sosial yang terkenal misalnya *Facebook*, *Instagram*, *Twitter*, *LinkedIn*, *TikTok*, dan *YouTube*. Saat ini, media sosial punya efek banyak bagi cara kita hidup dan bisnis saat ini, dan banyak digunakan untuk tujuan seperti komunikasi, promosi, pemasaran, dan hiburan. Pengguna Instagram dapat mengkritik foto atau video yang diposting orang-orang di bagian komentar. Komentar yang dibuat dalam bentuk kalimat digunakan sebagai input dan output digunakan dalam bentuk kalimat identifikasi yang mengandung *cyberbullying* dan *non-cyberbullying Instagram* termasuk *platform* media sosial yang memungkinkan pengguna berbagi video dan foto pendek dengan pengikut mereka baik secara publik maupun pribadi. Instagram memungkinkan pengguna untuk memperluas jaringan sosial mereka dengan mengikuti akun orang lain dan juga memperbolehkan pengguna untuk berinteraksi dengan konten yang diunggah melalui berbagai opsi seperti "like", "komentar", "repost", dan sebagainya. *Text mining* bertujuan untuk menemukan informasi berharga yang tersembunyi baik dari sumber informasi terstruktur dan tidak terstruktur. Analisis sentiment sangat diperlukan filter komentar di media sosial, proses analisis sentiment yaitu untuk mengklasifikasikan informasi sebagai informasi yang termasuk kedalam komentar negative ataupun positif. Algoritma *Support Vector Machine* dapat digunakan untuk menganalisis sentimen bullying online di kolom komentar Instagram. Hasil klasifikasi berupa kelas positif dan negatif, terbagi menjadi perasaan positif tentang *cyberbullying* dan perasaan negatif tentang *cyberbullying*. Penggunaan 400 *dataset* untuk melakukan training dan membentuk model klasifikasi menghasilkan akurasi sebesar 84,25% ,memperoleh *precision* sebesar 80,22%, memperoleh *Recall* sebesar 92,50% dan nilai *AUC* sebesar 0,928.

Kata kunci: Media Sosial, *Instagram*, Algoritma *Support Vector Machine*

Abstract:

Social Media includes a digital platform that allows users to interact, exchange information, and create content that can be accessed and shared by other users over the internet. Social media users can connect with people around the world and build an online social network. Facebook Instagram, Twitter, LinkedIn, TikTok, and YouTube are many well-known Social media. Today, social media has many effects on the way we live and do business today, and is widely used for purposes such as communication, promotion, marketing, and entertainment. Instagram users

can criticize photos or videos that people post in the comments section. Comments made in the form of sentences are used as input and output are used in the form of identification sentences containing cyberbullying and non-cyberbullying Instagram includes a social media platform that allows users to share short videos and photos with their followers both publicly and privately. Instagram allows users to expand their social network by following other people's accounts and also allows users to interact with uploaded content through various options such as "like", "comment", "repost", and so on. Text mining aims to find valuable information hidden from both structured and unstructured information sources. Sentiment analysis is needed to filter comments on social media, the process of sentiment analysis is to classify information as information that is included in negative or positive comments. The Support Vector Machine algorithm can be used to analyze online bullying sentiment in Instagram comments. The classification results in the form of positive and negative classes, divided into positive feelings about cyberbullying and negative feelings about cyberbullying. The use of 400 datasets to conduct training and form a classification model resulted in accuracy of 84.25%, obtaining precision of 80.22%, obtaining Recall of 92.50% and AUC value of 0.928.

Keywords: *Social Media, Instagram, Support Vector Machine Algorithm*

1. Pendahuluan

Media sosial termasuk suatu *platform* digital yang mengizinkan pemakainya untuk berinteraksi, bertukar informasi, serta membuat konten yang dapat diakses dan dibagikan oleh pengguna lainnya melalui internet. Pengguna media sosial dapat berhubungan dengan orang-orang di seluruh dunia dan membangun jaringan sosial online. Banyak media sosial yang terkenal misalnya *Facebook, Instagram, Twitter, LinkedIn, TikTok*, dan *YouTube*. Pengguna Instagram dapat mengkritik foto atau video yang diposting orang-orang di bagian komentar. Komentar yang dibuat dalam bentuk kalimat digunakan sebagai input dan output digunakan dalam bentuk kalimat identifikasi yang mengandung *cyberbullying* dan *non-cyberbullying* [1].

Cyberbullying merujuk pada tindakan menggunakan teknologi atau komunikasi *elektronik* untuk melakukan perundungan atau pelecehan terhadap seseorang. *Cyberbullying* dapat memiliki dampak negatif yang serius pada kesehatan mental dan kesejahteraan korban, dan menjadi masalah yang semakin meningkat di era digital saat ini. Penting bagi individu untuk menyadari konsekuensi yang mungkin terjadi akibat *cyberbullying*, dan bagi orang tua, pendidik, dan pembuat kebijakan untuk mengambil langkah-langkah untuk mencegah terjadinya *cyberbullying*.

Pasal 27(3) dan (4) UU Perubahan atas UU No. 19 Tahun 2016 tersebut sebelumnya UU No. 11 Tahun 2008 mengenai Pernyataan Informasi dan Transaksi Elektronik Distribusi sangat dilarang Informasi yang terkandung melalui Internet fitnah, hinaan dan Ancaman melalui Internet [2].

Instagram termasuk *platform* media sosial yang memungkinkan pengguna berbagi video dan foto pendek dengan pengikut mereka baik secara publik maupun pribadi. Instagram memungkinkan pengguna untuk memperluas jaringan sosial mereka dengan mengikuti akun orang lain dan juga memperbolehkan pengguna untuk berinteraksi dengan konten yang diunggah melalui berbagai opsi seperti "like", "komentar", "repost", dan sebagainya [3].

Support Vector Machine (SVM) menyertakan algoritme pembelajaran mesin populer untuk tugas klasifikasi dan regresi. SVM adalah algoritma pembelajaran terawasi yang dapat menangani data linier dan nonlinier. Ide dasar SVM adalah menemukan hyperplane dalam ruang berdimensi tinggi yang memisahkan kelas yang berbeda sejauh mungkin [4].

Analisis Sentimen sangat diperlukan Filter komentar di media sosial. Proses analisis sentimen adalah untuk mengklasifikasikan informasi sebagai dalam Emosi dan Emosi Positif negatif. Informasi akan dikategorikan sebagai Jika informasi yang disampaikan adalah kelas positif Untuk memiliki nilai atau setuju dengan sesuatu. Oleh karena itu, algoritma *support vector Machine* dipilih Mesin yang digunakan untuk penelitian ini. Analisis sentimen adalah proses yang dirancang untuk menentukan konten kumpulan data Jadilah positif, negatif atau netral dalam bentuk teks (dokumen, kalimat, sentimen, dll.) [5]. *Ciberbullying* pada *twitter* dengan menggunakan metode SVM dan *Information Gain* (IG) sebagai seleksi fiturnya mendapatkan akurasi sebesar 75% dan nilai presisi 70,27% setelah menggunakan *Information Gain* (IG) hasil akurasi yang didapat sebesar 76,66%, presisi 72,22% dapat dilihat bahwa dengan menggunakan svm dan seleksi fitur IG maka akurasi bertambah [6]. Metode KNN juga

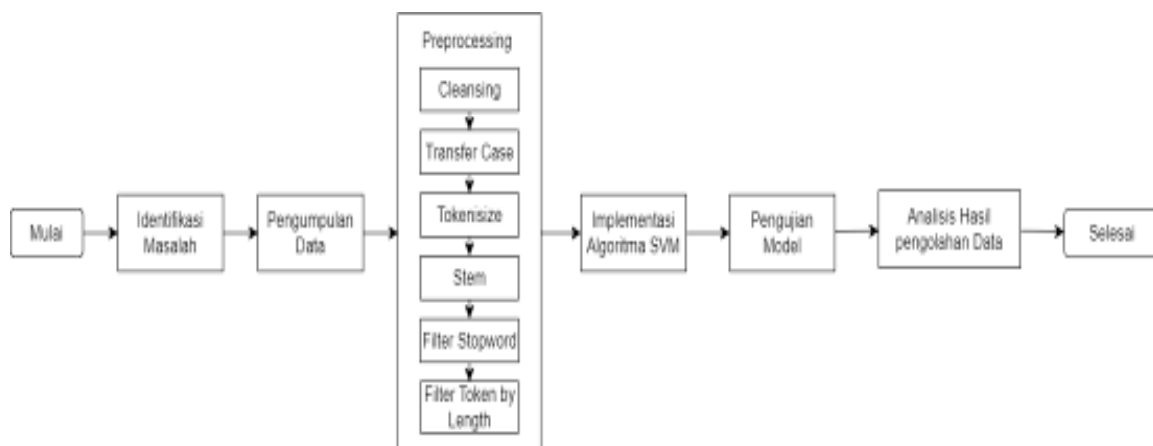
bisa mengklasifikasikan komentar bullying dari penelitian dengan menggunakan 1000 data, KNN telah diuji menggunakan 5 nilai K yang berbeda (7,9,11,13,15) dengan pembagian data 70:30, 80:20 dan 90:10 menggunakan confusion matrix, maka didapatkan nilai akurasi pada nilai K ke 13 yaitu sebesar 77% [7].

Text Mining (juga dikenal sebagai analisis teks) menggunakan pemrosesan bahasa alami (NLP) Teks berbasis lebih baik (tidak terstruktur) dalam dokumen dan database, analisis dan pembelajaran. Penambangan teks adalah tujuan Memproses teks menjadi informasi Dari mode ramalan dan Melalui tren model statistik. Tujuan penambangan teks adalah Opini analisis, emosi, evaluasi, Sikap, penilaian, emosional seseorang Tahu apakah itu terkait dengan tema, Layanan, organisasi, individu atau kegiatan Beberapa Gunakan teks Penambangan digunakan untuk pengelompokan, klasifikasi, Pengambilan informasi dan informasi Ekstrak [8]. Berdasarkan uraian latar belakang diatas penulis. Dalam penelitian untuk mengklasifikasikan data textmining dibutuhkan tools yang berupa Rapid Miner sebagai perhitungan dengan menggunakan algoritma SVM. Rapid Miner adalah perangkat lunak analisis data mandiri dan mesin penambangan data yang dapat diintegrasikan ke dalam produk sendiri. Rapid Miner ditulis dalam Bahasa java, sehingga bekerja di semua sistem operasi [10].

Penelitian ini bertujuan untuk mengukur tingkat cyberbullying di Indonesia dengan menganalisis komentar di media sosial Instagram dengan menggunakan *Support Vector Machine* (SVM) untuk melihat nilai *accuracy* pada *dataset* sebanyak 400 data.

2. Metode Penelitian

Metode penelitian yang digunakan untuk menganalisis sentiment cyberbullying di Instagram menggunakan metode *Support Vector Machine* (SVM) yang ditunjukkan dalam Gambar 1.



Sumber: Hasil Penelitian (2023)

Gambar 1. Metode Penelitian

Gambar 1. adalah alur penelitian yang dimulai dengan mengidentifikasi permasalahan seperti adanya komentar yang dapat menjadi sarana untuk *cyberbullying* dan penyebaran konten negatif di platform *Instagram*. Masalah tersebut membutuhkan solusi yang efektif untuk mengatasi dan mencegah konten *cyberbullying* di platform *Instagram*.

Langkah selanjutnya adalah mengumpulkan data yang berkaitan dengan komentar *cyberbullying* pada aplikasi *Instagram*. Pengumpulan data dilakukan dengan menggunakan teknik *scrapping* yang merupakan metode untuk mengambil data secara otomatis dari suatu sumber dalam hal ini komentar-komentar pada aplikasi *Instagram*. Pengumpulan data dilakukan dalam rentang waktu antara tanggal 1 Desember sampai 31 Desember 2019.

Setelah memperoleh data, tahapan awal dari pengolahan data adalah *preprocessing*. Tahap ini dimulai dengan mengambil data yang telah di peroleh kemudian membersihkan data, mengubah teks menjadi huruf kecil, mengganti kata-kata slang, menghapus kata-kata penghubung, dan mengaplikasikan stemming. Data yang telah melalui tahap *preprocessing* kemudian dilakukan proses implementasi menggunakan software *Rapidminer Studio* menggunakan metode algoritma SVM. SVM digunakan untuk mengklasifikasikan komentar

sebagai positif atau negatif berdasarkan sentimen yang terkandung di dalamnya. Pada proses implementasi penelitian ini, terdapat tiga proses utama. Pertama, proses pembobotan TF-IDF digunakan untuk menghitung bobot kata-kata dalam teks berdasarkan frekuensi kemunculannya dalam dokumen dan frekuensi kemunculannya dalam seluruh dokumen. Kedua, dilakukan proses data visualization untuk menyajikan hasil dalam bentuk grafik, diagram, atau plot guna memperoleh wawasan yang lebih intuitif dan mudah dipahami mengenai distribusi sentimen. Terakhir, dilakukan proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) di mana teks yang telah diolah dan dibobotkan akan digunakan untuk melatih model klasifikasi yang dapat memprediksi sentimen berdasarkan fitur-fitur teks.

Tahap selanjutnya dalam penelitian ini adalah melakukan pengujian model performance dari model yang telah dihasilkan menggunakan *confusion matrix*. Pengujian ini dilakukan untuk mengukur tingkat keakuratan model dalam mengklasifikasikan data. Terakhir dilakukan analisis hasil pengolahan data menggunakan kurva ROC yang bertujuan untuk mengetahui kategori dan nilai model performance hasil penelitian. *Support Vector Machine* (SVM) adalah teknik yang digunakan untuk melakukan prediksi, baik dalam kasus regresi maupun klasifikasi. SVM memiliki prinsip utama untuk menemukan *hyperplane* yang terbaik dan berfungsi memisahkan dua buah kelas pada ruang input [6]. SVM (*Support Vector Machine*) adalah metode yang digunakan untuk melakukan klasifikasi atau regresi. Terdapat beberapa variasi SVM yang berbeda, tetapi yang paling umum adalah SVM dengan fungsi kernel. Rumus dasar SVM untuk klasifikasi adalah sebagai berikut:

$$K(X_1, X_j) = \tanh(ax_1 + X_j + \beta)$$

Keterangan :

K = Kernel

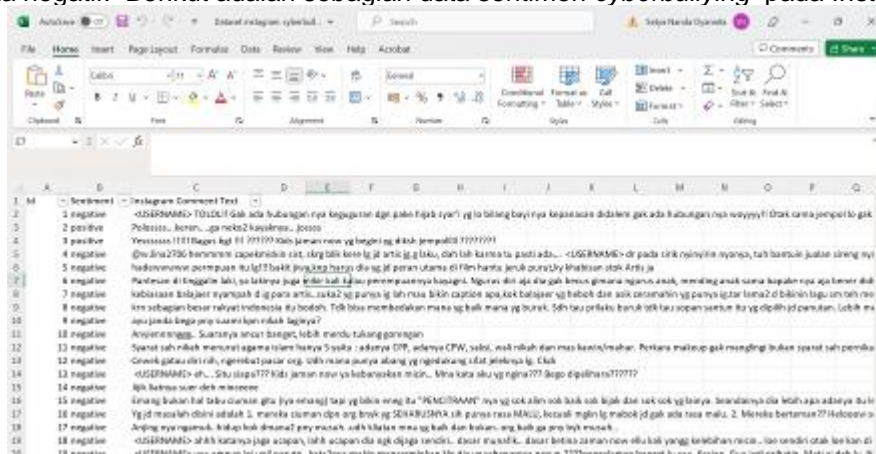
X & Y = Vector input space

d = quadratic

α = scalar parameter

3. Hasil dan Pembahasan

Pengumpulan data pada penelitian ini menggunakan teknik *scrapping* yaitu proses otomatis untuk mengumpulkan data dari suatu sumber dengan cara mengekstrak informasi yang diinginkan dari halaman yang tersedia <https://github.com/rizalespe/Dataset-Sentimen-Analisis-Bahasa> di berisi komentar-komentar *cyberbullying* pada *Instagram*. Hasil dari pengumpulan data yaitu sebesar 400 data yang sudah diberi label meliputi 200 data positif dan 200 data negatif. Berikut adalah sebagian data sentimen *cyberbullying* pada *Instagram* yaitu:



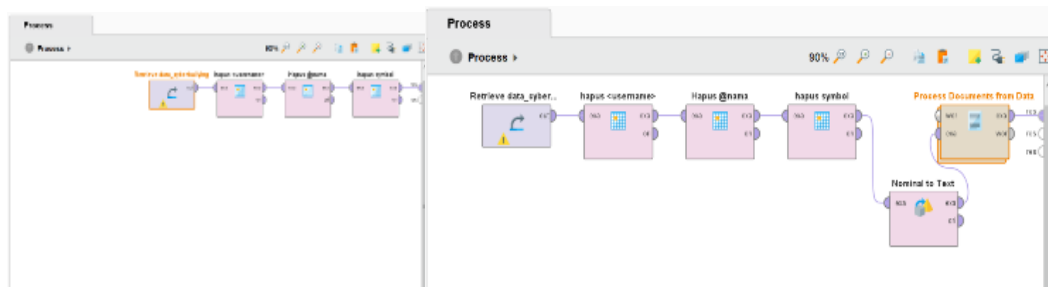
Sumber: Hasil Penelitian (2023)

Gambar 2. Data Komentar *Cyberbullying* di Instagram

Tahap *Preprocessing*

Preprocessing adalah tahapan awal dalam pengolahan data di mana data mentah atau tidak terstruktur diubah menjadi data yang lebih terstruktur dan siap digunakan untuk analisis

lebih lanjut. Pada penelitian ini dilakukan enam tahapan *Preprocessing* menggunakan bantuan *tool rapidminer* yaitu sebagai berikut :



Sumber: Hasil Penelitian (2023)

Gambar 3. Proses *Preprocessing*

Cleansing

Pada penelitian ini dilakukan proses cleansing menggunakan bantuan *tool rapidminer* untuk melakukan proses permbersihan data sentimen. Adapun proses dalam *cleansing* ini dapat ditujukan pada Gambar 4.2:

Hasil dari proses *cleaning* dapat dilihat pada tabel hasil data setelah dan sesudah *cleansing* yang ditujukan pada tabel 1.

Tabel 1. Hasil *Cleansing*

Sebelum	Sesudah	Kategori
<USERNAME> TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy!! Otak sama jempol lo gak sinkron sih ya jadinya asal nulis komentar!	TOLOL Gak ada hubungan nya keguguran dgn pake hijab syarbilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy Otak sama jempol gak sinkron sih jadinya asal nulis komentar	Negative
Polossss...keren....ga neko2 kayaknya...jossss	Polossss keren neko kayak nyajossss	Positive
Yesssssss !!!!! Bagus bgt !!! ????? Kids jaman now yg begini yg diksh jempolllll ????????	Yesssssss Bagus bgt Kids jaman now begini diksh jempolllll	Positive

Sumber: Hasil Penelitian (2023)

Transform Case

Pada penelitian ini akan diubah menjadi huruf kecil. Berikut adalah tabel dari hasil dari *Transform case* disajikan pada Tabel 2. Sebagai berikut:

Tabel 2. Hasil Data *Transform Case*

Sebelum	Sesudah	kategori
TOLOL Gak ada hubungan nya keguguran dgn pake hijab syar bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy Otak sama jempol gak sinkron sih jadinya asal nulis komentar	tolol gak ada hubungan nya keguguran dgn pake hijab syar bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy otak sama jempol gak sinkron sih jadinya asal nulis komentar	Negative
Polossss keren neko kayaknya jossss	polossss keren neko kayaknya jossss	Positive
Yesssssss Bagus bgt Kids jaman now begini diksh jempolllll	yesssssss bagus bgt kids jaman now begini diksh jempolllll	Positive

Sumber: Hasil Penelitian (2023)

Tokenisize

Pada proses *tokenize* penelitian ini setiap kata dalam dokumen dirubah menjadi berbentuk token atau kata – kata. Berikut adalah tabel dari hasil dari *tokenize* disajikan pada Tabel 3. yaitu :

Tabel 3. Hasil *Tokenisize*

Sebelum	Sesudah	Kategori
tolol gak ada hubungan nya keuguran dgn pake hijab syar bilang bayi nya kepanasan didalam gak ada hubungan nya woyyyy otak sama jempol gak singkron sih jadinya asal nulis komentar	tolol, gak, ada, hubungan, nya, keuguran, dgn, pake, hijab, syar, bilang, bayi, nya, kepanasan, didalam, gak, ada, hubungan, nya, woyyyy, otak, sama, jempol, gak, singkron, sih, jadinya, asal, nulis, komentar	Negative
polossss keren neko kayaknya jossss	polossss, keren, neko, kayaknya, jossss	Positive
yesssssss bagus bgt kids jaman now begini diksh jempolllll	yesssssss, bagus, bgt, kids, jaman, now, begini, diksh, jempolllll	Positive

Sumber: Hasil Penelitian (2023)

Stem

Kemudian selanjutnya dilakukan proses *stemming* dengan cara menghilangkan imbuhan atau awalan pada kata-kata sehingga hanya menyisakan kata dasarnya [25]. Berikut disajikan Tabel 4. hasil dari *stemming* yaitu:

Tabel 4. Hasil *Stem*

Sebelum	Sesudah	Kategori
tolol, gak, ada, hubungan, nya, keuguran, dgn, pake, hijab, syar, bilang, bayi, nya, kepanasan, didalam, gak, ada, hubungan, nya, woyyyy, otak, sama, jempol, gak, singkron, sih, jadinya, asal, nulis, komentar	tolol, gak, ada, hubung, nya, gugur, dgn, pake, hijab, syar, bilang, bayi, nya, panas, dalam, gak, ada, hubung, nya, woy, otak, sama, jempol, gak, singkron, sih, jadi, asal, nulis, komentar	Negative
polossss, keren, neko, kayaknya, jossss	polos, keren, neko, kayak, jos	Positive
yesssssss, bagus, bgt, kids, jaman, now, begini, diksh, jempolllll	yes, bagus, banget, kids, jaman, now, begini, dikasih, jempol	Positive

Sumber: Hasil Penelitian (2023)

Filter Stopword

Filter stopwords adalah suatu proses dalam analisis teks yang bertujuan untuk menghilangkan kata-kata yang tidak memberikan kontribusi penting dalam pemahaman isi dokumen. Kata-kata tersebut disebut *stopwords* atau kata penghubung [26]. Berikut disajikan Tabel 5. hasil dari *filter stopwords* yaitu:

Tabel 5. Hasil Data *Filter Stopwords*

Sebelum	Sesudah	Kategori
tolol, gak, ada, hubung, nya, gugur, dgn, pake, hijab, syar, bilang, bayi, nya, panas, dalam, gak, ada, hubung, nya, woyyyy, otak, sama, jempol, gak, singkron, sih, jadi, asal, nulis, komentar	tolol, hubung, gugur, pake, hijab, syar, bilang, bayi, panas, hubung, woy, otak, jempol, singkron, jadi, nulis, komentar	Negative
polos, keren, neko, kayak, jos	polos, keren, neko, jos	Positive
yes, bagus, banget, kids, jaman, now, begini, dikasih, jempol	yes, bagus, banget, kids, jaman, now, kasih, jempol	Positive

Sumber: Hasil Penelitian (2023)

Filter token by length

Lalu proses terakhir pada proses preprocessing yaitu *Filter tokens by length* yaitu suatu proses dalam analisis teks yang bertujuan untuk menghilangkan kata-kata yang memiliki jumlah

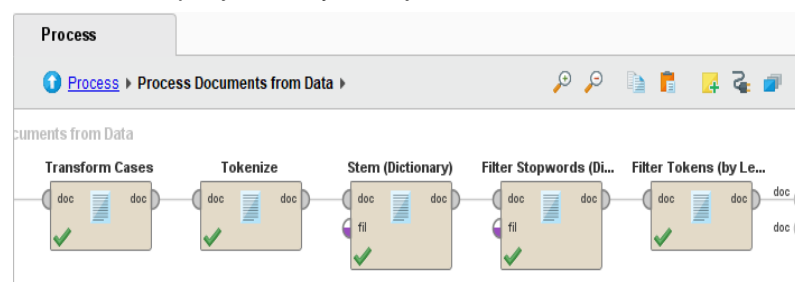
huruf tertentu [26]. Berikut disajikan tabel 6 berisi hasil dari proses *Filter tokens by length* dengan parameter huruf 4-15 huruf per kata :

Tabel 6. Hasil Data *Filter Tokens By Length*

Sebelum	Sesudah	Kategori
tolol, hubung, gugur, pake, hijab, syar, bilang, bayi, panas, hubung, woy, otak, jempol, singkron, jadi, nulis, komentar	tolol, hubung, gugur, pake, hijab, syar, bilang, bayi, panas, hubung, otak, jempol, singkron, jadi, nulis, komentar	<i>Negative</i>
polos, keren, neko, jos	polos, keren, neko	<i>Positive</i>
yes, bagus, banget, kids, jaman, now, kasih, jempol	bagus, banget, kids, jaman, kasih, jempol	<i>Positive</i>

Sumber: Hasil Penelitian (2023)

Adapun untuk membantu proses preprocessing data mulai dari tahap *Transform case* , *tokenize*, *stem*, *Filter Stopword* sampai *Filter token by length* peneliti akan menggunakan bantuan *tool rapidminer*. Berikut penjelasannya disajikan Gambar 4.:



Sumber: Hasil Penelitian (2023)

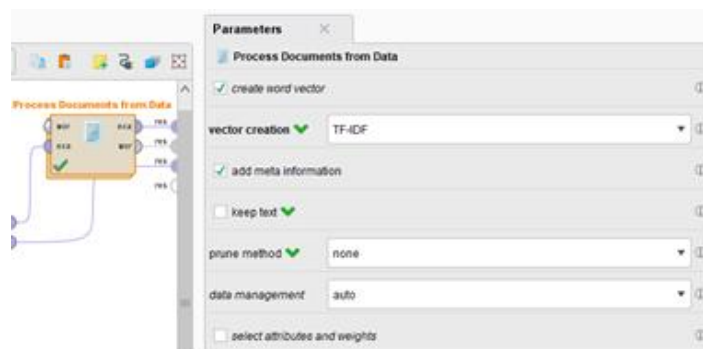
Gambar 4. Lanjutan Proses Processing 2

Implementasi

Pada proses *implementasi* penelitian ini terdapat tiga proses utama yaitu pembobotan TF-IDF, *Visualization*, dan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Hasil Proses tersebut disajikan pada penjelasan sebagai berikut:

Pembobotan TF-IDF

Adapun didalam operator pembobotan TF-IDF peneliti memasukan tahapan tahap *Transform case*, *tokenize*, *stem*, *Filter Stopword* sampai *Filter token by length* yang sudah dijelaskan sebelumnya. Proses dalam pembobotan TF-IDF menggunakan *tool rapidminer* disajikan dalam Gambar 5 berikut ini:



Sumber: Hasil Penelitian (2023)

Gambar 5. Proses Pembobotan TF-IDF

Adapun dengan menggunakan metode TF-IDF kata-kata yang sering muncul dalam satu dokumen akan memiliki bobot yang tinggi, sementara kata-kata yang muncul jarang dalam

satu dokumen namun sering muncul dalam seluruh dokumen akan memiliki bobot yang rendah. Berikut disajikan Gambar 6. hasil dari jumlah frekuensi kemunculan kata pada penelitian ini:

Row No.	word	in documents	total	in class (ne...	in class (po...
1	anak	41	55	25	30
2	orang	45	55	44	11
3	semoga	39	44	4	40
4	artis	36	43	29	14
5	cantik	36	42	2	40
6	kaya	30	34	22	12
7	suka	26	31	11	20
8	punya	24	30	18	12
9	baik	25	29	14	15
10	allah	23	26	4	22

Sumber: Hasil Penelitian (2023)

Gambar 6. Hasil Pembobotan TF-IDF

Data Visualization

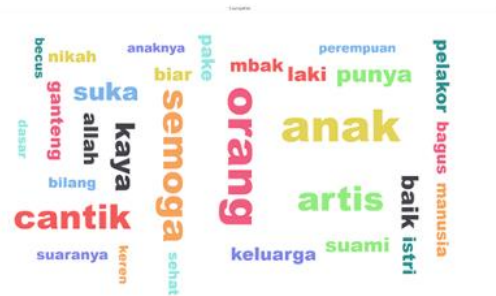
Setelah tahapan-tahapan sebelumnya selesai dilakukan langkah selanjutnya adalah melakukan visualisasi hasil data dengan menghubungkan operator *Process Documents from Data* ke operator *WordList to Data*. Operator ini bertugas untuk menghitung nilai bobot dan frekuensi kemunculan dari setiap kata pada *dataset* yang telah melalui tahapan *Transform case*, *tokenize*, *stem*, *Filter stopwords*, dan *Filter token by length*. Berikut adalah proses *WordList to Data* menggunakan tool *rapidminer*:



Sumber: Hasil Penelitian (2023)

Gambar 7. Proses Data Visualization

Setelah proses ini maka selanjutnya adalah melakukan visualisasi *wordcloud* pada hasil yang telah diperoleh untuk dapat lebih mudah memahami informasi yang dihasilkan yaitu :

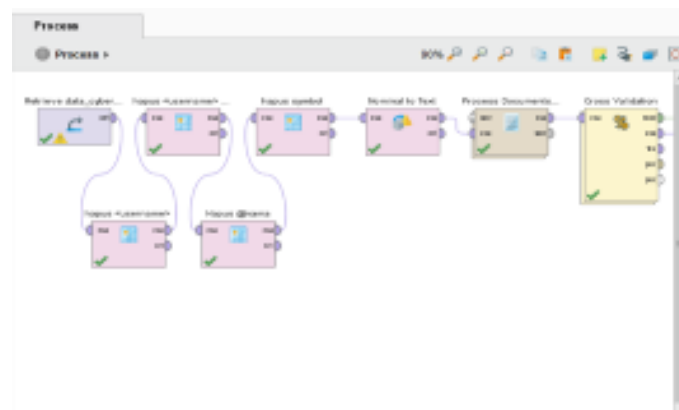


Sumber: Hasil Penelitian (2023)

Gambar 8. Hasil Data Visualization

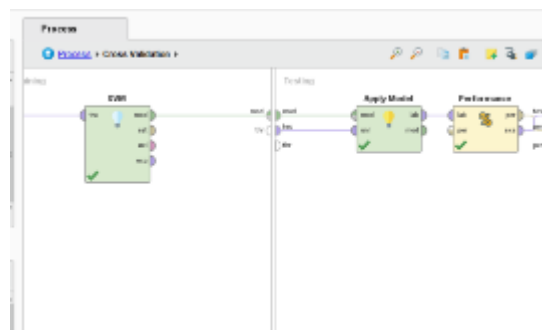
Klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*

Pada tahapan implementasi, dilakukan pemodelan klasifikasi teks menggunakan algoritma Support Vector Machine (SVM). Data dipisah menjadi data uji dan data latih menggunakan pengujian k-fold cross-validation dengan nilai $k = 10$, sesuai dengan penelitian sebelumnya yang menunjukkan bahwa $k = 10$ menghasilkan performa model terbaik. Data sentimen dibagi menjadi sepuluh kelompok, dan setiap kelompok digunakan bergantian sebagai data uji, sementara kelompok lainnya digunakan sebagai data latih dalam percobaan berikutnya. Proses klasifikasi menggunakan algoritma SVM disajikan dalam Gambar 9. dan Gambar 10:



Sumber: Hasil Penelitian (2023)

Gambar 9. Proses SVM

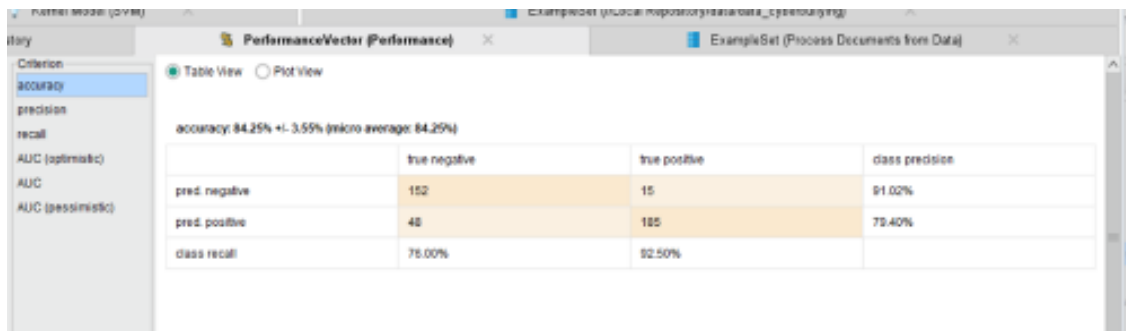


Sumber: Hasil Penelitian (2023)

Gambar 10. Lanjutan Proses SVM

Evaluasi Pengujian

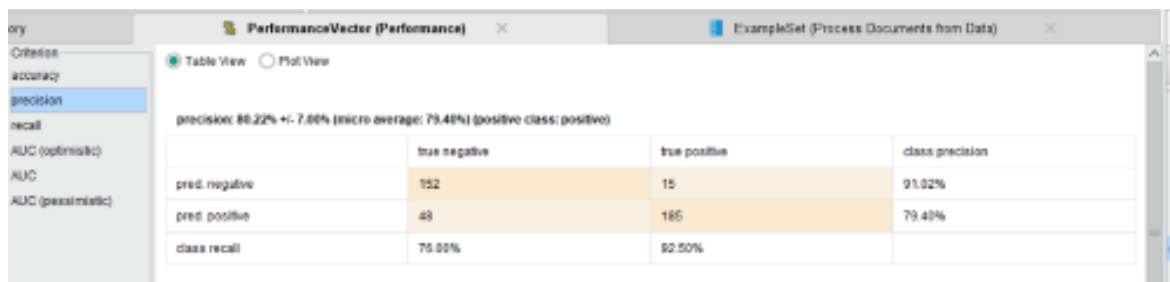
Pada tahapan ini, dilakukan pengujian performa model yang telah dihasilkan menggunakan confusion matrix. Pengujian ini bertujuan untuk mengukur tingkat keakuratan model dalam mengklasifikasikan data. Operator cross-validation digunakan untuk membagi data menjadi data uji dan data latih. Kemudian, data tersebut dihubungkan kembali dengan Support Vector Machine (SVM) untuk proses pemodelan. Terakhir, hasilnya dihubungkan dengan operator performance untuk mengetahui performa model menggunakan confusion matrix. Hasil dari proses ini disajikan dalam Gambar. 11:



Sumber: Hasil peneliti (2023)

Gambar 11. Hasil Accuracy

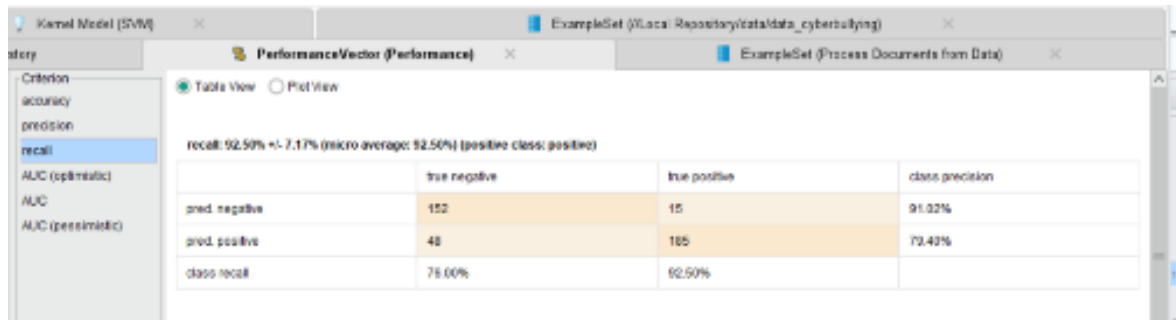
Hasil pemodelan menggunakan *Support Vector Machine* (SVM) pada Gambar 11 menunjukkan *accuracy* sebesar 84,25% dalam mengklasifikasikan komentar *cyberbullying* pada *Instagram*. Angka *accuracy* yang cukup tinggi tersebut menunjukkan bahwa model yang dibangun memiliki kemampuan yang baik dalam mengenali dan membedakan sentimen positif dan negatif dalam teks. Hasil ini menunjukkan bahwa model klasifikasi menggunakan metode *Support Vector Machine* (SVM) dapat digunakan sebagai alat yang efektif dalam mengklasifikasikan apakah suatu teks mengandung sentimen positif atau negatif terkait dengan data komentar *cyberbullying* pada *Instagram*.



Sumber: Hasil Penelitian (2023)

Gambar 12. Hasil Precision

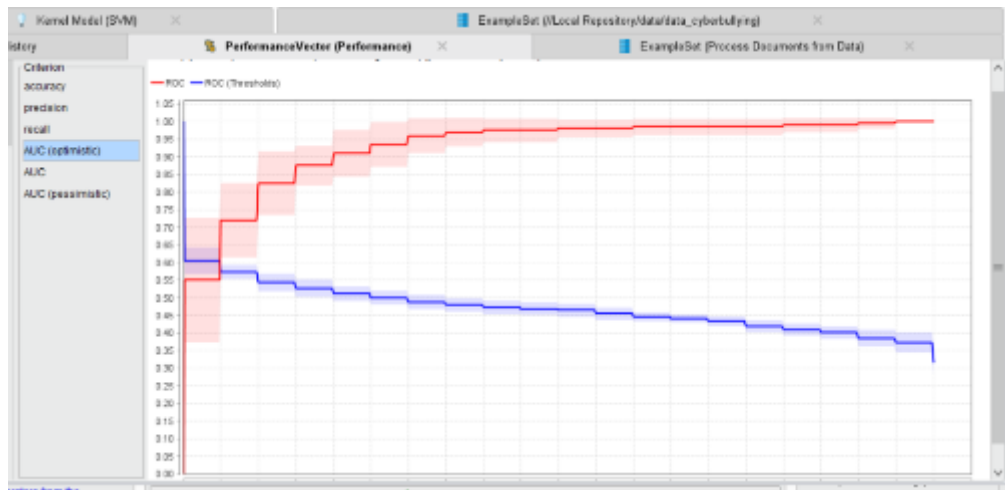
Gambar 12. menunjukkan hasil dari *precision* pada hasil pemodelan. Adapun hasil klasifikasi komentar *cyberbullying* pada *Instagram* menggunakan *Support Vector Machine* (SVM) memperoleh *precision* sebesar 80,22% yang menunjukan bahwa dari seluruh hasil klasifikasi positif sebanyak 80,22% di antaranya benar-benar positif.



Sumber: Hasil Penelitian (2023)

Gambar 13. Hasil Recall

Gambar 13. menunjukkan Nilai *recall* pada hasil pemodelan. Adapun hasil klasifikasi komentar *cyberbullying* pada *Instagram* menggunakan *Support Vector Machine* (SVM) memperoleh *Recall* sebesar 92,50% yang menunjukkan bahwa model tersebut mampu mengenali dengan baik sentimen negatif dengan tingkat kemampuan model dalam mengenali sentimen negatif sebesar 92,50%. Kemudian proses evaluasi selanjutnya yaitu ditampilkan pada kurva ROC yang bertujuan untuk mengetahui kategori dan nilai model performance hasil penelitian ini yang terdapat pada Gambar 12. sebagai berikut:



Sumber: Hasil Penelitian (2023)

Gambar 14. Kurva ROC

Gambar 14. pada *kurva ROC* menunjukkan bahwa terdapat 2 sumbu yaitu sumbu X dan sumbu Y. Sumbu Y merupakan *True Positive Rate* sebagai *Sensitivity (Recall)* sedangkan sumbu X merupakan *False Positive Rate* sebagai *Specitivity*. Selain itu terdapat 2 garis pada kurva yaitu garis merah dan garis biru. Garis merah merupakan representasi data training sedangkan garis biru merupakan representasi data sampel (*random forest*). Semakin tinggi *True Positive Rate* dan semakin kecil *False Positive Rate* maka *thresholdnya* semakin bagus. Pada gambar terlihat bahwa rentang dari *True Positive Rate* antara 0 sampai 1,05 sedangkan rentang nilai dari *False Positive Rate* antara 0 sampai 1,05 artinya keduanya memiliki kinerja yang sama karena garis kurva berwarna merah dan warna biru sama-sama menjauh dari garis bujur (0,0). Sehingga untuk membandingkan nilai kinerja kurva merah dan biru dalam bentuk angka maka dilakukan dengan membandingkan luas area dibawah kurva atau *Area Under Curve (AUC)*. Berdasarkan gambar tersebut dapat dilihat bahwa luas area kurva berwarna merah jauh lebih besar dibanding luas kurva berwarna biru yaitu menghasilkan nilai *AUC* sebesar 0,928 dan tergolong sebagai kategori klasifikasi paling baik atau *Excellent Classif.*

4. Kesimpulan

Algoritma Support Vector Machine dapat digunakan untuk menganalisis sentiment bullying online di kolom komentar Instagram. Hasil klasifikasi berupa kelas positif dan negative. Penggunaan 400 dataset untuk melakukan training dan membentuk model klasifikasi menghasilkan akurasi yang baik sebesar 84,25%, dan memperoleh precision sebesar 80,22%, serta recall sebesar 92,50% dan nilai AUC sebesar 0,928. Pada penelitian berikutnya bisa menggunakan data yang sama dengan metode yang lain, ataupun dengan menggunakan PSO ataupun GA untuk mengoptimalkan nilai akurasi menjadi lebih baik lagi dibanding hanya dengan menggunakan metode SVM.

Referensi

- [1] L. Afina, H. Raudhoti, A. Herdiani, and D. A. Romadhony, "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)," *J-Cosine*, vol. 4, no. 1, pp. 1–8, 2020.
- [2] S. P. Karyanti, M.Pd. & Aminudin, *Cyberbullying & Body Shaming*.
- [3] M. Jubaidi and N. Fadilla, "Pengaruh Fenomena Cyberbullying Sebagai Cyber-Crime Di Instagram Dan Dampak Negatifnya," *Shaut Al-Maktabah J. Perpustakaan, Arsip dan Dokumentasi*, vol. 12, no. 2, pp. 117–134, 2020.
- [4] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [5] A. Salim, W. Gata, M. H. Fakhri, C. S. Rhayu, and A. Budiarto, "Analisis Sentiment Instagram Menggunakan Metode Support Vector Machine (SVM) Berbasis Grid Search Algorithm (GSA)," pp. 466–472.
- [6] 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.
- [7] R. M. Candra and A. Nanda Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 45–52, 2020.
- [8] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018.
- [9] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2018.
- [10] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019.
- [11] A. H. Anshor and A. Safuwani, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization," *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 170–178, 2023.
- [12] I. R. Afandi, F. N. Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022.