



ANALISIS SENTIMEN CYBERBULLYING PADA KOMENTAR INSTAGRAM MENGGUNAKAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE(SVM).

Tegar Adi Saputra Tampubolon¹, Mariza Devega²

^{1,2}Universitas Lancang Kuning

(Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning)

'Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015)

e-mail.: [1tregaradisaputra45@gmail.com](mailto:tregaradisaputra45@gmail.com), [2marizadevega@unilak.ac.id](mailto:marizadevega@unilak.ac.id)

Abstrak

Instagram merupakan media sosial yang paling populer pada zaman sekarang. Pengguna yang dimulai dari anak-anak, remaja hingga orang dewasa turut mendongkrak popularitas Instagram. Namun, media sosial ini tidak lepas dari bahaya cyberbullying yang sering dilakukan oleh pengguna khususnya pada kolom **kmtir**. Cyberbullying merupakan *intimidasi* yang dilakukan seseorang pada orang lain yang dilakukan melalui *chatroom*, media sosial, *e-mail*, *website* dalam bentuk seperti pelecehan verbal, penghinaan, pencemaran nama baik, ancaman. Bahaya cyberbullying tentunya meresahkan banyak orang dikarenakan dampak yang ditimbulkan, maka dari itu dapat dilakukan suatu analisis sentimen pada kolom komentar Instagram yang berupaya untuk mengetahui sentimen dari setiap komentar. Untuk mengetahui setiap sentimen pada komentar digunakan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan metode klasifikasi *Support Vector Machine* (SVM). Dalam pengambilan data dilakukan dengan cara crawling data pada tolls <https://console.apify.com/> dan menarik data komentar pada sebuah postingan instagram sebanyak 1000 data komentar. Dari data tersebut akan di lakukan pemrosesan dengan perbandingan data latih dan data uji diantaranya:90%:10% dengan akurasi 96%, 80% : 20% dengan akurasi 96%, dan 70% : 30% dengan akurasi 95%.

Kata kunci: *Analisis sentimen*, *Cyberbullying*, *Instagram*, *Support VectorMachine*.

Abstract

Instagram is the most popular social media today. Users ranging from children, teenagers to adults also boosted the popularity of Instagram. However, this social media cannot be separated from the danger of cyberbullying that is often done by users, especially in the comment column. Cyberbullying is bullying carried out by someone on others carried out through chatrooms, social media, e-mail, websites in forms such as verbal harassment, insults, defamation, threats. The danger of cyberbullying is certainly troubling for many people because of the impact caused, therefore a sentiment analysis can be done in the Instagram comment column that seeks to find out the sentiment of each comment. To find out each sentiment in comments, the Term Frequency-Inverse Document Frequency (TF-IDF) feature and the Support Vector Machine (SVM) classification method are used. Data retrieval is done by crawling data on tolls https://console.apify.com/ and pulling comment data on an Instagram post as much as 1000 comment data. From this data, processing will be carried out by comparing training data and test data including: 90%: 10% with an accuracy of 96%, 80% : 20% with an accuracy of 96%, and 70% : 30% with an accuracy of 95%.

Kata kunci: *Analisis sentimen*, *cyberbullying*, *instagram*, *Support VectorMachine*.

1. PENDAHULUAN

Cyberbullying merupakan *intimidasi* yang dilakukan seseorang pada orang lain yang dilakukan melalui *chatroom*, media sosial, *e-mail*, *website* dalam bentuk seperti pelecehan verbal, penghinaan, pencemaran nama baik, ancaman, atau pengungkapan informasi pribadi yang tidak sesuai[1].

Cyberbullying ini sering sekali terjadi pada media sosial instagram, Menurut artikel [2], *instagram* merupakan aplikasi media sosial nomor 4 terpopuler di dunia menurut pengguna aktifnya. Dengan berbagai fitur yang ada di *instagram* menjadi sarana tempat banyak orang melakukan tindakan *cyberbullying*. Selain itu Permasalahan yang sering terjadi dalam penggunaan media sosial khususnya Instagram adalah *cyberbullying*, tidak banyak orang yang menyadari jika komentar ataupun ulasan yang di lontarkan pada seseorang atau pun komunitas merupakan tindakan *bullying*[3].

Dengan adanya permasalahan ini maka dibutuhkannya Analisis sentimen, Analisis sentimen sangat diperlukan dalam menyaring komentar komentar orang yang tidak bertanggung jawab di media sosial. Analisis sentimen merupakan proses komputasi yang melibatkan teknik analisis teks terhadap data *tekstual* dengan tujuan mengelola, memahami, dan mengelompokkan aspek emosional, baik itu *positif* maupun *negatif*. Dalam *konteks* ini, analisis sentimen menjadi krusial sebagai sarana untuk menggali informasi mengenai respons emosional terhadap berbagai isu, produk, atau layanan[4].

Berbagai macam penelitian tentang *cyberbullying* telah dilakukan. Salah satunya yaitu, Analisis sentimen *cyberbullying* pada komentar *facebook* dengan metode *klasifikasi Support Vector Machine* (SVM). Berdasarkan penelitian tersebut Analisis sentimen dengan melakukan *crawling* pada data *twitter* sebanyak 1000 data *tweet* yang berbahasa Indonesia menggunakan *tools RapidMiner*, *Klasifikasi* menggunakan *algoritma SVM (Support Vector Machine)* ini mencapai akurasi sebesar 92% ketika pengujian pada proposisi 80:20 yaitu 80% data latih dan 20% data *testing*[5].

Dalam proses analisis sentimen pada penelitian ini menggunakan metode *support vector machine*(SVM), karena metode *support vector machine* merupakan algoritma yang bekerja menggunakan pemetaan *nonlinear* untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi[6]. Dengan menerapka metode klasifikasi *support vector machine* sehingga dapat memberi pemahaman kepada para pengguna media sosial terkhususnya pengguna aplikasi Instagram agar memahami, apakah komentar yang mereka lontarkan itu mengandung *bullying* atau tidak. Oleh karna itu, peneliti mengangkat penelitian dengan judul “Analisis Sentimen Cyberbullying Pada Komentar Instagram Menggunakan Metode *Klasifikasi Support Vector Machine*”.

2. METODE PENELITIAN



Gambar 1 Metode Penelitian

2.1. Crawling Data

Pengambilan data pada penelitian ini dilakukan dengan crawling data menggunakan *web site console.apify.com*. Data yang diambil berupa data komentar pada salah satu postingan sepak bola Indonesia yang di posting oleh akun PSSI (Persatuan Sepak Bola Seluruh Indonesia).



Gambar 2 Tampilan Tools Crawling Data

Pada gambar diatas adalah tampilan web yang digunakan dalam proses crawling data dari postingan instagram.

2.2. Pre Processing

Pada tahap *pre-processing* data terdapat *text mining* yaitu untuk menganalisis *text* yang berguna untuk mencari informasi yang bermanfaat dengan tujuan tertentu. Penambahan *text* sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan yang kedua dengan ekstrasi informasi dan pengetahuan yang relevan. *Processing* yang digunakan dalam penelitian ini adalah *Tokenizing*, *Filtering* dan *Stemming*[7].

Berikut adalah tahapan tahapan pre processing:

1. Cleaning

Cleaning adalah proses untuk membersihkan karakter-karakter tertentu dari data. Data yang dimasukkan kemudian dibaca dalam bentuk matriks, di mana dilakukan iterasi pada setiap dokumen (komentar) dan setiap kata (*term*) di dalam dokumen tersebut [8].

content	cleaning
@and_mutia postingan nya arkhan waduh ternyata peny**** 😂	mutia postingan nya arkhan waduh ternyata peny
Oh ya kalian ga tau terima kasih apa ya sama Garuda muda mereka udah berjuang untuk negara kita bisa bosan kalian hina kaka kolo Kalina hina kaka berati Kalina di bawah kaka lebih di hina.coba kalin yang jadi yang di bilang alay gaya doang ga bisa kan seperti kaka jangan terlalu menghin orang lain	Oh ya kalian ga tau terima kasih apa ya sama Garuda muda mereka udah berjuang untuk negara kita bisa bosan kalian hina kakakolo Kalina hina kaka berati Kalina di bawah kaka lebih di hinacoba kalin yang jadi yang di bilang alay gaya doang ga bisa kan seperti kaka jangan terlalu menghin orang lain
lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out
umur 22 juga bakal flop,	umur juga bakal flop
@andrisaa hahaha gak tau malu ya bang orang penyakit hati mah kayak begini tau menghujat doang eee bang gak pernah ya di ajarin sopan santun sama orang tua mu mulut itu harus di jaga jangan semena mena menghujat kaka dong kamu aja yang sana maeh	hahaha gak tau malu ya bang orang penyakit hati mah kayak begini tau menghujat doang eee bang gak pernah ya di ajarin sopan santun sama orang tua mu mulut itu harus di jaga jangan semena mena menghujat kaka dong kamu aja yang sana maeh

Gambar 3 Hasil Cleaning

Pada gambar merupakan hasil dari proses *cleaning*, diana dalam proses ini menghilangkan komentar yang mengandung link URL, *username* (yang ditandai dengan karakter '@'), tanda baca, dan angka.

2. Case folding

Case Folding Merupakan proses mengubah semua huruf menjadi huruf kecil atau huruf besar agar konsistensi teks terjaga [9]. Misalnya, mengubah "Hello" menjadi "hello" atau sebaliknya. Tujuannya adalah untuk menyamakan format huruf dalam teks agar tidak ada perbedaan yang diakibatkan oleh perbedaan kapitalisasi.

content	cleaning	case_folding
@and_mutia postingan nya arkhan waduh ternyata peny**** 😂	mutia postingan nya arkhan waduh ternyata peny	mutia postingan nya arkhan waduh ternyata peny
Oh ya kalian ga tau terima kasih apa ya sama Garuda muda mereka udah berjuang untuk negara kita bisa bosan kalian hina kaka kolo Kalina hina kaka berati Kalina di bawah kaka lebih di hina.coba kalin yang jadi yang di bilang alay gaya doang ga bisa kan seperti kaka jangan terlalu menghin orang lain	Oh ya kalian ga tau terima kasih apa ya sama Garuda muda mereka udah berjuang untuk negara kita bisa bosan kalian hina kakakolo Kalina hina kaka berati Kalina di bawah kaka lebih di hinacoba kalin yang jadi yang di bilang alay gaya doang ga bisa kan seperti kaka jangan terlalu menghin orang lain	oh ya kalian ga tau terima kasih apa ya sama garuda muda mereka udah berjuang untuk negara kita bisa bosan kalian hina kakakolo kalina hina kaka berati kalina di bawah kaka lebih di hinacoba kalin yang jadi yang di bilang alay gaya doang ga bisa kan seperti kaka jangan terlalu menghin orang lain
lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out
umur 22 juga bakal flop,	umur juga bakal flop	umur juga bakal flop
@andrisaa hahaha gak tau malu ya bang orang penyakit hati mah kayak begini tau menghujat doang eee bang gak pernah ya di ajarin sopan santun sama orang tua mu mulut itu harus di jaga jangan semena mena menghujat kaka dong kamu aja yang sana maeh	hahaha gak tau malu ya bang orang penyakit hati mah kayak begini tau menghujat doang eee bang gak pernah ya di ajarin sopan santun sama orang tua mu mulut itu harus di jaga jangan semena mena menghujat kaka dong kamu aja yang sana maeh	hahaha gak tau malu ya bang orang penyakit hati mah kayak begini tau menghujat doang eee bang gak pernah ya di ajarin sopan santun sama orang tua mu mulut itu harus di jaga jangan semena mena menghujat kaka dong kamu aja yang sana maeh

Gambar 4 Hasil Case Folding

Pada gambar diatas adalah hasil dari proses *case folding*, dimana dalam proses diatas mengubah kalimat yang mengandung huruf kapital menjadi huruf kecil.

3. Tokenisasi

Tokenisasi adalah proses memecah teks menjadi bagian-bagian yang lebih kecil yang disebut sebagai "token". Setiap token dapat berupa kata, frasa, atau bahkan karakter tergantung pada konteks tokenisasi yang digunakan [10]

content	cleaning	case_folding	tokenize
@and_mutia postingan nya arkan waduh ternyat...	mutia postingan nya arkan waduh ternyata penye	mutia postingan nya arkan waduh ternyata penye	[mutia, postingan, nya, arkan, waduh, ternyat...]
Oh ya kalian ga tau terima kasih apa ya sama G...	Oh ya kalian ga tau terima kasih apa ya sama G...	oh ya kalian ga tau terima kasih apa ya sama g...	[oh, ya, kalian, ga, tau, terima, kasih, apa, ...]
lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	[lemparan, cacat, wkwkwk, gak, guna, kalau, out]
umur 22 juga bakal floop,	umur juga bakal floop	umur juga bakal floop	[umur, juga, bakal, floop]
@andrissaa hahaha gak tau malu ya bang orang p...	hahaha gak tau malu ya bang orang penyakit ha...	hahaha gak tau malu ya bang orang penyakit ha...	[hahaha, gak, tau, malu, ya, bang, orang, peny...

Gambar 5 Hasil Tokenisasi

Pada gambar diatas adalah hasil dari proses *Tokenisasi*, dimana dalam proses ini memisahkan kata pra kata dari satu kalimat komentar yang dipisahkan dengan tanda buka kurung.

4. Stopword Removal

Stopword Removal (Penghapusan Kata-kata Bantu) digunakan untuk Menghilangkan kata-kata yang umumnya tidak memberikan informasi penting dalam pemrosesan teks [11], seperti "dan", "atau", "yang", dan lain-lain.

content	cleaning	case_folding	tokenize	stopword removal
@and_mutia postingan nya arkan waduh ternyat...	mutia postingan nya arkan waduh ternyata penye	mutia postingan nya arkan waduh ternyata penye	[mutia, postingan, nya, arkan, waduh, ternyat...]	[mutia, postingan, nya, arkan, penye]
Oh ya kalian ga tau terima kasih apa ya sama G...	Oh ya kalian ga tau terima kasih apa ya sama G...	oh ya kalian ga tau terima kasih apa ya sama g...	[oh, ya, kalian, ga, tau, terima, kasih, apa, ...]	[oh, ya, ga, tau, terima, kasih, ya, garuda, m...]
lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	[lemparan, cacat, wkwkwk, gak, guna, kalau, out]	[lemparan, cacat, wkwkwk, gak, out]
umur 22 juga bakal floop,	umur juga bakal floop	umur juga bakal floop	[umur, juga, bakal, floop]	[umur, floop]
@andrissaa hahaha gak tau malu ya bang orang p...	hahaha gak tau malu ya bang orang penyakit ha...	hahaha gak tau malu ya bang orang penyakit ha...	[hahaha, gak, tau, malu, ya, bang, orang, peny...	[hahaha, gak, tau, malu, ya, bang, orang, peny...

Gambar 6 Hasil Stopword Removal

Pada gambar adalah hasil dari *stopword removal*, dimana kalimat pada hasil diatas akan di hapuskan dari data yang akann di proses.

5. Stemming

Pada proses ini menghapus kan kata bantu di awal dan di akhir [12] Contohnya, mengubah "memakan", "makan"(jadi yang di hapus dari memakan = 'me'), dan "dimakan" menjadi "makan"(jadi yang di hapus dari memakan = 'di').

content	cleaning	case_folding	tokenize	stopword removal	stemming_data
@and_mutia postingan nya arkan waduh ternyat...	mutia postingan nya arkan waduh ternyata penye	mutia postingan nya arkan waduh ternyata penye	[mutia, postingan, nya, arkan, waduh, ternyat...]	[mutia, postingan, nya, arkan, penye]	mutia postingan nya arkan penye
Oh ya kalian ga tau terima kasih apa ya sama G...	Oh ya kalian ga tau terima kasih apa ya sama G...	oh ya kalian ga tau terima kasih apa ya sama g...	[oh, ya, kalian, ga, tau, terima, kasih, apa, ...]	[oh, ya, ga, tau, terima, kasih, ya, garuda, m...]	oh ya ga tau terima kasih ya garuda muda udah ...
lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	lemparan cacat wkwkwk gak guna kalau out	[lemparan, cacat, wkwkwk, gak, guna, kalau, out]	[lemparan, cacat, wkwkwk, gak, out]	lempar cacat wkwkwk gak out
umur 22 juga bakal floop,	umur juga bakal floop	umur juga bakal floop	[umur, juga, bakal, floop]	[umur, floop]	umur floop
@andrissaa hahaha gak tau malu ya bang orang p...	hahaha gak tau malu ya bang orang penyakit ha...	hahaha gak tau malu ya bang orang penyakit ha...	[hahaha, gak, tau, malu, ya, bang, orang, peny...	[hahaha, gak, tau, malu, ya, bang, orang, peny...	hahaha gak tau malu ya bang orang sakit hati m...

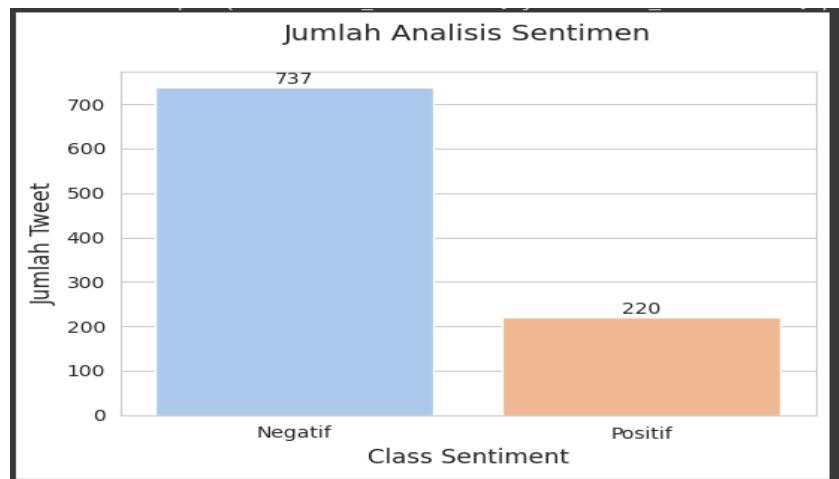
Gambar 7 Hasil Stemming

6. Normalisasi Kata

Normalisasi Kata adalah Proses untuk mengubah kata-kata ke dalam bentuk standar [13]. Ini bisa termasuk penggabungan kata-kata yang sama tetapi ditulis dengan variasi yang berbeda (misalnya, "tidak" dan "ga"), atau konversi singkatan ke bentuk lengkap (misalnya, "dr" menjadi "dokter").

2.3. Pelabelan Kata

Dalam proses pelabelan data ini akan dilakukan pelabelan data yang akan digunakan yaitu sebanyak 1000 data set. Dikarenakan begitu banyak data set yang harus diberi label maka akan ditentukan kata-kata yang mengandung *positif cyberbullying* (kata-kata yang benar positif mengandung unsur *bullying*) dan *negatif cyberbullying* (kata-kata yang tidak mengandung unsur *bullying*). Hasil dari pelabelan dapat diakses pada link *Google Drive* berikut <https://drive.google.com/file/d/1-Zv0q8caUiDtSW-G-6mPGheaGoIBN52v/view?usp=sharing>.



Gambar 8 Hasil Pelabelan [14]

Pada gambar diatas adalah tampilan hasil pelabelan, dimana ada 2 kelas yang telah dilabelkan adalah kelas positif *bullying* sebanyak 125 komentar dan negatif *bullying* sebanyak 832 komentar.

2.4. Pembobotan Kata

Langkah selanjutnya adalah pembobotan kata TF-IDF. TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik yang digunakan dalam pemrosesan teks dan informasi terstruktur lainnya untuk mengevaluasi seberapa penting suatu kata dalam suatu dokumen dalam suatu koleksi dokumen. Tujuan utama dari TF-IDF adalah untuk memberikan skor yang lebih tinggi kepada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang muncul dalam dokumen lain di dalam koleksi[15]. Jika dalam satu teks muncul 2 kali maka teks tersebut akan memiliki skor 2.

2.5. Klasifikasi SVM

Klasifikasi menggunakan metode Support Vector Machine (SVM) adalah proses penilaian kinerja model yang telah dibangun menggunakan algoritma SVM dalam konteks pembelajaran mesin atau pemrosesan data [16]. Berikut adalah hasil dari pemrosesan klasifikasi menggunakan metode SVM.

Accuracy: 0.92					
	precision	recall	f1-score	support	
Negatif	0.90	1.00	0.95	70	
Positif	1.00	0.69	0.82	26	
accuracy			0.92	96	
macro avg	0.95	0.85	0.88	96	
weighted avg	0.93	0.92	0.91	96	

Gambar 9 Hasil Klasifikasi Svm data latih 90%:10%

Pada gambar diatas adalah hasil klasifikasi Suport Vector Machine, menggunakan Perbandingan Data latih 90% dan Data uji 10%. Dimana dari 96 data uji telah terdeteksi hasil yang di dapatkan pada akurasi 92%.

Accuracy: 0.89					
	precision	recall	f1-score	support	
Negatif	0.87	1.00	0.93	143	
Positif	1.00	0.57	0.73	49	
accuracy			0.89	192	
macro avg	0.94	0.79	0.83	192	
weighted avg	0.90	0.89	0.88	192	

Gambar 10 Hasil Klasifikasi Svm data latih 80%:20%

Pada diatas adalah hasil klasifikasi Suport Vector Machine, menggunakan Perbandingan Data latih 80% dan Data uji 20%. Dimana dari 192 data uji telah terdeteksi hasil yang di dapatkan pada akurasi 89%.

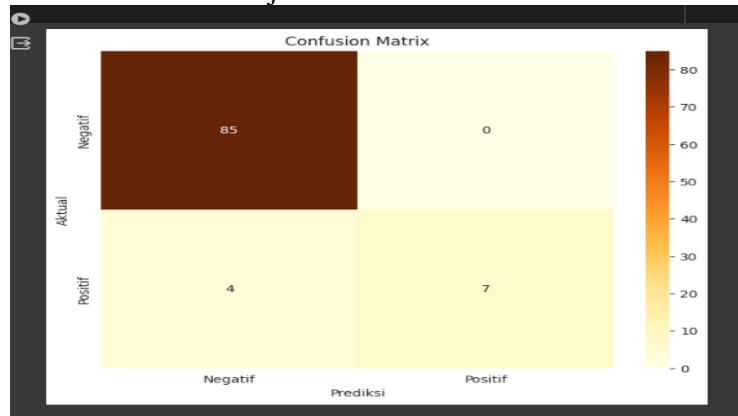
Accuracy: 0.90					
	precision	recall	f1-score	support	
Negatif	0.88	1.00	0.94	219	
Positif	1.00	0.57	0.72	69	
accuracy			0.90	288	
macro avg	0.94	0.78	0.83	288	
weighted avg	0.91	0.90	0.88	288	

Gambar 11 Hasil Klasifikasi Svm data latih 70%:20%

Pada gambar 5.12 diatas adalah hasil klasifikasi Suport Vector Machine, menggunakan Perbandingan Data latih 70% dan Data uji 30%. Dimana dari 288 data uji telah terdeteksi hasil yang di dapatkan pada akurasi 90%.

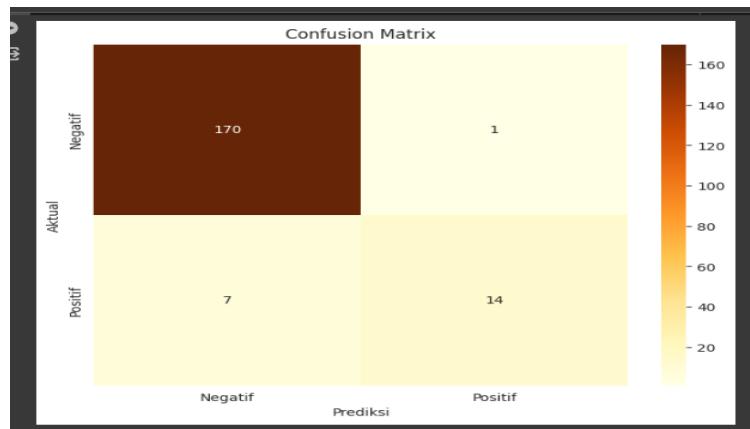
1. Hasil Confusion Matrix

Confusion matrix menyajikan jumlah prediksi yang benar dan yang salah yang dihasilkan oleh model klasifikasi pada suatu dataset yang telah diberi label [17]. Berikut adalah hasil dari 3 percobaan data latih dan data uji:



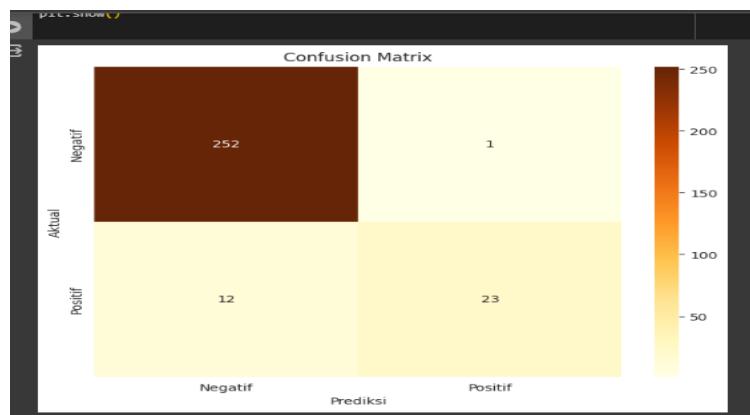
Gambar 12 Tampilan Confusion Matrix 90% :10%

Pada gambar diatas adalah hasil confusion matrix dengan True Positif 7, True Negatif 85, False Positif 0, dan False Negatif 4.



Gambar 13 Tampilan Confusion Matrix 80% :20%

Pada gambar diatas adalah hasil confusion matrix dengan True Positif 14, True Negatif 170, False Positif 1, dan False Negatif 7.



Gambar 14 Tampilan Confusion Matrix 70% :30%

Pada gambar 14 diatas adalah hasil confusion matrix dengan True Positif 23, True Negatif 252, False Positif 1, dan False Negatif 12.

3. Hasil dan Pembahasan

3.1. Hasil

Pada hasil penelitian ini merupakan hasil dari proses 3 kali pengujian data latih dan data uji diantaranya 90% Data latih 10% data uji, 80% data latih 20% data uji, dan 70% data latih 30% data uji. Penjelasan lebih detailnya pada tabel berikut:

Tabel 1 hasil perbandingan data training dan data testing

Data latih	Data uji	Accuracy Positif	Precision Positif	Recall Positif	F1-Score Positif
70%	30%	90%	100%	57%	72%
80%	20%	89%	100%	57%	73%
90%	10%	92%	100%	69%	82%

Hasil penelitian menunjukkan bahwa algoritma Penelitian ini menganalisis 957 komentar setelah pra pemrosesan, dimana algoritma *Support Vector Machine(SVM)* mampu mengklasifikasikan Komentar ke dalam Dua kelas yaitu kelas Positif bullying dan negatif bullying dengan akurasi sebesar 92%.

1. Perhitungan Manualisasi Accuracy menggunakan *Confusion Matrix*.

- #### a Perhitungan Manual Accuracy 90% : 10%

$$\begin{aligned}
 Accuracy &= \frac{(TP + TN)}{TP + TN + FP + FN} \times 100 \\
 &= \frac{(18 + 70)}{18 + 70 + 0 + 8} \times 100 \\
 &= \frac{88}{96} \times 100 = 92\%
 \end{aligned}$$

- a Perhitungan Manual Accuracy 80% : 20%

$$\begin{aligned}
 Accuracy &= \frac{(TP + TN)}{TP + TN + FP + FN} \times 100 \\
 &= \frac{(28 + 143)}{28 + 143 + 0 + 21} \times 100 \\
 &= \frac{171}{192} \times 100 = 89\%
 \end{aligned}$$

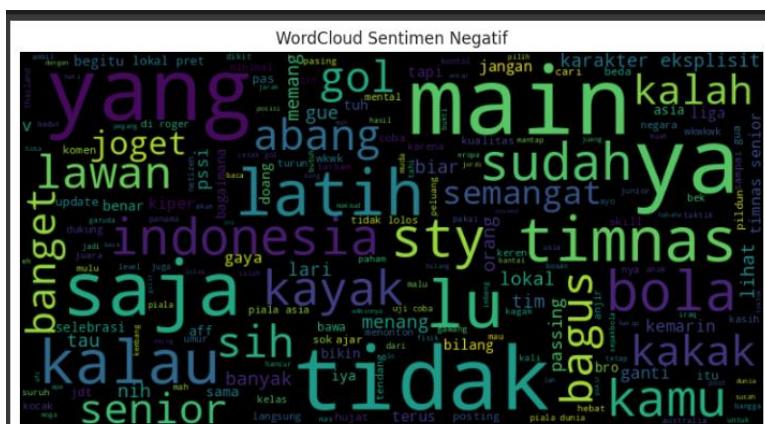
- b Perhitungan Manual Accuracy 70% : 30%

$$\begin{aligned}Accuracy &= \frac{(TP + TN)}{TP + TN + FP + FN} \times 100 \\&= \frac{(39 + 219)}{39 + 219 + 0 + 30} \times 100 \\&= \frac{258}{288} \times 100 = 90\%\end{aligned}$$

Dari ketiga perhitungan manual diatas menunjukan bahwa, hasil dari perhitungan secara manual SVM dengan menggunakan python itu menunjukan hasil akurasi yang sama.

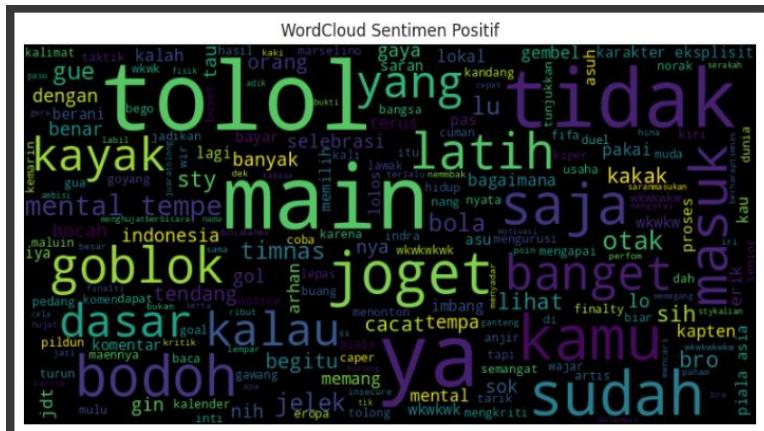
2. Visualisasi Hasil Analisis

Visualisasi hasil dari analisis ini akan di visualisasikan melalui gambar *Worcloud* di bawah ini:



Gambar 15 Tampilan Worcloud Sentimen negatif bullying

Pada Gambar diatas adalah *Worcloud* sentimen negatif *bullying*, dimana kata kata yang ada di dalam *wordcloud* itu adalah kata kata yang sering muncul dan tidak mengandung unsur *bullying*.



Gambar 16 Tampilan Worcloud Sentimen positif bullying

Pada Gambar diatas adalah *Worcloud* sentimen positif *bullying*, dimana kata kata yang ada di dalam *wordcloud* itu adalah kata kata yang sering muncul dan mengandung unsur *bullying*.

3.2 Pembahasan

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa **algoritma Support Vector Machine (SVM)** mampu mengklasifikasikan komentar Instagram terkait cyberbullying dengan tingkat akurasi yang cukup tinggi, terutama pada pembagian data latih dan uji 90:10 yang mencapai akurasi 92%. Hasil ini mengindikasikan bahwa algoritma SVM efektif dalam mendeteksi komentar yang mengandung unsur *bullying*.

Pada pembagian data 80:20 dan 70:30, meskipun akurasinya sedikit menurun, algoritma ini tetap menunjukkan performa yang baik dengan precision yang tetap tinggi. Namun, **recall** untuk komentar positif cyberbullying sedikit lebih rendah, menunjukkan bahwa model memiliki tantangan dalam mendeteksi semua komentar yang benar-benar mengandung unsur cyberbullying.

Word Cloud yang dihasilkan memperlihatkan kata-kata yang sering digunakan dalam komentar yang mengandung unsur *bullying* dan yang tidak. Kata-kata yang muncul dalam sentimen negatif lebih cenderung netral atau tidak menunjukkan unsur *bullying*, sementara kata-kata pada sentimen positif jelas menunjukkan indikasi intimidasi atau pelecehan.

Kedepannya, penelitian ini menyarankan untuk meningkatkan jumlah dataset guna meningkatkan kinerja model. Selain itu, penggunaan metode yang lebih kompleks seperti GRU atau Bi-LSTM bisa dipertimbangkan untuk meningkatkan akurasi analisis sentimen.

4. KESIMPULAN

Dari hasil penelitian Analisis Sentimen *Cyberbullying* pada komentar Instagram menggunakan metode klasifikasi *Support Vector Machine* (SVM), dari 1000 data komentar yang telah diproses pelabelan terklasifikasi 220 komentar positif *bullying* dan 737 komentar negatif *bullying*. Setelah proses pelabelan, dilakukan pengujian dengan berbagai komposisi data latih dan data uji. Spesifiknya, ketika menggunakan 90% data latih dan 10% data uji, akurasi yang didapat adalah 92%; dengan menggunakan 80% data latih dan 20% data uji, akurasi mencapai 89%; sedangkan menggunakan 70% data latih dan 30% data uji, akurasi yang dihasilkan adalah 90%. Dengan demikian, dapat disimpulkan bahwa model SVM mampu mengklasifikasikan komentar Instagram terkait *cyberbullying* dengan akurasi terbaik 92%.

Daftar Pustaka

- [1] F. S. Mutma, "Deskripsi Pemahaman Cyberbullying Di Media Sosial Pada Mahasiswa," *J. Komun.*, vol. 13, no. 2, pp. 165–182, 2019, doi: 10.21107/komunikasi.v13i2.5928.
- [2] I. Deretan and M. Sosial, "Daftar Media Sosial Terpopuler di Dunia April 2023 , Facebook Masih Juara," no. April, 2023.
- [3] V. P. Ramadhan and G. M. Namung, "Klasterisasi Komentar Cyberbullying Masyarakat di Instagram berdasarkan K-Means Clustering," *J-Intech*, vol. 11, no. 1, pp. 32–39, 2023, doi: 10.32664/j-intech.v11i1.846.
- [4] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," vol. 5, no. 2014, pp. 1018–1026, 2021, doi: 10.30865/mib.v5i3.3111.
- [5] R. Nurlaelly, D. Sartika Simatupang, and I. Lucia Kharisma, "Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM)," *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 376–384, 2023, [Online]. Available: <http://ejurnal.umri.ac.id/index.php/coscitech/indexhttps://doi.org/10.37859/coscitech.v4i2.5161>
- [6] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding)," *Ilm. Edutic*, vol. 5, no. 1, pp. 17–25, 2018.
- [7] F. A. Muttaqin and A. M. Bachtiar, "Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak " Dodo Kids Bowswr " Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) Jurnal Ilmiah Komputer dan Informatika" 2020.
- [8] E. Suryati, S. Styawati, and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding dan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi dan Sistem Informasi*, vol. 4, no. 1, pp. 96–106, Mar. 2023, doi: 10.33365/jtsi.v4i1.2445.
- [9] R. Wati and S. Ernawati, "Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python," *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, vol. 6, no. 2, pp. 240–245, Dec. 2021.
- [10] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [11] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 2, pp. 162–168, May 2020.
- [12] I. Kurniawan, A. L. Hananto, S. S. Hilabi, A. Hananto, B. Priyatna, and A. Y. Rahman, "Perbandingan Algoritma Naive Bayes dan SVM dalam Sentimen Analisis Marketplace pada Twitter," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 731–740, Mar. 2023.
- [13] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 1, pp. 147–156, Feb. 2021, doi: 10.25126/jtiik.202183944.
- [14] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, Mar. 2019.

- [15] Ratino, N. Hafidz, S. Anggraeni, and W. Gata, “Sentimen Analisis Informasi Covid-19 Menggunakan Support Vector Machine dan Naïve Bayes,” Jurnal JUPITER, vol. 12, no. 2, pp. 1–11, 2020.
- [16] R. Wahyudi and G. Kusumawardhana, “Analisis Sentimen pada Review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” Jurnal Informatika, vol. 8, no. 2, pp. 200–207, Sept. 2021.
- [17] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM,” Jurnal JDMSI, vol. 2, no. 1, pp. 31–37, 2021.



ZONAsi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](#)