

ANALISIS SENTIMENT *CYBERBULLYING* PADA SOSIAL MEDIA TWITTER MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Muh. Fitra Rizki, Karina Auliasari, Renaldi Primaswara Prasetya
Program Studi Teknik Informatika S1, Fakultas Teknologi Industri
Institut Teknologi Nasional Malang, Jalan Raya Karanglo km 2 Malang, Indonesia
1718053@scholar.itn.ac.id

ABSTRAK

Twitter merupakan salah satu media sosial yang saat ini menjadi trend, karena terdapat banyak sekali berita dan informasi yang dapat direspon dengan cepat dan tepat dari berbagai sudut pandang. Hal ini menjadikan Twitter tidak hanya berdampak positif, tetapi juga berdampak negatif bagi pengguna maupun non-pengguna Twitter, salah satunya adalah cyberbullying. Cyberbullying adalah bentuk intimidasi yang pelaku lakukan untuk melecehkan korbannya melalui perangkat teknologi. Korban yang mengalami Cyberbullying akan mengalami gangguan fisik hingga psikologis seperti kesepian, kegelisahan, depresi yang lebih tinggi, dan merasa hargadirinya rendah. Selain itu korban yang mengalami Cyberbullying juga akan merasakan tekanan sehingga menunjukkan keinginan bunuh diri yang lebih tinggi. Pada penelitian ini dilakukan proses analisis sentiment cyberbullying yang disampaikan oleh pengguna pada media sosial twitter dengan mengembangkan sistem berbasis web untuk mengklasifikasikan sentiment tersebut menggunakan metode support vector machine. Data inputan pada sistem ini berupa konten tweet yang diperoleh dari twitter dengan memasukkan keyword hashtag yang berpotensi menimbulkan cyberbullying seperti #cebong atau #kadrun dan tidak melebihi 100 data tweet. Sedangkan outputnya berupa klasifikasi sentiment cyberbullying atau non-cyberbullying dari setiap tweet yang sudah melewati proses text preprocessing dan pembobotan teks dengan TF-IDF. Dari hasil pengujian menunjukkan dengan menggunakan 100 data tweet, sistem mampu melakukan proses klasifikasi dengan rata-rata waktu 101100,2 milisecond dan kecepatan pemrosesan 0,000989 data per milisecond. Diperoleh pula hasil pengukuran evaluasi klasifikasi dengan menggunakan metode confusion matrix dengan nilai recall 64%, precision 58% dan tingkat accuracy sebesar 70%.

Kata Kunci: *Support Vector machine, Analisis Sentiment, cyberbullying, Text Preprocessing, Term Frequency-Inverse Document Frequency*

1. PENDAHULUAN

Media sosial adalah platform media yang berfokus pada kehadiran pengguna dan mempromosikan aktivitas dan kolaborasi mereka. Oleh karena itu, media sosial dapat dikatakan sebagai media online yang dapat mempererat hubungan antara pengguna dan koneksi sosial. Twitter merupakan salah satu media sosial yang saat ini selalu menjadi pusat trending di Indonesia, karena terdapat banyak sekali informasi yang dapat direspon dengan cepat dan tepat dari berbagai sudut pandang. Hal ini menjadikan Twitter tidak hanya berdampak positif, tetapi juga berdampak negatif bagi pengguna maupun non-pengguna Twitter, salah satunya adalah cyberbullying.

Data Kasus cyberbullying di Indonesia saat ini masih sulit ditemukan secara menyeluruh pada platform media sosial twitter ini. Namun baru-baru ini pemerintah Indonesia berkerjasama dengan POLRI tertanggal 24 februari 2021 telah meresmikan beroperasinya polisi virtual yang bertugas melakukan pelacakan dan memberikan peringatan kepada pengguna yang mengunggah postingan baik berupa text ataupun gambar yang berpotensi melanggar pidana pada UU ITE. Upaya yang dilakukan oleh Twitter dalam menangani cyberbullying diantaranya yaitu pembatasan reply, rethink method, pemblokiran,

mute dan pelaporan akun. Namun, fitur - fitur tersebut belum dapat mengklasifikasikan ke dalam tipikal kalimat atau kata yang mengandung cyberbullying atau Non-Cyberbullying.

Berdasarkan permasalahan yang diuraikan diatas tersebut, maka dibutuhkan adanya analisis sentiment untuk mengklasifikasikan tweet yang mengandung unsur cyberbullying atau Non-Cyberbullying. Sistem yang akan dibuat berbasis web dengan menggunakan metode klasifikasi dari Sentiment Analylis dengan algoritma Support Vector machine (SVM). Sehingga dengan adanya sistem ini nantinya diharapkan dapat membantu kinerja polisi virtual POLRI ini dalam mengidentifikasi tweet yang berpotensi melanggar pidana UU ITE tersebut dan secara tidak langsung akan meminimalisir pengguna untuk membuat konten Twitter yang mengandung unsur cyberbullying.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian terkait analisis sentiment, diantaranya dilakukan oleh Maria Mega Mala Olhang, Sentot Achmadi dan F.X Ariwibisono. di tahun 2020 yaitu mengembangkan analisis sentiment pengguna twitter terhadap COVID-19 di Indonesia menggunakan metode Naive Bayes Classifier (NBC). Pada penelitian ini, dilakukan proses menganalisis

sentimen masyarakat terhadap aspirasi yang disampaikan melalui twitter dengan memasukkan keyword seperti #coronavirusindonesia atau #covid-19 dengan jumlah data tidak melebihi 500 data tweet. Dari hasil penelitian tersebut menunjukkan hasil tingkat akurasi yang didapatkan dengan melakukan pengujian terhadap 75 dokumen tweet adalah pengukuran akurasi recall 32%, precision 80%, F-Measure 45% serta rata-rata akurasi 36% [1].

Ditahun yang sama, Amar P. Natasuwarna juga melakukan penelitian tentang analisis sentimen yaitu seleksi fitur Support Vector Machine pada keberlanjutan pembelajaran daring. Tahapan penelitian terdiri dari pengambilan data mentah, pre-processing data, seleksi fitur dengan Term Frequency dan TF-IDF, klasifikasi dengan Support Vector Machine (SVM), dan evaluasi menggunakan k-Fold Cross Validation dan Confusion Matrix. Data yang digunakan berjumlah 200 data tweet terdiri dari 100 komentar positif dan 100 komentar negatif menggunakan lima rasio perbandingan data latih dan data uji. Penelitian menghasilkan evaluasi tertinggi pada 8-Fold Cross Validation dengan accuracy sebesar 86,00%, precision sebesar 87,38%, dan recall sebesar 85,02% [2].

Bahkan terdapat pula beberapa penelitian tentang analisis sentimen sebelumnya yang membahas terkait cyberbullying pada beberapa platform media sosial seperti dalam penelitian yang dilakukan pada tahun 2018 oleh Wanda Athira Luqyana, dkk. Yang berjudul analisis sentimen cyberbullying pada komentar instagram dengan metode klasifikasi Support Vector Machine, dengan melakukan uji coba pada dokumen yang berisi 400 data yang diambil secara luring (offline) dengan total fitur 1799. Dokumen komentar dibagi menjadi 70% data latih dan 30% data uji. Berdasarkan pengujian yang dilakukan didapatkan parameter terbaik pada metode SVM yaitu dengan nilai degree kernel polynomial sebesar 2, nilai learning rate sebesar 0,0001, dan jumlah iterasi maksimum yang digunakan adalah 200 kali. Dari pengujian tersebut didapatkan hasil akurasi tertinggi sebesar 90% pada komposisi data latih 50% dan komposisi data uji 50% [3].

Pada tahun 2020 juga, Fajar Agus Maulana dan Iin Ernawati melakukan penelitian terkait analisis sentimen cyberbullying ini dengan memanfaatkan media sosial twitter dan menggunakan algoritma Naive Bayes Classifier. Hasil pengujian dengan data uji real time pada tanggal 12 Mei 2020 pukul 01.00 WIB mendapatkan nilai akurasi sebesar 76%. Metode ini cukup baik dalam mengklasifikasikan tweet positif dan negatif, namun pada proses pengujian penelitian ini dalam mendeteksi tweet yang mengandung unsur cyberbullying masih kurang baik, dikarenakan masih terdapatnya tweet yang tidak mengandung unsur cyber bullying didalam data latih yang memiliki label tweet negatif [4].

Dari penelitian-penelitian terkait analisis sentimen yang pernah dilakukan seperti pada contoh diatas, diketahui bahwa terdapat beberapa metode yang dapat digunakan untuk melakukan analisis sentimen ini seperti Naive Bayes Classifier dan Support Vector Machine yang dimana jika dilihat dari hasil penelitian yang dilakukan, terdapat perbedaan tingkat akurasi dari hasil pemodelan dengan kedua sistem tersebut.

Terkait hal tersebut, Nurrun Muchammad Shiddieqy, dkk. Melakukan penelitian studi literatur tentang perbandingan metode untuk proses analisis sentimen di twitter, dimana dalam penelitian tersebut mereka membandingkan beberapa hasil penelitian yang dilakukan sebelumnya dimana perbandingan dua metode yang bisa digunakan untuk melakukan proses analisis sentimen disertai tahapan yang akan dilalui. Metode yang dibandingkan adalah Naive Bayes Classifier dan Support Vector Machine. Dari hasil studi literatur ini dapat disimpulkan dan bisa dilihat bahwa hasil terbaik dapat diraih menggunakan metode Support Vector Machine. Namun, untuk tetap mendapatkan hasil klasifikasi terbaik, maka lebih baik dilakukan pengujian yang bersifat komparatif dengan beberapa metode klasifikasi pada penelitian yang sedang dilakukan [5].

2.2. Dasar Teori

2.2.1 Analisis Sentiment

Analisis Sentimen adalah klasifikasi dari opini dan sentimen yang diungkapkan dalam teks, yang dihasilkan oleh manusia melalui teknologi Data Mining. Analisis Sentimen menyediakan fitur ekstraksi otomatis dan kemampuan representasi dan kinerja yang lebih baik daripada teknik berbasis fitur tradisional [6]. Analisis Sentimen ditujukan untuk mencari pendapat orang lain. Ini tidak hanya berlaku untuk individu tetapi juga berlaku untuk organisasi. Contohnya saat ini, jika seseorang ingin membeli produk konsumen, tidak lagi terbatas untuk meminta pendapat teman dan keluarga seseorang karena ada banyak ulasan pengguna dan diskusi tentang produk di forum publik di website. Bagi sebuah organisasi, mungkin tidak perlu lagi melakukan survei, jajak pendapat, dan memfokuskan diri untuk mengumpulkan opini publik. Beberapa tahun terakhir, postingan pendapat di media sosial juga telah membantu membentuk bisnis, mempengaruhi sentimen publik dan emosi publik [7].

2.2.2 Cyberbullying

Cyberbullying dapat diartikan sebagai teknologi internet untuk menyakiti orang lain dengan cara sengaja dan diulang-ulang. Cyberbullying adalah bentuk intimidasi yang pelaku lakukan untuk melecehkan korbannya melalui perangkat teknologi. Korban yang mengalami Cyberbullying akan melibatkan gangguan fisik hingga psikologis seperti kesepian, kegelisahan, depresi yang lebih tinggi, dan merasa hargadirinya rendah. Selain itu korban yang

mengalami Cyberbullying juga akan merasakan tekanan sehingga menunjukkan keinginan bunuh diri yang lebih tinggi [8].

2.2.3 Text Preprocessing

Text Preprocessing merupakan tahap awal dari text mining untuk mengubah data sesuai dengan format yang dibutuhkan. Proses ini dilakukan untuk menggali, mengolah dan mengatur informasi dan untuk menganalisis hubungan tekstual dari data terstruktur dan data tidak terstruktur [9]. Proses preprocessing juga bertujuan agar data yang digunakan memiliki dimensi yang lebih kecil dan terstruktur, sehingga dapat diolah lebih lanjut. Tahapan dari preprocessing meliputi: case folding, cleansing, normalisasi bahasa, convert negation, stopwords removal, tokenizing.

2.2.4 Pembobotan TF-IDF

Algoritma TF-IDF (Term Frequency – Inverse Document Frequency) adalah salah satu algoritma yang dapat digunakan untuk menganalisa hubungan antara sebuah frase/kalimat dengan sekumpulan dokumen.

1. Term Frequency (TF)

TF (Term Frequency) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan. Semakin besar jumlah kemunculan suatu term (TF tinggi) dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar. Term Frequency (TF) dapat diformulasikan seperti pada persamaan (2.1) berikut ini:

$$TF = 1 + \log(F_{t,d}), F_{t,d} > 0 \quad (2.1)$$

Dimana nilai $F_{t,d}$ adalah frekuensi term (t) pada document (d). Jadi jika suatu kata atau term terdapat dalam suatu dokumen sebanyak 5 kali maka diperoleh bobot = $1 + \log(5) = 1.699$. Tetapi jika term tidak terdapat dalam dokumen tersebut, bobotnya adalah nol (0).

2. Inverse Document Frequency (IDF)

IDF (Inverse Document Frequency) merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. IDF menunjukkan hubungan ketersediaan sebuah term dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung term yang dimaksud, maka nilai IDF semakin besar. Sedangkan untuk Inverse Document Frequency (IDF) dihitung dengan menggunakan formula seperti pada persamaan (2.2) sebagai berikut:

$$IDF = \log(D/df) \quad (2.2)$$

Dimana D adalah jumlah semua dokumen dalam koleksi sedangkan df adalah jumlah dokumen yang mengandung term (tj).

Jenis formula TF yang biasa digunakan untuk perhitungan adalah TF murni (raw TF). Dengan demikian rumus umum untuk Term Weighting TF-

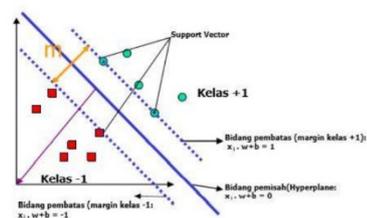
IDF adalah penggabungan dari formula perhitungan raw TF dengan formula IDF dengan cara mengalikan nilai TF dengan nilai IDF seperti pada persamaan (2.3) dibawah ini : [10].

$$W = tf \times idf \quad (2.3)$$

2.2.5 Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan suatu prediksi, baik dalam kasus klasifikasi atau regresi. Metode SVM memiliki prinsip dasar linier classifier yaitu kasus klasifikasi yang dapat dipisahkan secara linier, namun SVM yang dikembangkan dapat bekerja dengan problem non-linier dengan memasukkan konsep kernel pada ruang berdimensi tinggi. Pada ruang berdimensi tinggi, hyperplane yang akan dapat memaksimalkan jarak(margin) antara kelas data. Metode Support Vector Machine (SVM) berakar dari teori pembelajaran statistik yang dapat memberikan hasil yang lebih baik dari metode yang lain.

SVM dapat bekerja dengan data pada non linier dengan menggunakan pendekatan kernel pada fitur awal himpunan data. Fungsi kernel yang digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang lebih tinggi). Konsep SVM adalah pencarian hyperplane terbaik yang berfungsi sebagai pemisah data dari dua kelas pada input space. Hyperplane pemisah terbaik adalah hyperplane yang terletak di tengah diantara dua set objek dari dua kelas. Hyperplane terbaik dapat dicari dengan memaksimalkan margin atau jarak dari dua set objek dari dua kelas yang berbeda. Dapat diasumsikan bahwa kedua belah kelas dapat terpisah secara sempurna oleh hyperplane(linear separable). Akan tetapi, pada umumnya dua belah kelas pada input space tidak dapat terpisah secara sempurna(non linear separable). Untuk mengatasi masalah ini SVM dirumuskan ulang dengan memperkenalkan metode margin[2]. Seperti pada Gambar 1.



Gambar 1. Margin Hyperplane

2.2.6 Evaluasi Klasifikasi

Evaluasi merupakan tahapan terakhir dalam setiap data mining dan klasifikasi teks. Pada tahap ini akan mengevaluasi hasil percobaan, membandingkan dan menganalisis terhadap kinerja klasifikasi teks. Pada penelitian ini evaluasi yang adalah perhitungan Recall, Precision, specificity dan accuracy. Nilai pada semua perhitungan tersebut didasarkan pada hasil confusion matrix yang didapatkan setelah

melakukan pengujian model dan data testing. *Confusion matrix* dapat dilihat pada tabel 2.1.

Tabel 1. Tabel Confusion Matrix

Predicted Class	True Class	
	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Negatif (FN)	True Negatif (TN)

Keempat parameter diatas digunakan untuk menghitung 3 metode evaluasi yakni:

1. *Recall*, yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah seluruh dokumen relevan. *Recall* memiliki rumusan seperti pada persamaan (2.4).

$$Recall = \frac{TP}{TP + FP} \times 100\% \tag{2.4}$$

2. *Precision*, yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah dokumen yang terkenali. *Precision* memiliki rumusan seperti pada persamaan (2.5).

$$Precision = \frac{TP}{TP + FN} \times 100\% \tag{2.5}$$

3. *Accuracy*, yaitu seberapa akurat sistem dapat mengklasifikasikan secara benar. *Accuracy* memiliki rumusan seperti pada persamaan (2.6).

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \tag{2.6}$$

Nilai *recall*, *precision* dan *accuracy* dinyatakan dalam persen. Semakin tinggi persentase ketiga nilai tersebut menunjukkan semakin baik kinerja sistem klasifikasi teks otomatis.

3. METODE PENELITIAN

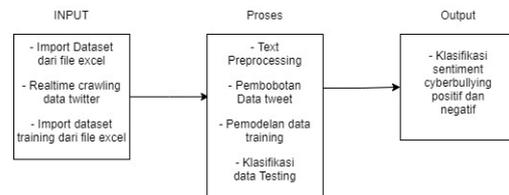
3.1. Analisis Sistem

Dalam pengembangan sistem analisis sentimen cyberbullying di twitter ini user akan melakukan crawling data dari twitter untuk mendapatkan dataset berupa konten tweet yang mengandung abusive word dan berpotensi menjadi konten cyberbullying dengan maksimal data tidak lebih dari 100 tweets Bahasa Indonesia dalam rentang waktu 7 hari terakhir sejak pengambilan data. Setelah dataset didapatkan, maka selanjutnya akan dilakukan proses text preprocessing, adapun tahapan preprocessing yang akan dilakukan pada penelitian ini yaitu Case folding, Cleansing, Normalisasi Bahasa, Convert Negation, Stopword Removal, Tokenisasi.

Sistem yang akan dibuat berbasis web dengan menggunakan metode klasifikasi dari Sentiment Analisis dengan algoritma Support Vector machine (SVM) akan menampilkan hasil klasifikasi sentiment dari setiap tweet dalam dataset yang dimasukkan. Sehingga dengan adanya sistem ini nantinya diharapkan dapat membantu kinerja polisi virtual POLRI ini dalam mengidentifikasi tweet yang berpotensi melanggar pidana UU ITE tersebut dan secara tidak langsung akan meminimalisir pengguna untuk membuat konten Twitter yang mengandung unsur cyberbullying.

3.2. Blok Diagram Sistem

Dalam blok diagram sistem, menunjukkan bagaimana alur dari sistem akan berjalan mulai dari input apa saja yang akan diberikan kemudian proses yang dilakukan sampai output hasil akhir yang akan diberikan. Bisa dilihat pada gambar diagram 2.

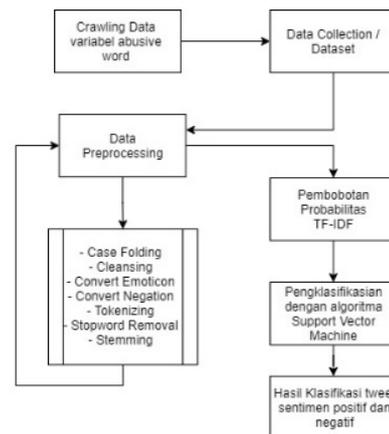


Gambar 2. Blok diagram system

3.3. Desain Arsitektur Sistem

Sistem ini dimulai dengan melakukan crawling data secara otomatis menggunakan twitter API untuk mendapatkan data collection atau dataset tweets dengan dalam Bahasa Indonesia. Sebelum melakukan tahap klasifikasi dilakukan tahap teks preprocessing yakni proses reduksi teks yang bertujuan membuang kata-kata atau term-term yang tidak memiliki kontribusi atau bobot yang mempengaruhi tahap selanjutnya.

Data yang telah melauai tahapan Preprocessing telah siap untuk diolah. Pada data mentah tersebut akan dilakukan proses pembobotan pada setiap kata (term) dan memberikan hasil akhir berupa bobot TF-IDF. Hasil dari pembobotan ini yang akan digunakan dalam proses klasifikasi dengan metode SVM. Output dari sistem ini merupakan data dengan nilai output positif atau negatif yang diklasifikasikan oleh sistem berdasarkan nilai probabilitas yang ada. Flowchart desain arsitektur sistem ini dapat dilihat pada gambar 3.

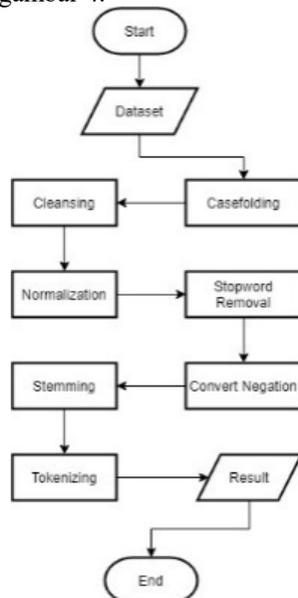


Gambar 3. Desain Arsitektur Sistem

3.4 Pemodelan Proses Text Preprocessing

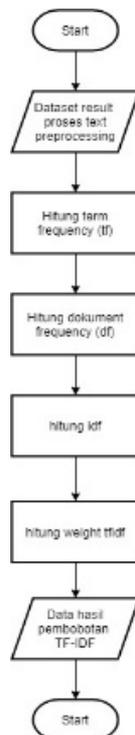
Sebelum melakukan proses klasifikasi sentiment cyberbullying dilakukan proses text preprocessing terlebih dahulu yakni proses reduksi teks yang bertujuan membuang kata-kata atau term-term yang tidak memiliki kontribusi atau bobot yang

mempengaruhi tahap selanjutnya. Proses text preprocessing yang dilakukan diantaranya yakni casefolding, cleansing, normalization, stopword removal, convert negation, stemming, dan tokenizing, akan menghasilkan query term yang akan digunakan pada tahap selanjutnya. Proses ini digunakan untuk mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi. Data yang dimasukkan pada tahap ini masih berupa data mentah. Proses text preprocessing terdiri dari beberapa tahap yang dapat dilihat pada gambar 4.



Gambar 4. Flowchart Text Preprocessing

3.5 Pemodelan Proses Pembobotan TF-IDF



Gambar 5. Flowchart Pembobotan TF-IDF

Pada penelitian ini pembobotan diperoleh dari frekuensi sebuah kata yang terdapat di dalam dokumen tweet atau jumlah kemunculan term dalam satu dokumen term frequency (tf) yang dapat dihitung dengan persamaan (2.1) dan sebuah kata di dalam kumpulan dokumen atau jumlah kemunculan term dalam koleksi dokumen inverse document frequency (idf). Nilai idf sebuah term (kata) dapat dihitung menggunakan persamaan (2.2). Untuk menghitung bobot (W) masing-masing dokumen terhadap setiap term (kata) dapat menggunakan persamaan (2.3). Adapun flowchart dari proses pembobotan tf-idf dapat dilihat pada Gambar 5.

3.6 Pemodelan Proses Support Vector Machine

Data pada ruang input (input space) berdimensi d dinotasikan dengan $x_i \in \mathbb{R}^d$ sedangkan label kelas dinotasikan dengan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, n$. Dimana n adalah banyaknya data. Untuk mencari nilai x_i dan y_i dapat dilakukan ketika sudah didapatkan nilai tiap kata (term) dari pembobotan tf-idf dan label tweet dari anotator. Hasil dari pembobotan tf-idf diubah ke dalam bentuk format data svm light, sedangkan hasil pelabelan dari anotator menjadi label data svm.

Untuk mendapatkan nilai a_i , langkah pertama adalah mengubah setiap kalimat menjadi nilai vektor (support vector) = $\begin{bmatrix} x \\ y \end{bmatrix}$. Kemudian nilai vektor dari setiap kalimat dimasukkan ke persamaan (3.6) kernel trick phi berikut:

$$D_i = \phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{x_n^2 + y_n^2} - x + (x - y)^2 \\ \sqrt{x_n^2 + y_n^2} - y + (x - y)^2 \end{bmatrix} \quad (3.6)$$

Nilai x didapatkan dari persamaan (3.7) kernel linear untuk x berikut:

$$\sum_{i=1, j=1}^1 x_i x_j^T, (i, j = 1, \dots, n) \quad (3.7)$$

Nilai y didapatkan dari persamaan (3.8) kernel linear untuk y berikut:

$$\sum_{i=1, j=1}^1 y_i y_j^T, (i, j = 1, \dots, n) \quad (3.8)$$

Untuk mendapatkan jarak tegak lurus yang optimal dengan mempertimbangkan vektor positif, maka hasil perhitungan dari substitusi nilai x dan nilai y ke persamaan (3.6) diberi nilai bias = 1. Kemudian cari parameter a_i , dengan terlebih dahulu mencari nilai fungsi setiap kalimat menggunakan persamaan (3.9), lalu mencari nilai a_i pada persamaan linear menggunakan persamaan (3.10) dengan memperhatikan $i, j = 1, \dots, n$ berikut :

$$\sum_{i=1, j=1}^1 a_i D_i^T D_j \quad (3.9)$$

$$\sum_{i=1, j=1}^1 a_i D_i^T D_j = y_i \quad (3.10)$$

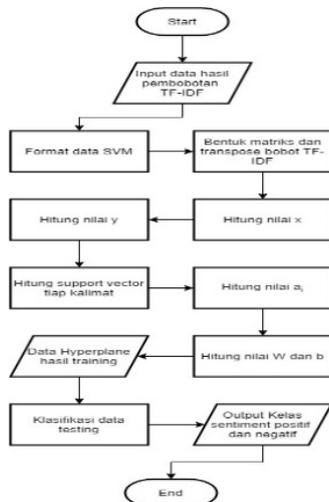
Setelah parameter a_i didapatkan, kemudian masukkan ke persamaan (3.11) berikut :

$$\tilde{W} = \sum_{i=1}^n a_i D_i \quad (3.11)$$

Hasil yang didapatkan menggunakan persamaan (3.11), selanjutnya digunakan persamaan (3.12) untuk mendapatkan nilai w dan b :

$$y = w \cdot x + b \quad (3.12)$$

Sedemikian sehingga didapatkanlah nilai w dan nilai b atau nilai hyperplane untuk mengklasifikasikan kedua kelas. Untuk alur pemodelan metode *support vector machine* ini didalam sistem dapat dilihat pada gambar 6.



Gambar 6. Flowchart Metode SVM

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Sistem

Sistem ini diimplementasikan pada platform web browser menggunakan bahasa pemrograman PHP dan didesain dengan menggunakan HTML, CSS dan Javascript serta menggunakan MySQL sebagai Database Management system. Dalam proses implementasi, akan digunakan 100 data testing yang sudah didapatkan dalam proses crawling baik realtime maupun dalam dataset spreadsheet excel. Untuk data training digunakan dataset dari spreadsheet excel yang telah diberikan label sentiment cyberbullying positif dan negatif oleh anotorator sebanyak 150 data tweet.

4.2. Halaman Import Data Train

Pada halaman import dataset training ini disajikan tombol import dataset yang dapat digunakan user untuk menginputkan dataset tweet yang berupa file spreadsheet excel yang telah diberikan label sentiment cyberbullying positif dan negatif pada masing-masing tweet pada dataset oleh anotorator. Halaman import dataset training dapat dilihat pada gambar 7.

Import Training Data

No	Tweet	label
1	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!	Cyberbullying
2	RT USER: USER siapa yang telat ngasih tau elu/edan sarap gue bergaul dengan cigax jiffa callis sama siapa noh licew juga?	Non-Cyberbullying
3	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter	Non-Cyberbullying
4	USER USER AKU ITU AKU/nyiku TAU MATAMU SIPIT TAPI DILAT DARI MANA ITU AKU!	Non-Cyberbullying
5	USER USER Kaum cebong kapir udah keliatan dongkronya dari awal tambah dongkro lagi hahahah!	Cyberbullying
6	USER Ya bani taplak dkk !v0y0fy98y84yf0y0fy98y84yf0y0fy98y84yf0y0fy98y84	Cyberbullying
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jaban	Non-Cyberbullying

Gambar 7. Halaman Import Data Train

4.3. Halaman Import Data Testing

Pada halaman import dataset testing ini disajikan tombol import dataset yang dapat digunakan user untuk menginputkan dataset tweet yang berupa file spreadsheet excel yang didapatkan melalui proses crawling dengan tools. Halaman import dataset testing dapat dilihat pada gambar 8.

Import Dataset

No	Tweet
1	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!
2	RT USER: USER siapa yang telat ngasih tau elu/edan sarap gue bergaul dengan cigax jiffa callis sama siapa noh licew juga?
3	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter
4	USER USER AKU ITU AKU/nyiku TAU MATAMU SIPIT TAPI DILAT DARI MANA ITU AKU!
5	USER USER Kaum cebong kapir udah keliatan dongkronya dari awal tambah dongkro lagi hahahah!
6	USER Ya bani taplak dkk !v0y0fy98y84yf0y0fy98y84yf0y0fy98y84yf0y0fy98y84
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jaban
8	Gue baru aja kelar re-watch Aldoah Zero!!! paling kampret emang endingnya! 2 karakter utama cowonya kena friendzone bray! XD URL
9	Nah admin belajar satu lagi port terbaik nak makan Ais Kepala Milo, Ais Kepala Horlicks atau Cendol Topping kaw kaw. 877; Diket mano tu ? Geni

Gambar 8. Halaman Import Data Testing

4.4. Halaman Realtime Crawling

Pada halaman realtime crawling ini disajikan tombol crawling yang dapat digunakan user untuk melakukan crawling data twitter secara realtime dengan menginputkan query tertentu. Realtime crawling data twitter ini memanfaatkan akses dari twitter API untuk mendapatkan data berdasarkan query yang diinputkan. Halaman realtime crawling dapat dilihat pada gambar 9.

Real-time Crawling Twitter

Crawling

Query:

Close

No	Tweet
1	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!
2	RT USER: USER siapa yang telat ngasih tau elu/edan sarap gue bergaul dengan cigax jiffa callis sama siapa noh licew juga?
3	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter
4	USER USER AKU ITU AKU/nyiku TAU MATAMU SIPIT TAPI DILAT DARI MANA ITU AKU!
5	USER USER Kaum cebong kapir udah keliatan dongkronya dari awal tambah dongkro lagi hahahah!
6	USER Ya bani taplak dkk !v0y0fy98y84yf0y0fy98y84yf0y0fy98y84yf0y0fy98y84
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jaban
8	Gue baru aja kelar re-watch Aldoah Zero!!! paling kampret emang endingnya! 2 karakter utama cowonya kena friendzone bray! XD URL
9	Nah admin belajar satu lagi port terbaik nak makan Ais Kepala Milo, Ais Kepala Horlicks atau Cendol Topping kaw kaw. 877; Diket mano tu ? Geni

Gambar 9. Halaman Realtime Crawling

4.5. Halaman Text Preprocessing

Pada halaman text preprocessing ini diajikan tombol untuk melakukan proses text preprocessing pada dataset training yang telah diinputkan pada halaman sebelumnya. Pada proses text preprocessing ini dilakukan beberapa tahap yang telah disebutkan pada analisis dan perancangan sistem sebelumnya. Setelah proses selesai maka data hasil akan ditampilkan pada table halaman text preprocessing. Halaman text preprocessing dapat dilihat pada gambar 10.

Dengan data hasil pengujian skenario kedua yang menggunakan 66 data dan percobaan sebanyak 10 kali maka dilakukanlah perhitungan rata-rata waktu dan kecepatan proses dengan hasil bahwa untuk memproses 66 data, sistem membutuhkan rata-rata waktu 121821,7 ms dan kecepatan pemrosesan data 0,000542 data per milisecond.

3. Pengujian Kecepatan dengan 100 data

Tabel 3. Pengujian Kecepatan dengan 100 data

Trial	Waktu
1	122037
2	139617
3	107951
4	113102
5	42310
6	116324
7	108801
8	108396
9	49035
10	103429

Dengan data hasil pengujian skenario pertama yang menggunakan 100 data dan percobaan sebanyak 10 kali maka dilakukanlah perhitungan rata-rata waktu dan kecepatan proses dengan hasil bahwa untuk memproses 100 data, sistem membutuhkan rata-rata waktu 101100,2 ms dan kecepatan pemrosesan data 0,000989 data per milisecond.

4.9. Pengujian Evaluasi Klasifikasi

Pengujian evaluasi klasifikasi bertujuan untuk mengecek kebenaran hasil klasifikasi dengan menghitung berapa angka yang diperoleh. Acuan yang dipakai untuk tahap evaluasi klasifikasi ini yakni perhitungan tabel confusion matrix. Dalam pengujian ini dilakukan dua skenario yaitu skenario 1 dengan 100 data train dan skenario 2 dengan 200 data train. Dengan data test dari dokumen tweet sebanyak 100 dokumen yang telah didapatkan dengan perbandingan label dari anotator dan label yang didapatkan dari proses pada sistem yang dapat dilihat pada tabel 4. dan 5.

Tabel 4. Perhitungan Confusion Matrix Skenario 1

Predicted Class	True Class		Total
	Positif	Negatif	
Positif	23	13	36
Negatif	17	47	64
Total	40	60	100

Setelah melakukan perhitungan confusion matrix maka didapatkanlah nilai-nilai yang akan digunakan dalam proses perhitungan recall dengan persamaan (2.4), precision dengan persamaan (2.5) dan accuracy dengan persamaan (2.6). Berikut perhitungan untuk masing evaluasi yang dilakukan:

$$Recall = \frac{TP}{TP + FP} \times 100\% = \frac{23}{23 + 13} \times 100\% = 64\%$$

$$Precision = \frac{TP}{TP + FN} \times 100\% = \frac{23}{23 + 17} \times 100\% = 58\%$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100\% = \frac{47 + 23}{47 + 23 + 17 + 13} \times 100\% = 70\%$$

Berdasarkan tabel 4., diperoleh dokumen dengan prediksi positif dan faktanya positif sebanyak 23 tweet, dokumen dengan prediksi positif dan faktanya negatif sebanyak 13 tweet, dokumen dengan prediksi negatif dan faktanya negatif sebanyak 47 tweet, serta dokumen dengan prediksi negatif dan faktanya positif sebanyak 17 tweet. Berdasarkan hasil perhitungan evaluasi rata-rata akurasi dengan dokumen sebanyak 100 tweet diperoleh hasil recall 64%, precision 58%, dan accuracy sebesar 70%.

Tabel 5. Perhitungan Confusion Matrix Skenario 2

Predicted Class	True Class		Total
	Positif	Negatif	
Positif	25	10	35
Negatif	15	50	65
Total	40	60	100

Setelah melakukan perhitungan confusion matrix maka didapatkanlah nilai-nilai yang akan digunakan dalam proses perhitungan recall dengan persamaan (2.4), precision dengan persamaan (2.5) dan accuracy dengan persamaan (2.6). Berikut perhitungan untuk masing evaluasi yang dilakukan:

$$Recall = \frac{TP}{TP + FP} \times 100\% = \frac{25}{25 + 10} \times 100\% = 71\%$$

$$Precision = \frac{TP}{TP + FN} \times 100\% = \frac{25}{25 + 15} \times 100\% = 63\%$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100\% = \frac{50 + 25}{50 + 25 + 15 + 10} \times 100\% = 75\%$$

Berdasarkan tabel 4.5, diperoleh dokumen dengan prediksi positif dan faktanya positif sebanyak 25 tweet, dokumen dengan prediksi positif dan faktanya negatif sebanyak 10 tweet, dokumen dengan prediksi negatif dan faktanya negatif sebanyak 50 tweet, serta dokumen dengan prediksi negatif dan faktanya positif sebanyak 15 tweet. Berdasarkan hasil perhitungan evaluasi rata-rata akurasi dengan dokumen sebanyak 200 tweet diperoleh hasil recall 71%, precision 63%, dan accuracy sebesar 75%.

Sistem analisis sentiment cyberbullying yang dibangun telah mampu mentransformasikan dan mengklasifikasikan sentiment cyberbullying yang berupa teks dari twitter dan menampilkan informasi berupa data konten tweet dan hasil klasifikasi tweet tersebut apakah mengandung unsur cyberbullying atau non-cyberbullying. Tingkat keakuratan yang dihasilkan dengan menggunakan metode support vector machine mencapai nilai 70% sehingga dapat dikatakan metode ini dapat digunakan sebagai metode analisis sentiment yang baik. Tingkat akurasi akan semakin baik bila data latih yang digunakan lebih banyak.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian terkait analisis sentiment cyberbullying pada media sosial twitter menggunakan metode support vector machine ini maka dapat disimpulkan bahwa:

1. Berdasarkan hasil dari pengujian blackbox testing menunjukkan bahwa sistem yang dibangun telah mampu mentransformasi dan mengklasifikasikan konten tweet yang mengandung unsur cyberbullying dan non-cyberbullying didalamnya.
2. Penggunaan metode *support vector machine* pada penelitian telah mampu melakukan klasifikasi dengan baik.
3. Penggunaan 100 data dokumen tweet untuk melakukan training dan membentuk model klasifikasi menghasilkan nilai *recall* 64%, *precision* 58% dan tingkat *accuracy* sebesar 70%
4. Penggunaan 200 data dokumen tweet untuk melakukan training dan membentuk model klasifikasi menghasilkan nilai *recall* 71%, *precision* 63% dan tingkat *accuracy* sebesar 75%.
5. Jumlah data training dan ketepatan proses *text preprocessing* dalam menghilangkan dan membersihkan *noise* dokumen tweet mempengaruhi tingkat akurasi dan kinerja klasifikasi sentimen cyberbullying menggunakan metode *support vector machine* ini.
6. Hasil dari penelitian ini juga dapat digunakan untuk mengedukasi masyarakat khususnya pengguna twitter agar lebih mempertimbangkan konten opini tweet maupun komentar yang mereka tuliskan saat menggunakan media sosial.

5.2. Saran

Berdasarkan penelitian yang telah dilakukan, maka penulis dapat memberikan saran-saran untuk pengembangan selanjutnya antara lain:

1. Menambah jumlah data training yang digunakan untuk membentuk model klasifikasi agar tingkat akurasi yang didapatkan lebih tinggi.
2. Meningkatkan proses pembersihan dokumen tweet pada *text preprocessing* agar hasil lebih maksimal.
3. Menambahkan metode ekstraksi fitur lainnya agar mendapatkan bobot yang maksimal dan fitur yang tepat.
4. Meningkatkan kinerja algoritma *support vector machine* pada sistem agar waktu yang dibutuhkan untuk proses klasifikasi lebih cepat.
5. Sistem dapat dikembangkan untuk analisis sentiment dengan topik berbeda lainnya.

DAFTAR PUSTAKA

- [1] Maria Mega Mala Olhang, Sentot Achmadi, F.X Ari Wibisono, 2020. *Analisis Sentimen Pengguna Twitter Terhadap COVID-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (NBC)*. JATI (Jurnal Mahasiswa Teknik Informatika) Vol. 4 No. 2 September 2020 Teknik Informatika ITN Malang.
- [2] Amar P. Natasuwarna, 2020. *Seleksi Fitur Support Vector Machine Pada Analisis Sentimen Keberlanjutan Pembelajaran Daring*. Techno.COM Vol. 19 No. 4 November 2020: 437 – 448 Sistem informasi STMIK Pontianak.
- [3] Wanda Athira Luqyana, Imam Cholissodin, Rizal Setya Perdana. 2018. *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer vol. 2 No. 11 November hlm. 4704-4713 e-ISSN: 2548-964X, Universitas Brawijaya.
- [4] Fajar Agus Maulana, Iin Ernawati. 2020. *Analisa Sentimen Cyberbullying di Jejaring Sosial Twitter dengan Algoritma Naive Bayes*. Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) ISBN 978-623-93343-1-4, Jakarta-Indonesia.
- [5] Nurrun Muchammad Shiddieqy Hadna, Paulus Insap Santosa, Wing Wahyu Winarno. 2016. *Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter*. Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016) ISSN: 2089-9815, Yogyakarta.
- [6] Araque, O. et al. 2017. *Enhancing deep learning sentiment analysis with ensemble techniques in social applications*. Expert Systems with Applications. Elsevier Ltd, 77, pp. 236–246. doi: 10.1016/j.eswa.2017.02.002.
- [7] Zhang, L., Wang, S. and Liu, B. 2018. *Deep learning for sentiment analysis: A survey*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), pp. 1–25. doi: 10.1002/widm.1253.
- [8] Kowalski, R. M. et al. (2016) ‘Cyberbullying among college students with disabilities’, Computers in Human Behavior. Elsevier Ltd, 57, pp. 416–427. doi: 10.1016/j.chb.2015.12.044.
- [9] A. F. Hidayatullah and A. SN, “Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter,” Seminar Nasional Informatika, vol.2 no.4, pp. 1–8, Okt 3,2015.
- [10] <https://informatikalogi.com/term-weighting-tf-idf/> , diakses tanggal 05 April 2021, pukul 13:00 WITA.