ANALISIS SENTIMEN CYBERBULLYING PADA KOMENTAR FACEBOOK DENGAN METODE KLASIFIKASI *SUPPORT VECTOR MACHINE*

Rafli Muhammad Kamal[1,] Ednawati Rainarli[2]
[1,2] Program Studi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia
Jl. Dipati Ukur 114 Bandung
E-mail : raflimkamal97@email.unikom.ac.id[1],ednawati.rainarli@email.unikom.ac.id[2]

ABSTRAK

Seiring maraknya penggunaan facebook sebagai media sosial, tentu semakin beragamlah komentar yang ditulis seseorang dalam suatu postingan. Terkadang tanpa disadari para pengguna facebook menuliskan komentar yang mengandung unsur bullying. Tentunya akan berdampak buruk pada dirinya sendiri maupun orang lain, oleh karena itu perlu adanya analisa terkait komentar facebook. Pendekatan machine learning yang dapat digunakan untuk mendeteksi cyberbullying adalah analisis sentimen. Dalam penelitian ini akan membahas analisis sentimen dengan menggunakan metode Support Vector Machine (SVM). Prosesnya dengan mengklasifikasikan sentimen yang positif (tidak mengandung unsur bullying) atau negatif (yang mengandung unsur bullying). Di dalam tahapan praproses ditambahkan proses normalisasi kata dengan tujuan untuk mengatasi penggunaan kata yang tidak baku dari komentar yang akan diproses. Untuk pengujian akurasi dilakukan 2 kali pengujian. Pengujian I menggunakan 100 data latih dan 100 data uji dan Pengujian II menggunakan 100 data latih dan 50 data uji yang berasal dari data komentar para pengguna facebook. Hasil pengujian akurasi yang dilakukan menunjukan bahwa SVM bisa memiliki tingkat persentase cukup tinggi pada kasus analisis sentiment bisa mencapai 96% dengan menggunakan fungsi kernel RBF. Dapat disimpulkan bahwa metode klasifikasi Support Vector Machine bekerja baik pada kasus analisis sentimen cyberbullying pada komentar facebook.

Kata kunci : analisis sentimen, komentar, preprocessing, TF-IDF, support vector machine, RBF

1. PENDAHULUAN

Facebook merupakan salah satu media sosial yang banyak digunakan di Indonesia. Permasalahan yang seringkali terjadi adalah tindakan *cyberbullying* yang muncul pada facebook. Tidak banyak pengguna facebook menyadari bahwa ulasan atau komentar yang dituliskan kepada seseorang ataupun kelompok merupakan tindakan bullying. Bebasnya pengguna facebook dalam memposting status dan berkomentar menjadi salah satu alasan munculnya konten-konten yang bersifat melukai dan berakibat perundungan (bullying) di media sosial.

Cara untuk mendeteksi komentar yang mengandung unsur bullying dapat menggunakan pendekatan machine learning yaitu dengan analisis sentimen. Analisis sentimen sangat diperlukan untuk menyaring komentar-komentar di media sosial. Proses yang dilakukan dalam analisis sentimen adalah dengan mengklasifikasikan informasi ke dalam kelas sentimen positif dan kelas sentimen negatif. Informasi akan diklasifikasikan ke dalam kelas positif apabila informasi yang disampaikan bernilai baik atau setuju terhadap sesuatu. Sebaliknya, informasi diklasifikasikan ke dalam kelas negatif apabila informasi yang disampaikan bernilai tidak baik atau tidak setuju[1].Untuk mengklasifikasikan komentar maka dibutuhkan pendekatan machine learning yang dapat memisahkan antara komentar yang mengandung cyberbullying dan tidak mengandung cyberbullying. Oleh karena itu, dipilihlah algoritma support vector machine untuk penelitian ini.

Pada penelitian sebelumnya yang berguna dalam mendukung pelaksanaan penelitian terkait adalah penelitian mengenai analisis sentimen yang menggunakan metode Support Vector Machine oleh Petrik[2]. Penelitian lain membahas mengenai kategorisasi teks Bahasa Indonesia oleh Wulandini dan Nugroho, didapat algoritma SVM memiliki akurasi yang paling tinggi yaitu 92,5 % dibanding algoritma yang lain seperti K-Nearest Neighbors, Naïve Bayes Classification, Information Fuzzy Networks[3].

Didapatkan kesimpulan bahwa penggunaan metode Support Vector Machine (SVM) memberikan hasil akurasi paling baik dibandingkan dengan metode lainnya. Berdasarkan hal itu dalam penelitian ini akan digunakan metode metode Support Vector Machine untuk mendeteksi adanya sentimen serta mengetahui nilai akurasi pada metode yang digunakan..

2. TINJAUAN PUSTAKA

2.1 Perundungan(*Bullying*)

Seiring berkembangnya zaman, teknologi pun ikut berkembang. Berkembangnya teknologi memberi pengaruh terhadap kehidupan sosial. Seperti pada tindakan bullying. Mulanya tindakan bullying menyerang secara fisik maupun psikologi secara langsung, namun kini tindakan tersebut dapat dilakukan pada dunia maya yang dikenal dengan *cyberbullying*. *Cyberbullying* merupakan suatu tindakan tidak menyenangkan yang dilakukan secara sengaja dan terus menerus melalui teks elektronik[4].

Cyberbullying merupakan salah satu bentuk serangan bersifat dengki untuk memberi kepuasan atau kesenangan pelaku dengan cara merendahkan orang lain. Perempuan lebih banyak terlibat dalam *cyberbullying*, baik sebagai pelaku maupun sebagai korban. Dimana 50% korban *cyberbullying* umumnya tidak mengetahui identitas pelaku bully meski hanya gender. *Cyberbullying* merupakan tindakan yang dapat menyerang psikologi, emosional, dan trauma sosial[5].

Hal ini menjadi permasalahan yang sangat serius ketika *cyberbullying* menjadi hal yang lebih berbahaya daripada bully pada umumnya. Karena dampak yang lebih berbahaya ketika perasaan takut dan depresi lalu berubah menjadi frustasi dan pemarah, bahkan parahnya dapat mengakibatkan korban melakukan tindakan bunuh diri. Pemilihan kata-kata yang diucapkan menjadi salah satu kunci apakah seseorang mengarah pada tindakan *bullying* atau tidak[6].

2.2 Text Mining

Text mining adalah disiplin keilmuan yang berfokus pada pencarian informasi, data mining, machine learning, statistik, dan komputasi linguistik[7].

Sumber data yang digunakan pada text mining adalah kumpulan dari teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen.

Beberapa tahapan proses pokok dalam text mining, yaitu pemrosesan awal teks (text preprocessing), transformasi teks (text transformation) atau (Feature Generation), pemilihan fitur (feature selection), dan penemuan pola text atau data mining (pattern discovery).

2.2.1 Text Preprocessing

Preprocessing merupakan tahap awal dari text mining untuk mengubah data sesuai dengan format yang dibutuhkan. Proses ini dilakukan untuk menggali, mengolah dan mengatur infomasi dan untuk menganalisis hubungan tekstual dari data terstruktur dan data tidak terstruktur[8]. 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.

a) Case Folding

Tahapan awal adalah case folding yang merupakan tahapan preprocessing yang dilakukan untuk menyeragamkan karakter pada data (dokumentasi/teks). Karena tidak semua dokumen teks hanya menggunakan huruf kapital, Semua huruf dapat dirubah menjadi huruf besar (uppercase) atau huruf kecil (lower case) [1]

b) Cleansing

Cleansing merupakan proses untuk pembersihan kata selain karakter ‘a’ sampai ‘z’ dan spasi berlebih akan dihilangkan. Pembersihan kata bertujuan untuk mengurangi noise. Untuk menanggulangi kelebihan spasi setelah cleansing, maka dilakukan penghapusan spasi yang berlebihan dari sebelum dan sesudah kata (remove whitespace).

c) Normalisasi Bahasa

Pada tahap preprocessing dilakukan normalisasi bahasa terhadap kata yang tidak baku. Tahapan ini bertujuan untuk mengembalikan bentuk penulisan dari masing-masing kata yang sesuai dengan KBBI. Proses ini dilakukan dengan mencocokkan setiap kata pada dokumen data latih dan data uji dengan kata yang ada pada kamus[9].

d) Convert Negation

Kata yang bersifat negasi, akan merubah nilai sentimen dari suatu komentar. Ketika banyak kata negasi adalah ganjil, maka sentimen komentar tersebut akan dirubah. Kata “tidak” akan digunakan untuk mengganti kata yang bersifat negasi diluar kamus KBBI[10]. Untuk menanggulangi kelebihan spasi setelah convert negation, maka dilakukan penghapusan spasi yang berlebihan dari sebelum dan sesudah kata (remove whitespace).

e) Stopword Removal

Stopword Removal bertujuan untuk menghilangkan kata (term) yang dianggap tidak dapat memberikan pengaruh dalam menentukan suatu kategori tertentu dalam suatu dokumen. Sebelum dilakukan proses stopword removal, terlebih dahulu dibuat kata-kata yang termasuk ke dalam stopword. Stopword merupakan daftar kata umum yang tidak memiliki arti penting dan tidak digunakan. Pada proses ini kata umum akan dihapus untuk mengurangi jumlah kata yang disimpan oleh sistem.

f) Tokenisasi

Tokenisasi adalah proses untuk memotong dokumen menjadi pecahan kecil yang dapat berupa bab, sub-bab, paragraf, kalimat, dan kata (token). Pada proses ini akan menghilangkan whitespace.

2.3 Pembobotan TF IDF

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode yang digunakan untuk menghitung bobotan setiap kata yang telah diekstrak. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model *term frequency (tf)* dan *inverse document frequency*

(idf), dimana *term frequency (tf)* merupakan proses untuk menghitung jumlah kemunculan term dalam satu dokumen dan *inverse document frequency (idf)* digunakan untuk menghitung term yang muncul di berbagai dokumen(komentar) yang dianggap sebagai term umum, yang dinilai tidak penting[11]. Proses awal yang dilakukan dalam pembobotan TF-IDF dilakukan dengan menghitung *term frequency* ($tf_{i,d}$). Dimana t menunjukkan term dalam dokumen d yang berfungsi untuk menunjukkan kemunculan term t pada dokumen d. Hal ini berpengaruh dalam bobot term yang akan semakin tinggi ketika banyak term yang muncul dalam suatu dokumen. Nilai dari *tf* akan dihitung bobotnya dengan rumus *weighting term frequency (*W_{tf}*)*. Rumus tersebut ditunjukkan pada Persamaan 1.

$$\boldsymbol{W_t f_t, d} = \begin{cases} \mathbf{1 + log10\ tf_{t,d}\ \ if\ tf_{t,d} > 0} \\ \mathbf{0,\ otherwise} \end{cases} \quad (1)$$

Banyaknya kata yang muncul pada dokumen, umumnya merupakan nilai *term frequency* dari kata yang tidak penting. Untuk menghindari pembobotan pada kata tidak penting maka digunakan pembobotan *document frequency* yang bermaksud untuk menghitung jumlah dokumen yang mengandung term t.

Dari nilai term pada setiap dokumen yang telah ditemukan akan dilakukan proses kebalikan dari pembobotan *document frequency*. Proses pembobotan ini disebut dengan *inverse document frequency*, yang menyatakan bahwa frekuensi dari term yang rendah pada banyak dokumen akan memberikan bobot paling tinggi. Perhitungan ini ditunjukkan dengan Persamaan 2.

$$\boldsymbol{idft} = \mathbf{log}\ \frac{N}{dft} \quad (2)$$

Dimana :
idf_t = bobot inverse dari nilai df
N = jumlah dokumen pada kumpulan dokumen
df_t = jumlah dokumen yang mengandung term
Perhitungan pembobotan TF-IDF merupakan perkalian yang dilakukan dari pembobotan term frequency dengan inverse document frequency. Hal ini ditunjukkan pada Persamaan 3.

$$\boldsymbol{W_{t,d} = W_t tf_{t,d} \times idf_t} \quad (3)$$

Keterangan:
$W_t f_{t,d}$ = bobot kata dalam setiap dokumen
$tf_{t,d}$ = jumlah kemunculan kata t pada dokumen d
idf_t = bobot inverse dari nilai df
$W_{t,d}$ = Pembobotan TF-IDF

2.4 Support Vector Machine

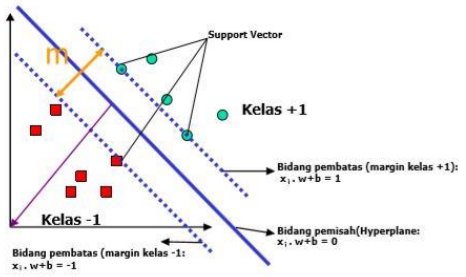
Support Vector Machine (SVM) adalah suatu teknik untuk melakukan suatu prediksi, baik dalam kasus klasifikasi atau regresi. Metode SVM memiliki prinsip dasar *linier classifier* yaitu kasus klasifikasi yang dapat dipisahkan seara linier, namun SVM yang dikembangkan dapat bekerja dengan problem *non-linier* 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 [2].

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[12]. Seperti pada Gambar 1.



Gambar 1. Margin Hyperplane

2.4.1 Klasifikasi Data Linear Separable

Data Linear Separable merupakan data yang dapat dipisahkan misalkan {x1 ... ... xn} adalah data set {+1,-1} adalah label dari kelas dari data ke xn. Pada Gambar 1. **Margin Hyperplane**

dapat dilihat berbagai alternatif bidang pemisah yang dapat memisahkan semua data set sesuai dengan kelasnya. Namun, bidang pemisah terbaik tidak hanya dapat memisahkan data tetapi juga memiliki margin paling besar.

Data latih dinyatakan $(y_i x_i)$ dimana i= 1,2, ..., N, dan $x_i = (x_{i1} x_{i2} \ldots, x_{iq})^T$ merupakan atribut (fitur) set data latih ke-i, yi € {-1,+1} menyatakan label kelas. *Hyperplane* klasifikas linier SVM [3] dinotasikan:

$$\boldsymbol{w}.\boldsymbol{x_i} + \boldsymbol{b} = \quad (4)$$

Keterangan :
w dan b = parameter model.
$w.x_i$ merupakan inner-product dalam antara w dan x_i
Jika x_i masuk ke dalam -1 maka memenuhi pertidaksamaan sebagai berikut:

$$\boldsymbol{w}.\boldsymbol{x_i} + \boldsymbol{b} \leq -\boldsymbol{1} \quad (5)$$

Jika x_i masuk ke dalam +1 maka memenuhi pertidaksamaan sebagai berikut:

$$\boldsymbol{w}.\boldsymbol{x_i} + \boldsymbol{b} \leq +\boldsymbol{1} \quad (6)$$

jika data dalam kelas -1 (xa) terdapat di hyperplane maka persamaan akan terpenuhi untuk dinotasikan dengan seperti berikut:

$$\boldsymbol{w}.\boldsymbol{x_a} + \boldsymbol{b} = \boldsymbol{0} \quad (7)$$

data kelas +1 (xb) akan memenuhi persamaan sebagai berikut:

$$\boldsymbol{w}.\boldsymbol{x_b} + \boldsymbol{b} = \boldsymbol{0} \quad (8)$$

Dengan mengurangi persamaan sebagai berikut:

$$\boldsymbol{w}.(\boldsymbol{x_b} - \boldsymbol{x_a}) = \boldsymbol{0} \quad (9)$$

$x_b - x_a$ merupakan vektor paralel di posisi hyperplane dan diarahkan dari
$x_a \; ke \; x_b$, maka arah w tegak lurus terhadap hyperplane saat inner product bernilai nol.
Formula untuk memberikan label -1 untuk kelas pertama, dan +1 untuk kelas kedua seperti berikut:

$$\boldsymbol{Y} = \begin{cases} +\boldsymbol{1}, \boldsymbol{jika}\; \boldsymbol{w}.\boldsymbol{z} + \boldsymbol{b} \; > \boldsymbol{0} \\ -\boldsymbol{1}, \boldsymbol{jika}\; \boldsymbol{w}.\boldsymbol{z} + \boldsymbol{b} \; < \boldsymbol{0} \end{cases} \quad (10)$$

Hyperplane untuk kelas -1 (garis putus putus) adalah data pada support vector yang memenuhi persamaan:

$$\boldsymbol{w}.\boldsymbol{x_b} + \boldsymbol{b} = \; -\boldsymbol{1} \quad (11)$$

Hyperplane untuk kelas +1 (garis putus putus) memenuhi persamaan:

$$\boldsymbol{w}.\boldsymbol{x_b} + \boldsymbol{b} = \; +\boldsymbol{1} \quad (12)$$

Margin dapat dihitung seperti berikut ini.

$$\boldsymbol{w}.(\boldsymbol{x_b} - \boldsymbol{x_a}) = \boldsymbol{2} \quad (13)$$

Untuk mencari margin (jarak) hyperplane terbaik yang terletak di tengah tengah dua bidang pembatas kelas sama dengan memaksimalkan margin atau jarak antara dua set objek kelas yang berbeda. Maka margin dapat dihitung dengan

$$||\boldsymbol{w}||\boldsymbol{x}\; \boldsymbol{d} = \boldsymbol{2} \text{ atau } \boldsymbol{d} = \frac{\boldsymbol{2}}{||\boldsymbol{w}||} \quad (14)$$

2.4.2 Klasifikasi Data Linear Non Separable

Untuk menyelesaikan problem non-linear, SVM dimodifikasi dengan memasukkan fungsi kernel. Dalam non-linear SVM, pertama-tama data x dipetakan oleh fungsi Φ (x) ke ruang vektor yang berdimensi lebih tinggi. Hyperplane yang memisahkan kedua kelas tersebut dapat dikontruksikan. Selanjutnya bahwa fungsi Φ memetakan tiap data pada input space tersebut ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), sehingga kedua kelas dapat dipisahkan secara linear oleh sebuah hyperplane.

Selanjutnya proses pembelajaran pada SVM dalam menemukan titik-titik support vector, hanya bergantung pada dot product (perkalian titik) dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu $\Phi(x_i).\Phi(x_j)$. Karena umumnya transformasi Φ ini tidak diketahui, dan sangat sulit untuk difahami secara mudah, maka perhitungan dot product (perkalian titik) dapat digantikan dengan fungsi kernel K ($x_i \;\; x_j$) yang mendefinisikan secara implisit transformasi Φ. Hal ini disebut sebagai Kernel Trick, yang dirumuskan:

$$\boldsymbol{K}(\boldsymbol{x_i}\boldsymbol{x_j}) = \; \boldsymbol{\phi}\,(\boldsymbol{x_i}).\boldsymbol{\phi}\,(\boldsymbol{x_j}) \quad (15)$$

Dengan *kernel trick* ini, hanya perlu mengetahui fungsi kernel yang dipakai untuk menentukan *support vector*. Tidak perlu mengetahui wujud dari fungsi nonlinier ∅. Pada kasus SVM non-linear ada beberapa fungsi kernel yang umum digunakan yaitu:

a. **Linear**

$$K(x, x_i) = x_i^T x \quad (16)$$

b. **Polynomial Kernel**

$$K(x, x_i) = (x_i^T x + 1)^d \quad (17)$$

c. **RBF**

$$K(x, x_i) = \exp\{-\gamma||x - x_i||^2, \gamma > 0 \quad (18)$$

d. **Sigmoid Kernel**

$$K(x, x_i) = \tanh[K \; + x_i^T x + \theta] \quad (19)$$

Penggunaan fungsi kernel akan menetukan feature space di mana fungsi klasifier akan dicari. Sepanjang fungsi kernelnya legitimate, SVM akan beroperasi secara benar meskipun tidak tahu seperti apa map yang digunakan, sehingga lebih mudah menemukan fungsi kernel daripada mencari map seperti apa yang tepat untuk melakukan mapping dari input space ke feature space. Pada penerapan metoda kernel, tidak perlu tahu map apa yang digunakan untuk satu per satu data, tetapi lebih penting mengetahui bahwa *dot product* (perkalian titik) dua titik di feature space bisa digantikan oleh fungsi kernel.

Dan prediksi pada set data dengan dimensi fitur yang baru diformulasikan dengan

$$\boldsymbol{f}(\boldsymbol{\phi}(\boldsymbol{x})) = (\boldsymbol{sign}(\boldsymbol{w}.\boldsymbol{\phi}(\boldsymbol{x}) + \boldsymbol{b}) \quad (20)$$

$$= \textstyle\sum_{\boldsymbol{i=1}, \boldsymbol{xi} \in \boldsymbol{SV}}^{\boldsymbol{n}} (\boldsymbol{sign}(\boldsymbol{\alpha_i}\boldsymbol{y_i}\, \emptyset(\boldsymbol{x}).\emptyset(\boldsymbol{x_i}) + \boldsymbol{b})$$

$$= \sum_{i=1, xi \in SV}^{n} (sign(\alpha_i y_i\, K(x.x_i) + b)$$

3. METODE PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini :

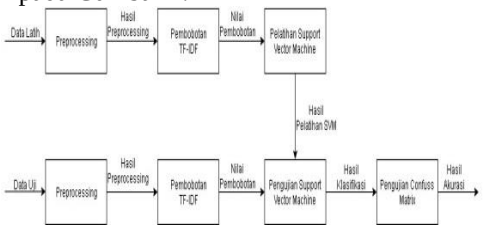
- Identifikasi Masalah
- Pengumpulan data
- Analisis Proses
- Pengujian
- Penarikan Kesimpulan

4. HASIL DAN PEMBAHASAN

Untuk melihat penggunaan metode support vector machine pada kasus analisis sentimen, maka berikut ini adalah analisis proses dari analisis sentimen cyberbullying pada komentar fanspage facebook dengan metode klasifikasi support vector machine, yang meliputi pengumpulan data masukkan, preprocessing, lalu kasifikasi support vector machine.

Tahapan yang digunakan untuk mengetahui adanya kalimat yang mengandung bullying dalam teks pada penelitian ini dibagi menjadi dua tahap, yaitu tahap pelatihan dan pengujian.

Berikut gambaran tahapan proses yang akan dilakukan pada Gambar 2.



Gambar 2. Tahapan Proses yang dilakukan

4.1 Analisis Data Masukkan

Data masukan yang digunakan dalam penelitian ini adalah komentar di facebook dari postingan facebook yang telah ditentukan. Data masukan yaitu 150 data komentar yang akan digunakan, Dimana terdiri dari, 100 data komentar sebagai data latih, dan 50 data komentar menjadi data uji.

Data yang digunakan dalam penelitan ini terdiri dari dua jenis data; yaitu data latih dan data uji.

4.2 Analisis Preprocessing

Analisis preprocessing merupakan tahap awal dan salah satu langkah yang penting dalam sebuah pengklasifikasian sebuah teks. Adapun tahapan preprocessing yang akan dilakukan pada penelitian ini yaitu *Case folding*, *Cleansing*, Normalisasi Bahasa, *Convert Negation, Stopword Remova*l, Tokenisasi. Preprocessing ini dilakukan pada data latih yang menjadi data masukan, hasil dari preprocessing dapat dilihat pada Tabel 1.

Tabel 1 Hasil Preprocessing

Kata
wasit indonesia sulit fokus sudah dapat tekanan semoga mental wasit jugabisadiperbaiki padahal fase grup mantap

4.3 Analisa Pembobotan TF-IDF

Proses awal dilakukan perhitungan term (kata tunggal) pada setiap dokumen, sehingga akan mendapatkan frekuensi term. Selanjutnya adalah menghitung df, karena df adalah banyaknya dokumen dimana munculnya suatu term. Setelah memperoleh nilai df, maka dilakukan perhitungan idf .

Diambil contoh pada kata “wasit”. Maka didapatkan banyak dokumen (N) = 7, dan df = 3. Maka perhitungannya seperti berikut.

$idf_t = \log(\frac{7}{3})$ = 0,368

Selanjutnya untuk mendapatkan bobot term, maka dilakukan perhitungan tf dan idf.

Diperoleh tf = 3, dan idf = 0,477. Maka perhitungannya menjadi seperti berikut.

$Wt = 3 * 0{,}368 = 1{,}104$

Sehingga kata “wasit” memiliki bobot sebesar 1,104.

4.4 Klasifikasi *Support Vector Machine*

Pada tahap klasifikasi *Support Vector Machine* tahap pelatihan dilakukan untuk mendapatkan *hyperplane* yang didapat dari data latih yang telah dimasukkan kemudian data *hyperplane* tersebut dimasukkan ke dalam database. Tahap selanjutnya yang dilakukan adalah tahap pengujian. Data uji dimasukkan melalui proses yang sama seperti data latih yaitu proses *preprocessing*, Pembobotan TF-IDF, lalu hasil dari pengujian ini adalah klasifikasi sentimen berupa komentar facebook dari data uji yang dimasukkan. Lalu untuk pengujian sistem dilakukan metode confuss matrix. Hasil dari pengujian tersebut akan menghasilkan akurasi dari metode *Support Vector Machine.*

4.4.1 Pelatihan *Support Vector Machine*

Pelatihan SVM bertujuan untuk menemukan vektor α, nilai W dan konstanta b untuk mendapatkan hyperplane terbaik. Pada penelitian kali ini, data yang digunakan sebagai pelatihan adalah data komentar. Dalam proses pelatihan SVM, setiap model klasifikasi dilatih dengan dua kelas ke-i dan kelas-j.

Data masukan yang akan digunakan untuk proses pelatihan adalah data dari 2 kelas yang berbeda yang telah melalui preprocessing. Sesuai dengan data masukan, data komentar diberi kelas positif dan negatif, lalu diberikan label kelas 1 atau -1 yang dimana kelas -1 merupakan kelas negatif sedangkan kelas 1 merupakan kelas positif.

$$f(\emptyset(x)) = (sign(w.\emptyset(x)) + b)$$

$$= \sum_{i=1, xi \in SV}^{n} (sign(\alpha_i y_i K(x.x_i)) + b)$$

dimana i= 1,2,.3,..., n=jumlah support vector.
Maka didapatkan bidang pemisah pelatihan:
$f(x_{testing}) = (sign$ (0,255(0,645) + 0,122(0,05311) + 0,243(0,24446) - 0,060(0,2876) - 0,102(0,21687) - 0,123(0,24654) , $K(x_i x_{testing})$) + 0,125)

4.1.2 Pengujian *Support Vector Machine*

Setelah mendapatkan nilai α dan b sebagai model fitur, dan nilai parameter gamma 0,5 (γ) dari proses pelatihan, selanjutnya menguji data uji ke dalam kelas +1 atau -1 dengan model fitur yang sudah di dapat. Data yang digunakan untuk dilakukan pengujian adalah data latih P1,P2, P3,P4,P5,P6sebagai hasil pelatihan data.
menentukan kelas mana komentar yang telah di*testing* menggunakan fungsi hyperplane.

$$f(x_{testing}) = (sign \sum_{i=1}^{n} \alpha_i y_i K(x_i\, x_{testing}) + b)$$

Sehingga

$$f(x_{testing}) = (sign\ (\alpha_1 y_1 K(x_{1,} x_{testing}) + \alpha_2 y_2 K(x_{2,} x_{testing}) + \alpha_3 y_3 K(x_{3,} x_{testing}) + \alpha_4 y_4 K(x_{4,} x_{testing}) + \alpha_5 y_5 K(x_{5,} x_{testing}) + \alpha_6 y_6 K(x_{6,} x_{testing}) + b)$$

$= (sign(0{,}255(\ \ 1) + 0{,}122(0{,}05311) + 0{,}243(0{,}24446) - 0{,}060(0{,}02876) - 0{,}102(0{,}21687) - 0{,}123(0{,}24654) + 0{,}125)$
$= sign(0{,}0819)$
$= +1$
Setelah salah satu kelas komentar melalui tahapan *testing*, menghasilkan fungsi hyperplane +1 yaitu (positif).

4.1.3 Pengujian Akurasi

Hasil pengujian akurasi merupakan gambaran dari skenario pengujian yang dilakukan yaitu pengujian dalam penggunaan nilai variabel di algoritma *support vector machine* serta pengujian akurasi terhadap hasil deteksi teks yang dihasilkan. Pengujian dilakukan dengan menggunakan nilai γ berkisar antara 0,1 hingga 1.

1. Pengujian akurasi I dengan 100 data latih dan 100 data uji(sama dengan data latih). Hasil skenario dapat dilihat pada Tabel 2.

Tabel 2 Hasil Pengujian Akurasi I

γ	Akurasi	*Precision*	*Recall*	*F-measure*
0,1	0,53	0	0	0
0,2	0,71	0,50	0,38	0,55
0,3	0,92	0,97	0,85	0,90
0,4	0,96	0,97	0,94	0,96
0,5	0,97	0,97	0,96	0,97
0,6	0,98	0,98	0,98	0,97
0,7	0,98	0,98	0,98	0,97
0,8	0,99	0,97	0,99	0,98
0,9	0,99	0,97	0,99	0,98
1	100	100	100	100

2. Pengujian akurasi II dengan 100 data latih dan 50 data uji (berbeda dengan data latih). Hasil skenario dapat dilihat pada Tabel 3.

Tabel 3 Hasil Pengujian Akurasi II

γ	Akurasi	*Precision*	*Recall*	*F-measure*
0,1	0,52	0,27	0,52	0,35
0,2	0,60	0,69	0,60	0,53
0,3	0,64	0,69	0,64	0,60
0,4	0,68	0,70	0,68	0,66
0,5	0,65	0,65	0,64	0,62
0,6	0,64	0,65	0,64	0,63
0,7	0,64	0,65	0,64	0,63
0,8	0,67	0,66	0,67	0,63
0,9	0,69	0,67	0,69	0,67
1	0.75	0.73	0.75	0.74

Berdasarkan hasil pengujian akurasi I pada Tabel 2 nilai optimal yang didapat dengan data latih sebanyak 100 latih dan 100 data uji(sama dengan data latih) lalu menggunakan nilai gamma 1 menghasilkan akurasi 100%. Sedangkan berdasarkan hasil pengujian akurasi II pada Tabel 3 nilai optimal yang didapat dengan data latih sebanyak 100 latih dan 50 data uji(berbeda dengan data latih) lalu menggunakan nilai gamma 1 menghasilkan akurasi 75%.

Analisis sentimen cyberbullying dangan metode Support Vector Machine dengan Kernel RBF dilakukan pengujian akurasi sebanyak 2 kali. Pengujian akurasi I menggunakan 100 data latih dan 100 data uji(sama dengan data latih) dan Pengujian akurasi II menggunakan 100 data latih dan 50 data uji(berbeda dengan data latih). Penelitian pernah dilakukan Imelda dan Affandes[13]. Dari hasil uji coba yang dilakukan keduanya,

Aplikasi menunjukkan akurasi stabil pada rentang nilai $0 \leq C \leq 3$ dan $0.01 \leq \gamma \leq 10$ pada data yang belum dilakukan pemilihan fitur dan akurasi stabil pada rentang nilai $0 \leq C \leq 300$ dan $0.01 \leq \gamma \leq 10$. Dengan pencapaian nilai akurasi yang baik maka, penggunaan parameter dapat membuat akurasi menjadi stabil dan hasil ini dapat diterapkan untuk membantu pengguna Twitter untuk melakukan filter terhadap tweet iklan yang terdapat pada akun Twitter mereka[13].

Berdasarkan hasil evaluasi pengujian dengan perhitungan yang dilakukan dengan nilai precision, recall dan F-measures tertera pada pengujian akurasi ke I di Tabel 4.10 dan pengujian akurasi ke II Tabel 4.11. menghasilkan perbedaan nilai yang cukup signifikan. Nilai akurasi yang dihasilkan pada pengujian akurasi I berada di rentang 53% hingga 100%. Sedangkan di pengujian akurasi ke II nilai akurasi berada di rentang 52% hingga 75%. Hasil ini menunjukkan bahwa jumlah data uji yang digunakan dapat berpengaruh pada tingkat akurasi yang dihasilkan untuk mendeteksi teks. Juga berdasarkan penelitian diatas penggunaan parameter gamma dari rentang 0,1 hingga 1 dapat membuat pengaruh terhadap akurasi, yaitu dapat membuat stabil akurasi.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Penelitian deteksi teks bullying di media sosial dengan mengimplementasikan Support Vector Machine pada komentar Facebook berbahasa Indonesia mendapatkan beberapa kekurangan diantaranya :

1. Data set yang diambil merupakan data komentar Facebook yang memiliki emoticon berbeda dengan twitter / belum memiliki kamus sehingga tidak dapat diproses lebih lanjut.
2. Banyaknya kata-kata typo sehingga terjadi kesalahan dalam beberapa proses preprosesing sehingga ikut terolah dikarenakan tidak terdeteksi pada stopwords.
3. Tidak adanya sumber atau referensi secara resmi dalam list kata-kata yang mengandung bullying berbahasa Indonesia.
4. Tahapan preprocessing mempengaruhi analisis sentimen karena dapat menghasilkan fitur—fitur kata yang berguna saat proses klasifikasi.
5. Pengujian pengklasifikasian analisis sentimen dengan menggunakan 150 dokumen, dimana 100 dokumen digunakan sebagai data latih dan 50 dokumen sebagai data uji, menghasilkan akurasi tertinggi sebesar 75% dengan menggunakan nilai gamma 0,5 dan C 1.

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa penerapan metode Support Vector Machine adalah metode yang memiliki akurasi yang bagus untuk digunakan dalam kasus analisis sentimen.

5.2 Saran

Saran untuk pengembangan deteksi teks bullying di media sosial berbahasa Indonesia dari penelitian ini adalah :

1. Data set yang digunakan bisa ditambahkan untuk mengetahui apakah adanya peningkatan dalam pembelajaran machine learning. Data set dapat diganti dari media sosial lain dan bahasa typo dapat di atasi sehingga meminimalisir kesalahan pada tahap preprosesing.
2. Mengubah metode yang digunakan untuk dapat membandingkan kinerja metode dalam mendeteksi teks bullying bahasa Indonesia.

UCAPAN TERIMA KASIH

Terima kasih kepada Ibu dan Bapak, Kakak, Adik yang saya tercinta, Bapak Alif Finandhita sselaku wali kelas saya, ibu Ednawati Rainarli sebagai pembimbing saya, ibu Kania Evita Dewi sebagai reviewer, ibu Nelly Indriani sebagai ketua Jurusan Teknik Informatika,.

DAFTAR PUSTAKA

[1] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia pada Twitter dengan Metode Support Vector Machine dan Lexion Based Feature,” Jur. Pengembangan Teknologi Informatika dan Ilmu Komputer ,Universitas Brawijaya, vol. 1, no. 12, pp. 1725–1732, Agu 3, 2017.

[2] Ridwannuloh M, Iwan, “Analisis Sentimen Pada Posting Official Akun Twitter Telkom Speedy Menggunakan Naive Bayes Classifer,” 20 Juli 2014[Online].Available: https://repository.unikom.ac.id/29384/.[Accessed: 27-Juni-2019]

[`3] 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.

[4] S. Stauffer, M. A. Heath, S. M. Coyne, and S. Ferrin, “High School Teachers Perceptions of Cyberbullying Prevention and Intervention Strategies,” Psychology in the Schools,vol. 49, no. 4, pp. 352-367, 4 April 2012.

[5] R. M. Kowalski and S. P. Limber, “Electronic Bullying Among Middle School Students,” Jurnal Adolescene Healing, vol. 41, no. 6., pp. 22–30, 3 June 2007.

[6] M. S. Hajmohammadi, R. Ibrahim, and Z. Ali Othman, “Opinion Mining and Sentiment Analysis: A Survey,” International Jurnal Computer Technology,United States, vol. 2, no. 3c, pp. 171–178, 2 Juni 2015.

[7] P. van der Putten, J. N. Kok, and A. Gupta, “Why the Information Explosion Can be Bad for Data Mining, and How Data Fusion Provides a Way Out”, In Proceedings of the Second SIAM International Conference on Data Mining, Arlington, VA, USA,pp. 128–138, 2013.

[8] J. Han, M. Kamber, J. Pei, “Getting to Know Your Data. Data Mining: Concepts And Techniques: Concepts and Techniques”, pp. 39–82, 2011.

[9] A. Hamzah,”Analisis Sentimen pada Twitter dengan Metode Support Vector Machine”, “Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST),Yogyakarta, ISSN: 1979-911X,” Snast, vol. 3, pp. 211–216, 2014.

[10] H. Tuhuteru and A. Iriani, “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier,” Jurnal Informatika: Jurnal Pengembangan IT, Jakarta, vol. 3, no. 3, pp. 394–401, 2018.

[11] Hemalatha, I., GP Saradhi Varma, and A. Govardhan, "Preprocessing the Informal Text for Efficient Sentiment Analysis." International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) ,pp58-61,2012.

[12] A. Nugroho, “Analisis Sentimen pada Media Sosial Twitter Menggunakan Naive Bayes Classifier dengan Ekstrasi Fitur N-Gram,” J-SAKTI (Jurnal

Sains Komputer dan Informatika),Malang, vol. 2, no. 2, p. 200, 201
[13] I. Made Budi Surya Darma, "Penerapan Sentimen Analisis Acara Televisi pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur," Jur. Pengembangan Teknologi Informasi dan Ilmu Komputer, Jakarta,vol. 2, no. 3, pp. 998–1007, 2018.
[14] K. E. Dewi, N. Indriani, and E. Rainarli, "Evaluasi Sentence Extraction pada Peringkasan Dokumen Otomatis," 2017