

Analisis Sentimen *Cyberbullying* pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine

Wanda Athira Luqyana¹, Imam Cholissodin², Rizal Setya Perdana³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹wandathira@gmail.com, ²imamcs@ub.ac.id, ³rizalespe@ub.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 komentar. Dengan data statistik yang telah didapatkan, bahwa 42% remaja berusia 12-20 tahun telah menjadi korban *cyberbullying*. 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. Analisis sentimen merupakan suatu cabang ilmu dari text mining yang digunakan untuk mengekstrak, memahami, dan mengolah data teks. Untuk mengetahui setiap sentimen pada komentar digunakan fitur Term Frequency-Inverse Document Frequency (TF-IDF) dan metode klasifikasi Support Vector Machine (SVM). 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%.

Kata kunci: Instagram, *cyberbullying*, analisis sentimen, *support vector machine*, *svm*

Abstract

Instagram is the most popular social media in these recent days. The users who start from kids, teenagers to adults, have the role in boosting the popularity of Instagram. However, this social media could not be separated from the dangers of cyberbullying which is done often by the users, especially in the comment column. The dangers of cyberbullying are certainly worried many people because of the impact it has. Therefore, a sentiment analysis in Instagram comment column can be done in order to find out the sentiments in each comment. Sentiment analysis is a branch of text mining science which is used to extract, understand, and cultivate the data. This research used Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machine (SVM) classification method to examine the sentiments in each comment. Data consisted of 400 data which taken offline have a total 1799 features. The comment document is divided into 70% of training data and 30% of test data. Based on the tests performed, the best parameters obtained in the SVM method are the degree of polynomial kernel 2, the average of learning rate of 0.0001, and the maximum number of iterations which is 200 times. From these result, it obtained that the highest accuracy is 90%, 50% in the training data composition and 50% composition of test data.

Keywords: *instagram, cyberbullying, sentiment analysis, support vector machine, svm*

1. PENDAHULUAN

Instagram merupakan media sosial yang pada dasarnya berfungsi untuk berbagi foto maupun video pada sesama pengguna. Sejak September 2017 pengguna aktif instagram telah

tercatat kurang lebih 800 juta orang (Yusuf, 2017). Pengguna Instagram tentunya tak hanya orang dewasa namun anak-anak dan remaja ikut menggunakan media sosial ini.

Pemanfaatan yang beragam dari pengguna Instagram dibuat sebagai akun pribadi, baik

untuk orang biasa maupun artis hingga menjadi sarana bisnis perseorangan. Namun dari manfaat penggunaan Instagram masih banyak pengguna yang belum memahami etika-etika dalam bersosialisasi pada dunia maya. Permasalahan tersebut terjadi dengan adanya tindakan *cyberbullying*.

Cyberbullying yang merupakan tindakan *bullying* yang sering terjadi di dunia maya. Banyak pengguna yang masih belum menyadari bahwa ulasan atau komentar yang dilontarkan merupakan tindakan *cyberbullying*. Hal ini tentunya menjadi salah satu peringatan bagi pengguna, orangtua, kerabat, maupun pemerintah agar dapat meminimalisir perbuatan merugikan tersebut.

Cyberbullying yang telah terjadi di Indonesia dialami oleh seorang remaja asal Indonesia yang berusia 13 tahun di bulan April 2017. Hal tersebut terjadi ketika remaja itu berniat untuk membagikan tiket secara gratis kepada khalayak umum dengan beberapa syarat. Namun hal ini memberikan dampak negatif untuk dirinya. Berbagai kabar tidak baik tersebar secara cepat di dunia maya dan pada akhirnya muncul komentar negatif yang bersifat menyakiti hati anak. Sebagai dampaknya remaja yang berkepribadian riang, menjadi murung dan selalu menangis (Muttya, 2017). Permasalahan *cyberbullying* pada kolom komentar Instagram menjadi hal yang penting untuk dikaji sebagai pemrosesan teks

Analisis sentimen sangat diperlukan dalam menyaring komentar-komentar di media sosial. Analisis sentimen pada komentar dilakukan untuk mengetahui komentar yang bersifat negatif dan komentar yang bersifat positif. Dari analisis tersebut dapat dilakukan tindakan preventif baik untuk korban maupun pelaku.

Dalam melakukan analisis sentimen diperlukan metode yang menunjang klasifikasi. Metode yang digunakan adalah *Support Vector Machine* (SVM), berdasarkan hasil penelitian analisis sentimen sebelumnya yang dilakukan oleh Putranti dan Winarko. Pada penelitian tersebut dilakukan analisis sentimen pada objek Twitter dengan mengimplementasikan metode *Maximum Entropy* dan SVM dan menghasilkan nilai akurasi sebesar 86,81% (Putranti & Winarko, 2014). Penelitian analisis sentimen lainnya dilakukan dengan membandingkan hasil akurasi metode Naïve Bayes dan SVM. Berdasarkan penelitian tersebut didapatkan SVM adalah metode klasifikasi yang baik untuk analisis sentimen dengan hasil akurasi 98% (K

& Shetty, 2017).

Dari berbagai referensi penelitian yang telah dilakukan, metode *Support Vector Machine* menjadi salah satu pilihan metode yang akan digunakan peneliti dalam analisis sentimen. Hal ini menjadi topik penelitian untuk memberikan solusi terhadap *cyberbullying* yang terjadi di media sosial Instagram.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence*. Proses yang dilakukan oleh analisis sentimen untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat (Akbari, et al., 2012). Selain itu analisis sentimen merupakan bidang ilmu yang menganalisis pendapat, sikap, evaluasi, dan penilaian terhadap suatu peristiwa, topik, organisasi, maupun perseorangan (Liu, 2012).

2.2. Text Mining

Text mining adalah ilmu yang bertujuan untuk memproses teks agar menjadi informasi yang diperoleh dari peramalan pola dan kecenderungan melalui pola statistik. (Jiawei, et al., 2012). *Text mining* bertujuan untuk menganalisis pendapat, sentiment, evaluasi, sikap, penilaian, emosi seseorang sehingga dapat diketahui apakah berkenaan dengan suatu topik, layanan, organisasi, individu, atau kegiatan tertentu (Liu, 2012). Penggunaan dari *text mining* dilakukan untuk klasterisasi, klasifikasi, *information retrieval*, dan *information extraction* (Berry & Kogan, 2010).

2.2.1. Pre-processing

Pre-processing 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 (Nugroho, 2016).

2.2.1.1. Case Folding

Case folding merupakan tahapan awal pada *Pre-processing* yang bertujuan untuk mengubah setiap bentuk kata menjadi sama. Hal ini dilakukan dengan mengubah kata menjadi *lower case* atau huruf kecil.

2.2.1.2. Data Cleaning

Data cleaning merupakan proses

pembersihan kata dengan menghilangkan delimiter koma (,), titik (.), dan tanda baca lainnya. Pembersihan kata bertujuan untuk mengurangi *noise*.

2.2.1.3. Normalisasi Bahasa

Pada tahapan *Pre-processing* dilakukan normalisasi bahasa terhadap kata tidak baku. Tahapan ini bertujuan untuk mengembalikan bentuk penulisan dari masing-masing kata yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Proses ini dilakukan dengan mencocokkan setiap kata pada dokumen data latih maupun data uji dengan kata yang ada pada kamus Bahasa tidak baku (Darma, 2017).

2.2.1.4. Stopword Removal

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 (Manning, et al., 2009).

2.2.1.5. Stemming

Stemming merupakan proses untuk mencari *stem* (kata dasar) dari kata hasil *stopword removal* (*filtering*). Terdapat dua aturan dalam melakukan *stemming* yaitu dengan pendekatan kamus dan pendekatan aturan (Utomo, 2013).

2.2.1.6. 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. Cyberbullying

Berkembangnya teknologi memberikan 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 (Stauffer, et al., 2012).

Berdasarkan sumber lain mengatakan bahwa *cyberbullying* merupakan tindakan *bullying* yang dilakukan pada dunia *cyber*. Dalam tindakan *cyberbullying* dapat dibagi menjadi beberapa kriteria dan dilakukan secara berulang-ulang. Terdapat beberapa aspek yang

memenuhi pada *cyberbullying* seperti, *flaming*, *harrassment*, *cyberstalking*, dan lainnya (Pratiwi, 2017).

Pemilihan kata yang dilontarkan menjadi kunci utama apakah seseorang mengarah pada tindakan *bullying* atau tidak. Contoh kata yang tergolong pada kata positif dan kata negatif digambarkan pada Tabel 1 dan Tabel 2:

Tabel 1. Contoh kata positif

Positif	
Cekatan	Berani
Sesuai	Suci
Indah	Teliti
Wibawa	Unggul
Sabar	Pesona

Tabel 2. Contoh kata negatif

Negatif	
Abnormal	Pengecut
Aneh	Banci
Bodoh	Buruk
Gila	Jelek
Khianat	Munafik

2.4. Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk menghitung bobot setiap kata yang telah diekstrak. Penggunaan metode ini umumnya dilakukan untuk menghitung kata umum yang ada pada *information retrieval*. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model *term frequency* (*tf*) dan *inverse document frequency* (*idf*). *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 (Akbari, et al., 2012).

Tahapan pembobotan dengan TF-IDF adalah:

1. Hitung *term frequency* $tf_{i,d}$
2. Hitung *weighting term frequency* (W_{tf})

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

3. Hitung *document frequency* (*df*)

4. Hitung bobot *inverse document frequency* (*idf*)

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

5. Hitung nilai bobot TF-IDF

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

Keterangan:

$tf_{t,d}$ = frekuensi term

$W_{tf_{t,d}}$ = bobot frekuensi term

df = jumlah frekuensi dokumen yang mengandung term

N = jumlah total dokumen

$W_{t,d}$ = bobot TF-IDF

2.5. Lexicon Based Features

Lexicon Based Features merupakan suatu kesepakatan dalam pendekatan yang meliputi frase, bentuk ekspresi, atau konten yang berupa teks yang umumnya terdapat pada obrolan, dialog, *post*, *review*, dan lainnya.

Lexicon Based Features merupakan pendekatan yang menggunakan suatu kamus sentimen berisi kata positif dan kata negatif yang dibandingkan dan dicocokkan dengan kata pada kalimat untuk diketahui tingkat polaritasnya (Peng, 2011).

2.5.1. Normalisasi Min-Max

Berdasarkan penelitian yang dilakukan oleh Rofiqoh dilakukan pembobotan *lexicon* menggunakan metode normalisasi *min-max* dengan nilai maksimum 0,9 dan nilai minimum 0,1 (Rofiqoh, 2017). Hal ini diperuntukan untuk menormalisasi data sehingga data tersebut berada pada *range* tertentu (Junaedi, et al., 2011). Tujuan dalam menormalisasi data ialah untuk meminimalisir kesalahan pada proses *data mining* (Wirawan & Eksistyanto, 2015). Rumus matematika yang digunakan dalam metode ini adalah:

$$v'_i = \frac{v_i - \min_a}{\max_a - \min_a} (newmax - newmin) + newmin \quad (4)$$

Keterangan:

v'_i = hasil normalisasi data ke-*i*

v_i = data yang dinormalisasi data ke-*i*

\min_a = data minimum pada kumpulan data *a*

\max_a = data maksimum pada kumpulan

data *a*

$newmax$ = nilai normalisasi maksimum

$newmin$ = nilai normalisasi minimum

2.5.2. Skor Sentimen

Pembobotan *Lexicon Based* lainnya dilakukan oleh Peng dengan mempertimbangkan skor sentimen dari setiap komentar. Tahapan yang perlu dilalui pada pembobotan *Lexicon Based* menggunakan perhitungan skor sentimen adalah sebagai berikut (Peng, 2011):

1. Memuat kamus, termasuk kata-kata dan *POS tags*
2. *Parse* komentar ke dalam *POS tags*. Hanya kata yang memiliki *tag* yang benar yang akan dihitung.
3. Hitung skor sentimen pada setiap komentar. Skor sentimen dihitung dengan mencari jumlah kata bersentimen positif dan kata bersentimen negatif. Skor sentimen didapatkan dari jumlah polaritas sentimen positif dikurangi dengan jumlah polaritas sentimen negatif.
4. Kalimat yang merupakan kalimat negatif diperhitungkan dengan menambahkan tanda minus ke skor sentimen.

2.6. Algoritme Support Vector Machine

Support Vector Machine atau SVM merupakan salah satu teknik untuk memprediksi yang baik dalam pengklasifikasian dan regresi (Santosa, 2007). Penggunaan algoritme SVM yang bertujuan untuk klasifikasi teks dengan menggunakan bobot indeks *term* sebagai fitur, dirintis oleh Thorsten Joachim. Pembelajaran SVM telah dipopulerkan sejak tahun 1992 oleh Boser, Guyon, dan Vapnik (Paramita, 2008).

SVM merupakan metode yang dapat menyelesaikan permasalahan secara linier maupun permasalahan *non-linier*. Dalam menyelesaikan permasalahan *non-linier* digunakan konsep kernel pada ruang kerja berdimensi tinggi, dengan mencari *hyperplane* yang dapat memaksimalkan margin antar kelas data. *Hyperplane* berguna dalam memisahkan 2 kelompok *class +1* dan *class -1* dimana setiap *class* memiliki *pattern* masing-masing.

Dalam mengambil keputusan dengan metode SVM digunakan fungsi kernel $K(x_i, x_d)$. Kernel yang digunakan pada penelitian ditunjukkan pada Persamaan 5:

$$K(x_i, x_d) = (X_i^T X_j + C)^d, \gamma > 0 \quad (5)$$

Pemrosesan yang dilakukan pada data latih digunakan algoritme *sequential training* karena

merupakan algoritme yang sederhana tanpa memakan waktu yang banyak (Vijayakumar, 1999) dengan tahapan perhitungan:

1. Inisialisasi terhadap berbagai parameter, seperti α_i , γ , C , dan ε .

α_i = alfa, untuk mencari *support vector*

γ = konstanta gamma untuk mengontrol kecepatan

C = variabel *slack*

ε = epsilon digunakan untuk mencari nilai error

2. Hitung matriks *Hessian* yang didapat dari perkalian antar kernel polynomial dan y yang merupakan vector bernilai 1 dan -1. Persamaan dari matriks *Hessian* adalah:

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (6)$$

3. Lakukan perhitungan berikut hingga interaksi data i hingga j :

- a. $E_i = \sum_j^i \alpha_j D_{ij} \quad (7)$

- b. $\delta \alpha_i = \min(\max[\gamma(1 - E_i), \alpha_i], C - \alpha_i)$

(8)

- c. $\alpha_i = \alpha_i + \delta \alpha_i \quad (9)$

4. Lakukan ketiga langkah diatas secara berulang hingga mencapai batas maksimum iterasi
5. Proses *sequential learning* dari tahap 1 hingga 4 akan mendapatkan nilai dari support vector (SV), dimana nilai $SV = (\alpha_i > threshold_{SV})$. Setelah itu, perlu dilakukan perhitungan pada nilai *bias* b yang diperoleh dari Persamaan 10.

$$b = -\frac{1}{2} (\sum_{i=0}^N \alpha_i y_i K(x_i, x^-) + \sum_{i=0}^N \alpha_i y_i K(x_i, x^+)) \quad (10)$$

6. Untuk mengetahui hasil klasifikasi teks pada kelas sentimen tertentu maka dilakukan proses perhitungan fungsi $f(x)$. Jika hasil dari fungsi tersebut bernilai negatif, maka dokumen terklasifikasi pada sentimen kelas negatif *cyberbullying*. Jika nilai fungsi bernilai positif, maka dokumen terklasifikasi pada kelas sentimen positif *cyberbullying*. Fungsi $f(x)$ diperoleh pada Persamaan 11.

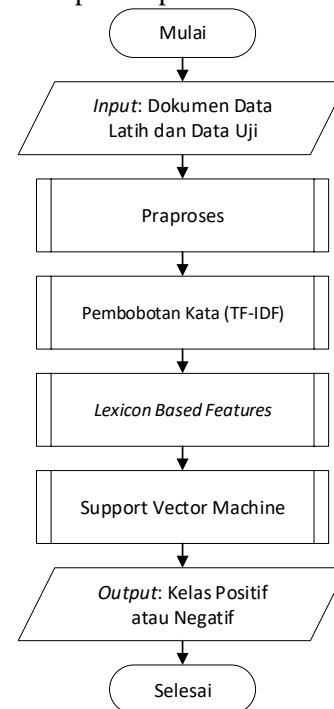
$$f(x) = \sum_{i=0}^m \alpha_i y_i K(x_i, x) + b \quad (11)$$

3. METODE USULAN

3.1 Alur Proses Sistem

Tahapan dalam melakukan analisis sentimen dengan metode klasifikasi SVM dimulai dengan *input* data yang berupa data latih dan data uji yang kemudian diproses pada tahapan *pre-processing* hingga proses klasifikasi

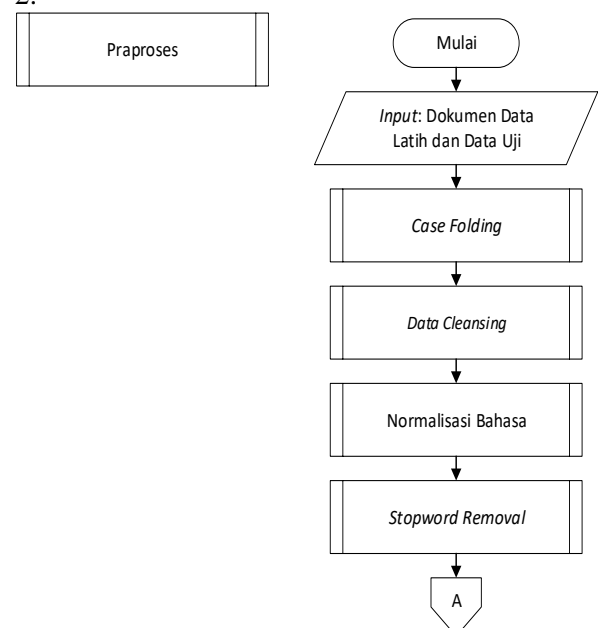
untuk menentukan kelas prediksi. Secara umum sistem dideskripsikan pada Gambar 1.

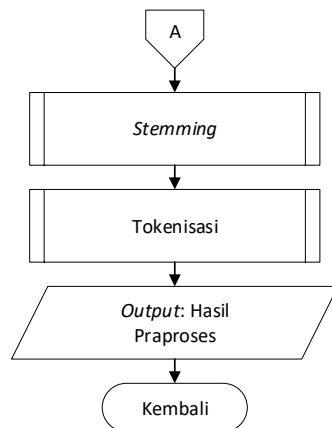


Gambar 1. Deskripsi Umum Sistem

3.2. Pre-processing

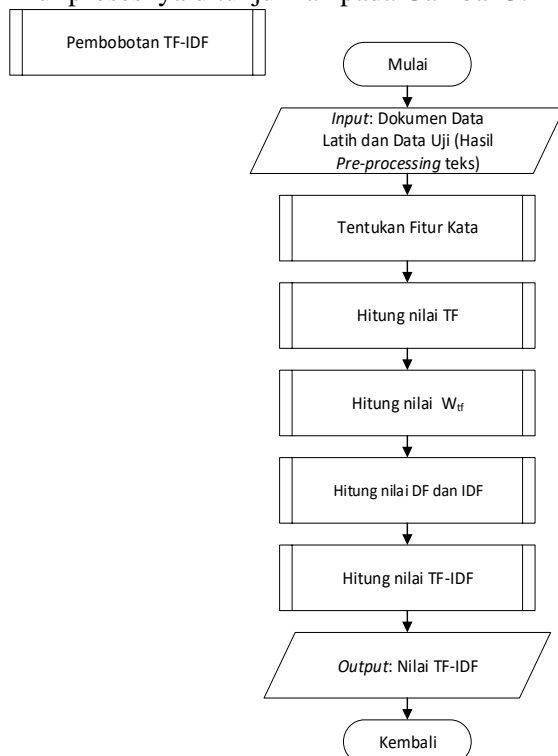
Pre-processing merupakan tahapan awal yang akan dilalui dalam memproses teks. Pada penelitian ini akan dilakukan tahapan *Pre-processing* dengan tahapan *case folding*, *data cleaning*, normalisasi kata tidak baku, *stopword removal*, *stemming*, dan tokenisasi. Alur proses pada tahapan ini akan ditunjukkan pada Gambar 2.



Gambar 2. Alur proses *Pre-processing*

3.3. Pembobotan TF-IDF

Data yang telah melalui tahapan *Pre-processing* 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. Alur prosesnya ditunjukkan pada Gambar 3.

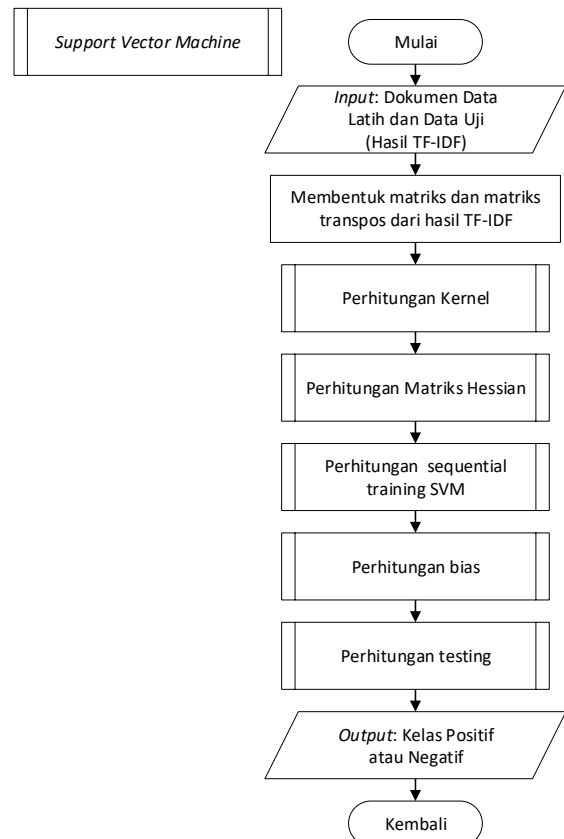


Gambar 3. Alur proses Pembobotan TF-IDF

3.4. Support Vector Machine

Metode *Support Vector Machine* merupakan metode yang digunakan untuk analisis sentimen pada penelitian ini. Hasil yang akan ditentukan dengan metode ini adalah klasifikasi kelas positif dan kelas negatif yang

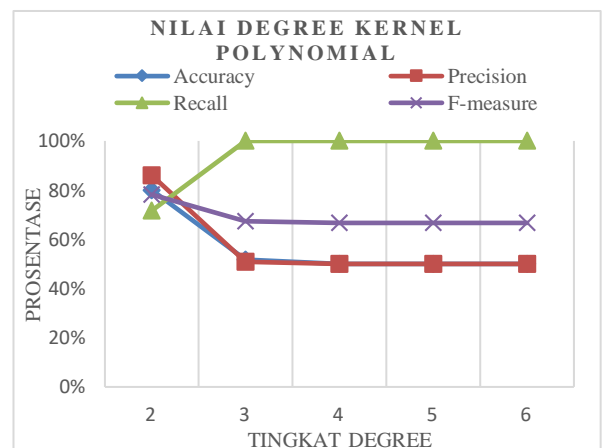
didapat berdasarkan bobot pada setiap fitur dokumen teks. Alur proses metode *Support Vector Machine* ditunjukkan pada Gambar 4.



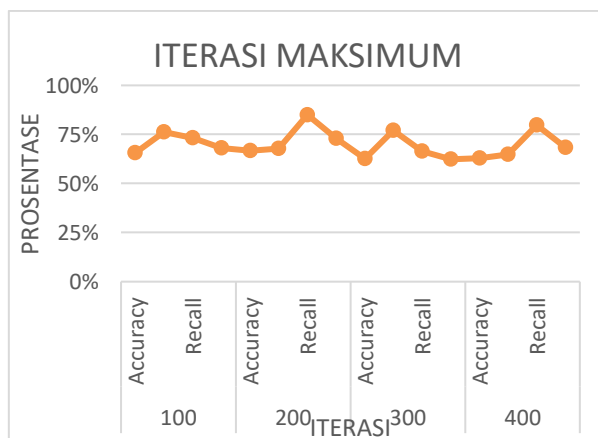
Gambar 4. Alur proses Support Vector Machine

4. PENGUJIAN DAN ANALISIS

Pengujian yang telah dilakukan pada beberapa parameter *Support Vector Machine*. Parameter SVM yang diuji adalah nilai *degree* kernel *polynomial*, iterasi maksimum, dan konstanta *learning rate*.

Gambar 5. Grafik hasil pengujian pengaruh nilai *degree*

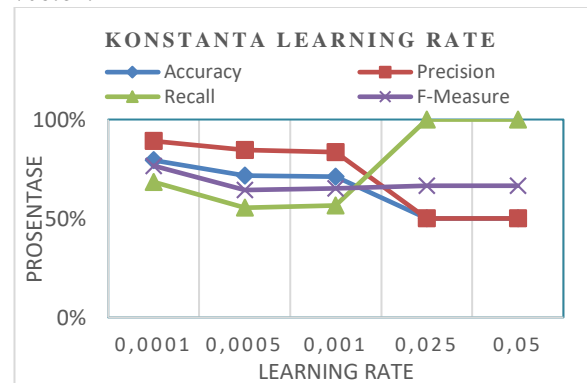
Pada Gambar 5 ditunjukkan nilai akurasi terbaik terletak pada nilai *degree* = 2, yaitu 80% dan diikuti oleh nilai *precision*, *recall*, dan *f-measure*. Nilai akurasi pada *degree* 3 hingga 6 cenderung konstan, yang mana nilai akurasinya menunjukkan pada prosentase 50%. Nilai *recall* yang didapatkan pada *degree* 3 hingga 6 dapat mencapai 100% menunjukkan bahwa sistem telah berjalan efektif, dikarenakan hasil yang diberikan sistem terhadap data yang relevan lebih besar atau seimbang. Dengan begitu dapat disimpulkan bahwa nilai *degree* yang paling optimal ketika pada *degree* 2 dengan memperoleh tingkat akurasi tertinggi sebesar 80%. Dengan meningkatnya nilai *degree* pada kernel polynomial berpengaruh terhadap hasil perhitungan matriks Hessian yang berfungsi dalam mencari nilai optimum pada setiap dokumen data, dimana hasil dari matriks hessian digunakan untuk menghitung besar nilai *error rate* pada setiap dokumen dan berpengaruh terhadap pembentukan nilai *support vector*.



Gambar 6. Grafik hasil pengujian iterasi maksimum

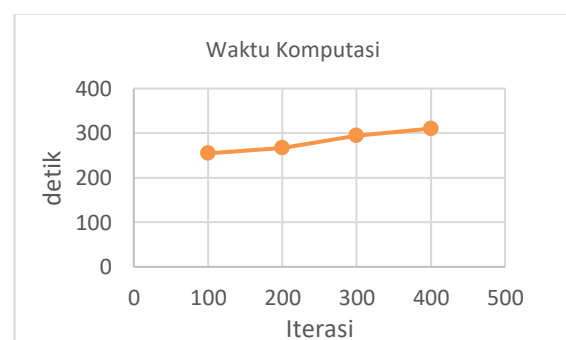
Dari pengujian iterasi maksimum didapatkan bahwa nilai akurasi paling baik ketika iterasi maksimum sebanyak 200 kali. Hal ini ditunjukkan oleh Gambar 6, dimana iterasi maksimum 200 memiliki tingkat akurasi sebesar 66,67% dan akurasi mengalami penurunan ketika iterasi maksimumnya adalah 300 kali dengan tingkat akurasi 62,50%. Namun peningkatan jumlah iterasi secara terus-menerus tidak menandakan perbaikan akurasi. Penurunan tingkat akurasi terjadi karena pada tahapan *sequential learning* akan mengalami perubahan nilai α_i . Perubahan nilai α_i yang berpengaruh pada penurunan tingkat akurasi ini karena nilai α_i yang menjadi tidak konvergen yang dapat dibuktikan dengan perubahan nilai α_i . Nilai dari

α_i tersebut menjadi pembentukan *support vector*.



Gambar 7. Grafik hasil pengujian iterasi maksimum

Hasil pengujian yang ditunjukkan oleh Gambar 7 menunjukkan nilai konstanta *learning rate* terbaik ketika 0,0001. Hal ini ditunjukkan dengan tingkat akurasi yang mencapai 79,38%. Konstanta *learning rate* berfungsi untuk mengontrol kecepatan pada proses *training* dan bergantung pada jumlah iterasi untuk mencapai konvergensi. Dari hasil pengujian ditunjukkan bahwa nilai optimal konstanta *learning rate* adalah 0,0001 dan terjadi ketika iterasi maksimum mencapai 200 kali. Perubahan nilai *learning rate* pada sistem memengaruhi nilai $\delta\alpha$. Karena *learning rate* menjadi salah satu kandidat nilai yang berpengaruh dalam pembentukan nilai alfa dan pembentukan himpunan *support vector*. Semakin meningkatnya nilai *learning rate* berdampak pada proses pelatihan yang semakin cepat, sehingga tingkat ketelitian menjadi berkurang.



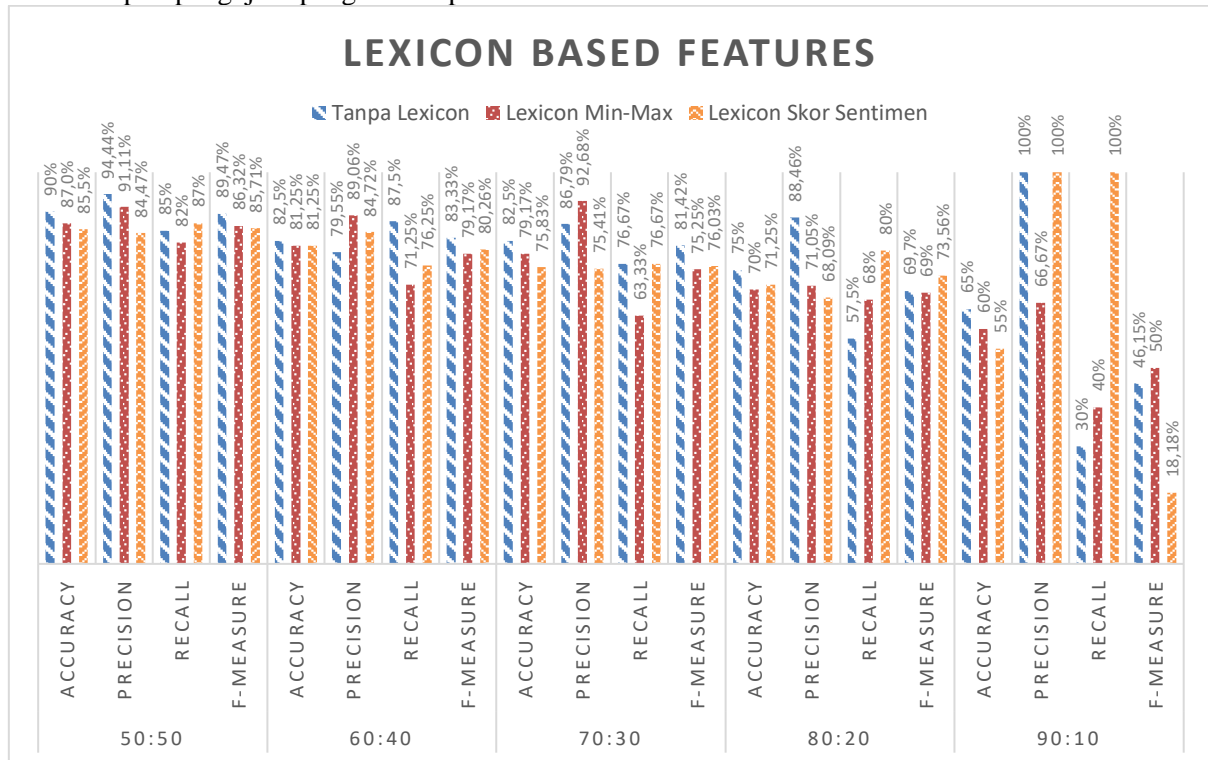
Gambar 8. Grafik hasil pengujian iterasi maksimum terhadap waktu komputasi sistem

Pengujian waktu komputasi (*running time*) pada sistem dilakukan terhadap iterasi maksimum yang dilakukan. Seiring dengan bertambahnya iterasi maksimum, maka akan meningkatkan waktu komputasi yang

direpresentasikan dalam satuan waktu detik. Ditunjukkan pada Gambar 8, bahwa waktu komputasi terbaik sistem dalam menyelesaikan klasifikasi adalah dengan menggunakan 100 kali iterasi maksimum. Waktu yang dibutuhkan yaitu selama 254.6124172 detik.

Lexicon Based Features bertujuan untuk mencari pengaruh implementasi *Lexicon Based Features* terhadap tingkat akurasi sistem. Tahapan pengujian ini akan membandingkan sistem ketika *Lexicon Based Features* diimplementasikan dan ketika *Lexicon Based Features* tidak diimplementasikan.

Tahapan pengujian pengaruh implementasi



Gambar 9. Grafik hasil pengujian pengaruh penerapan tanpa *lexicon based features*

Dari hasil pengujian diketahui bahwa proses klasifikasi yang dilakukan tanpa mengimplementasikan metode *Lexicon Based Features* memiliki tingkat akurasi yang lebih baik dibandingkan dengan proses klasifikasi yang mengimplementasikan metode *Lexicon Based Features*. Pengaruh implementasi dari metode *Lexicon Based Features* yaitu pada penggunaan kamus lexicon yang masih umum (tidak terfokus pada *cyberbullying*), namun data yang digunakan merupakan data yang bersifat variatif dan kompleks. Kamus lexicon yang digunakan hanyalah berbentuk kata, namun dalam mengenali sifat sentimen dari kalimat *cyberbullying* dibutuhkan kamus yang lebih spesifik yaitu kamus yang terdapat bentuk frase. Prosentase akurasi terbaik mencapai 90% pada sistem yang tidak mengimplementasikan *Lexicon Based Features*.

Didapatkan bahwa rata-rata akurasi *Lexicon Based Features* dengan normalisasi *min-max* lebih baik dibandingkan dengan *Lexicon Based*

Features dengan perhitungan skor sentimen. Pada satu komposisi data, akurasi yang didapatkan pada kedua cara implementasi *Lexicon Based Features* tidak terdapat perbedaan. Namun ketika pengujian yang dilakukan pada komposisi data latih dan data uji lainnya memberikan tingkat akurasi yang cukup berbeda. Penyebab perbedaan yang terjadi dikarenakan hasil yang didapatkan ketika menghitung matriks Hessian. Pengaruh perbedaan hasil yang signifikan dari perhitungan matriks Hessian dapat memberikan dampak dalam pembentukan *support vector*.

Selain pengujian implementasi *Lexicon Based Features* dan tanpa *Lexicon Based Features*, dapat dilihat bahwa terdapat pengujian pada komposisi data latih dan data uji. Tingkat akurasi terbaik sebesar 90% yang didapatkan pada 50% komposisi data latih dan 50% komposisi data uji pada sistem yang tidak mengimplementasikan algoritme *Lexicon Based Features*. Namun seiring dengan bertambahnya

komposisi data latih, tingkat akurasi semakin menurun. Hal ini terjadi karena adanya *over-fitting*, yaitu pada proses *training* (pelatihan) data telah dimodelkan dengan sangat baik, sehingga memungkinkan *noise* data telah dipelajari dan menyebabkan data uji tidak dapat diprediksi dengan baik.

Selain itu terdapat perubahan nilai *precision* dan *recall* pada pengujian komposisi data latih dan data uji. Perubahan nilai *precision* disebabkan oleh perubahan jumlah prediksi data yang diklasifikasikan sebagai kelas positif. Meningkatnya nilai *precision* dikarenakan rendahnya hasil prediksi kelas data yang tidak disesuaikan dengan keinginan (*query*). Namun meningkatnya nilai *recall* berkebalikan dengan nilai *precision*. Hal ini dikarenakan dengan seiring peningkatan nilai *recall*, kelas data yang ditemukan telah relevan dengan keinginan (*query*). Peningkatan nilai *recall* menunjukkan bahwa sistem telah berjalan dengan efektif.

5. KESIMPULAN DAN SARAN

Dari hasil pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan dalam analisis sentimen *cyberbullying* pada komentar Instagram yang menerapkan metode SVM. Hal yang disimpulkan dari penelitian ini adalah sebagai berikut:

1. Support Vector Machine dapat digunakan dalam menganalisis sentimen *cyberbullying* pada kolom komentar Instagram. Hasil klasifikasi berupa kelas positif dan negatif yang dibedakan menjadi sentimen positif *cyberbullying* dan sentimen negatif *cyberbullying*.
2. Berdasarkan hasil pengujian yang dilakukan didapatkan tingkat akurasi terbaik sebesar 90%, *precision* sebesar 94,44%, 85% *recall* sebesar dan *f-measure* sebesar 89,47% dengan komposisi data latih 50% dan data uji 50% dan tanpa mengimplementasikan algoritme *Lexicon Based Features*.

Berdasarkan penelitian yang telah dilakukan, masih terdapat beberapa kekurangan yang perlu diperbaiki maupun dikembangkan dari penelitian ini. Saran yang diberikan untuk dilakukan pada penelitian berikutnya adalah:

1. Data yang digunakan untuk proses klasifikasi didapatkan secara *real time* yang kemudian dapat dimasukkan sebagai

tambahan data untuk data latih baik yang bersentimen positif *cyberbullying* maupun yang bersentimen negatif *cyberbullying*.

2. Data diklasifikasikan menjadi tiga kelas sentimen, yaitu sentimen positif *cyberbullying*, netral, dan negatif *cyberbullying*. Hasil klasifikasi dari ketiga kelas tersebut dapat diketahui polaritas dari setiap sentimennya.
3. Dapat diimplementasikan suatu metode optimasi dalam ekstraksi fitur yang berguna dalam mengidentifikasi teks berdasarkan makna perkata, frase, dan kalimat.
4. Tahapan evaluasi sistem dapat mempertimbangkan konsep *macro average* dan *micro average* untuk mengevaluasi kinerja sistem, sehingga dapat meningkatkan kinerja dalam mengklasifikasi multi-label.
5. Sistem dapat dikembangkan menjadi suatu produk yang bersifat preventif bagi pengguna baik yang membaca kolom komentar maupun yang mengunggah suatu komentar.

6. DAFTAR PUSTAKA

- Akbari, M. I. H. A. D., Astri Novianty S.T., M. & Casi Setianingsih S.T., M., 2012. Analisis Sentimen Menggunakan Metode Learning Vector Quantization. Telkom University .
- Berry, M. & Kogan, J., 2010. Text Mining Application and Theory. In: Wiley: United Kingdom .
- Darma, I. M. B. S., 2017. Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur.
- Jiawei, H., Kamber, M. & Pei, J., 2012. Data Mining: Concepts and Techniques Third Edition. MA: Morgan Kaufmann.
- Junaedi, H., Budianto, H., Maryati, J. & Melani, Y., 2011. Data Transformation Pada *Data mining*. *Prosiding Konferensi Nasional "Inovasi dalam Desain dan Teknologi" - IDeaTech*.
- K, S. T. & Shetty, J., 2017. Sentiment Analysis of Product Reviews: A Review.

- International Conference on Inventive Communication and Computational Technologies.
- Liu, B., 2012. Sentiment Analysis and Opinion Mining. In: Chicago: Morgan & Claypool Publisher.
- Manning, C., Raghavan, P. & Schütze, H., 2009. An Introduction to Information Retrieval. Cambridge: Cambridge University Press.
- Muttya, A., 2017. Kompas.com. [Online] Tersedia di: <<http://entertainment.kompas.com/read/2017/04/28/060000410/gara-gara.tiket.bts.putri.uya.kuya.dapat.ancaman>> [Diakses 23 Agustus 2017].
- Nugroho, G. A. P., 2016. Analisis Sentimen Data Twitter Menggunakan K-Means Clustering.
- Paramita, 2008. Penerapan Support Vector Machine untuk Ekstraksi Informasi dari Dokumen Teks. Laporan Tugas Akhir, Program Studi Teknik Informatika, STEI Institut Teknologi Bandung.
- Peng, W., 2011. Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization. s.l.:s.n.
- Pratiwi, A., 2017. *Cyberbullying* [Interview] (18 Oktober 2017).
- Putranti, N. D. & Winarko, E., 2014. Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. IJCCS, Volume 8, pp. 91-100.
- Santosa, B., 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta : Graha Ilmu.
- Stauffer, S., Heath, M. A., Coyne, S. M. & Ferrin, S., 2012. High School Teachers Perceptions of *Cyberbullying* Prevention and Intervention Strategies. Psychology in the Schools, Volume 49.
- Utomo, M. S., 2013. Implementasi Stemmer Tala pada Aplikasi Berbasis Web. Jurnal Teknologi Informasi DINAMIK, Volume 18, pp. 41-45.
- Vijayakumar, W. S., 1999. Sequential Support Vector Classifiers and Regression. International Conference on Soft Computing, Issue SOCO'99, pp. 610-619.
- Yusuf, O., 2017. Naik 100 Juta, Berapa Jumlah Pengguna Instagram Sekarang?. [Online] Tersedia di: <<http://tekno.kompas.com/read/2017/09/29/06304447/naik-100-juta-berapa-jumlah-pengguna-instagram-sekarang>> [Diakses 11 Februari 2018].