

Identifikasi *Cyberbullying* pada Kolom Komentar Instagram dengan Metode *Support Vector Machine* dan *Semantic Similarity*

(*Cyberbullying* Identification on Instagram Comment Using *Support Vector Machine* and *Semantic Similarity*)

Lintani Afina Hajar Raudhoti*, Anisa Herdiani, dan Ade Romadhony

Jurusan Teknik Informatika, Fakultas Informatika, Universitas Telkom

Jl. Telekomunikasi no.1, Terusan Buah Batu, Bandung, Jawa Barat-INDONESIA

Email: lintaniafina@student.telkomuniversity.ac.id, aderomadhony@telkomuniversity.ac.id, anisaherdiani@telkomuniversity.ac.id

**Penulis korespondensi*

Abstract—Instagram is a popular social media platform where users can share photos and videos, and also post comments on other users' postings. Although there are many benefits on sharing information and posting comments, the freedom of using a social media platform also has a negative effect. One of non-constructive actions performed by social media users is cyberbullying, a misuse of technology through social media to embarrass or threaten others. Cyberbullying could affect the social media users, especially the target/victim, hence we have to build a system that can limit the negative posts. In this research, we tackled the cyberbullying detection on Instagram comments as a classification problem and employed the SVM classifier. As a supervised machine learning approach, the SVM method has limitation on processing new words that is unseen in the training data. Therefore, we employed a semantic information derived from pre-trained word embeddings to gather similar words that appear in the training data to substitute the unknown words in the testing data. The experimental results showed that the use of semantic similarity information improved the classification accuracy by 7%, from 67% to 74%.

Key Words: cyberbullying, support vector machine, semantic similarity.

I. PENDAHULUAN

Popularitas media sosial yang semakin meningkat tidak lepas dari fenomena *cyberbullying*. Facebook, twitter, ask.fm, dan instagram terdaftar dalam lima platform media sosial yang memiliki persentase tertinggi pada peristiwa *cyberbullying*, dengan Instagram menduduki peringkat pertama[1]. *Cyberbullying* adalah suatu tindakan penyalahgunaan teknologi melalui media sosial untuk mempermalukan atau mengancam orang lain[2].

Pengguna instagram dapat menulis komentar dalam kolom komentar pada unggahan foto atau video seseorang. Komentar yang ditulis berupa kalimat dijadikan sebagai input dengan output berupa identifikasi kalimat yang mengandung *cyberbullying* dan non-*cyberbullying*.

Berdasar beberapa penelitian yang menyelesaikan persoalan identifikasi *cyberbullying* dengan pendekatan pembelajaran mesin, metode yang mempunyai kinerja paling baik adalah SVM, dibanding metode *Naïve Bayes* dan

KNN[3]. Pada penelitian [3] terdapat persoalan terkait akurasi pada klasifikasi karena jumlah data yang digunakan sedikit yaitu 685 komentar. Peningkatan akurasi dapat dilakukan antara lain dengan menambahkan informasi semantik seperti yang dilakukan pada penelitian klasifikasi pada topik ulasan produk dengan kenaikan akurasi dari 88.2% menjadi 89.9%[4].

Berdasar hasil penelitian [3][4] identifikasi *cyberbullying* pada penelitian ini menggunakan metode SVM sebagai metode klasifikasi pada data latih dengan menambahkan informasi semantik pada data uji. Informasi semantik yang akan digunakan berasal dari basis pengetahuan korpus, yaitu berupa nilai kemiripan antar kata, sesuai dengan yang digunakan pada penelitian[4]. Penambahan kata-kata yang mempunyai nilai kemiripan tinggi dengan kata-kata pada data pengujian dapat membantu meningkatkan akurasi identifikasi, ketika terdapat kondisi kata pada data pengujian yang tidak tercakup dalam data pelatihan[4]. Penelitian [4] menggunakan basis pengetahuan WordNet dalam topik ulasan produk. Namun, penerapan *word2vec* lebih dipilih dalam penelitian ini karena komentar instagram mengandung bahasa terkini yang tidak terdapat dalam WordNet. Selain itu, *word2vec* bekerja mirip dengan pikiran manusia menggunakan asosiasi kata untuk membantu komputer mencari kemungkinan kombinasi kata dari basis pengetahuan korpus[5].

II. TINJAUAN PUSTAKA

A. *Cyberbullying*

Instagram merupakan salah satu platform media sosial yang menawarkan fitur berbagi foto dan video dengan sebuah kolom komentar pada setiap postingan yang dibuat. Selain itu, aktivitas yang dapat dilakukan oleh pengguna instagram seperti meninggalkan komentar pada kolom komentar dan berinteraksi dengan pengguna lain. Platform ini menduduki peringkat pertama pada fenomena *cyberbullying*. Terdapat sepuluh kategori *cyberbullying*, yaitu: (1) ancaman kekerasan fisik melalui teks. (2) penghinaan terkait kebencian. (3) pemanggilan nama (termasuk homophobia). (4) ancaman kematian. (5) berakhirnya hubungan platonis. (6) pelecehan seksual melalui teks. (7) penuntutan/instruksi. (8) mengancam untuk merusak

suatu hubungan. (9) ancaman yang ditujukan untuk keluarga. (10) ancaman melalui pesan[6]. Setiap komentar yang memenuhi salah satu dari sepuluh kategori tersebut diidentifikasi sebagai komentar yang mengandung *cyber-bullying*[6].

B. Support Vector Machine

Support Vector Machine merupakan salah satu metode klasifikasi data linear dan nonlinear. Algoritma ini menggunakan pemetaan nonlinear untuk mengubah data latihan ke dimensi yang lebih tinggi. Pada dimensi baru dicari *hyperplane* pemisah yang memisahkan tupel dari satu kelas dengan kelas yang lain. *Hyperplane* terbentuk dari *support vector* (vektor terpenting dari tupel latihan) dan *margin* (didefinisikan oleh *support vector*). Pembelajaran yang dilakukan pada SVM adalah menemukan maksimum margin (*Maximum Marginal Hyperplane*). Semakin besar margin semakin akurat saat melakukan klasifikasi menggunakan data baru[7].

Proses klasifikasi pada metode SVM dapat dilakukan dengan menggunakan pengaturan jenis kernel seperti *linear*, *polynomial*, *rbf*, dan *sigmoid*. Jenis kernel yang populer digunakan adalah *polynomial*. Pada *kernel polynomial* terdapat parameter *degree* yang dapat diubah untuk mencari nilai optimal pada proses klasifikasi yang ditunjukkan pada Persamaan (1)[8]:

$$k(x_i, x_h) = (x_i^T x_h + 1)^d \quad (1)$$

C. TFIDF

Pembobotan TFIDF merupakan perhitungan statistik untuk mengambil fitur kata yang penting pada suatu dokumen. TF-IDF menggabungkan model *term frequency* (tf) dan *inverse document frequency* (idf). *Term frequency* (tf) mengacu jumlah kemunculan term dalam satu dokumen sedangkan pada *inverse document frequency* (idf) menurunkan bobot kata yang sering muncul dengan tujuan untuk mengurangi efek kata yang tidak penting/unik. Nilai idf dapat dihitung menggunakan Persamaan (2) [9]:

$$idf_t = \log \left(\frac{N}{df_t} \right) \quad (2)$$

Keterangan :

N : jumlah dokumen

DF : jumlah dokumen yang terdapat term (t)

Bobot TFIDF pada term t dalam dokumen d ($W_{t,d}$) adalah perkalian antara bobot tf dan bobot idf menggunakan Persamaan (3) [9]:

$$W_{t,d} = \log(1 + tf_{t,d}) * \log \left(\frac{N}{df_t} \right) \quad (3)$$

Keterangan :

d : dokumen ke-d

t : kata (term) ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : frekuensi kemunculan kata (term) pada dokumen d

IDF : nilai IDF dari term t

D. Semantic Similarity

Semantic Similarity merupakan interaksi dua elemen yang mempertimbangkan hubungan taksonomi[10]. Kemiripan semantik tersebut tidak selalu berupa kemiripan makna, melainkan dapat berupa antonim, *hyponim*, *hyponim*, *holonym*, dan *meronym* pada dua konsep[11]. Sebuah contoh yang dapat menggambarkan kemiripan semantik salah satu nya kata “happy” dan “glad” memiliki kemiripan semantik karena dua kata tersebut tersusun dalam konsep taksonomi yang sama[4].

Semantic similarity juga dapat digunakan untuk mengkarakterisasi sebuah kata dari keragaman kosa kata dalam korpus. Dua kata dapat dibandingkan berdasarkan letaknya dalam hal ini ruang multi-dimensi. Kata-kata diwakili oleh vektor dan dibandingkan kedekatannya antar vektor kata lainnya. Perhitungan kedekatan antar vektor dapat menggunakan rumus *cosine-similarity* ditunjukkan pada Persamaan 4 [10].

$$sim_{cos}(u, v) = \frac{\sum_n^{k=1} u_k v_k}{\sqrt{\sum_n^{k=1} u_k^2} \sqrt{\sum_n^{k=1} v_k^2}} \quad (4)$$

E. Word2vec

Word2vec adalah model pemetaan kata menjadi vektor yang dirumuskan pertama kali oleh Mikolov[12]. *Word2vec* memiliki dua algoritma yaitu *continuous bag-of-words* (CBOW) dan *skipgram*. CBOW memiliki prinsip memprediksi kata output dari vektor kata di sekitarnya, dengan probabilitas seperti pada Persamaan 5.

$$P(W_t | \tau(W_{t-k}, W_{t-k+1}, \dots, W_{t+k-1}, W_{t+k})) \quad (5)$$

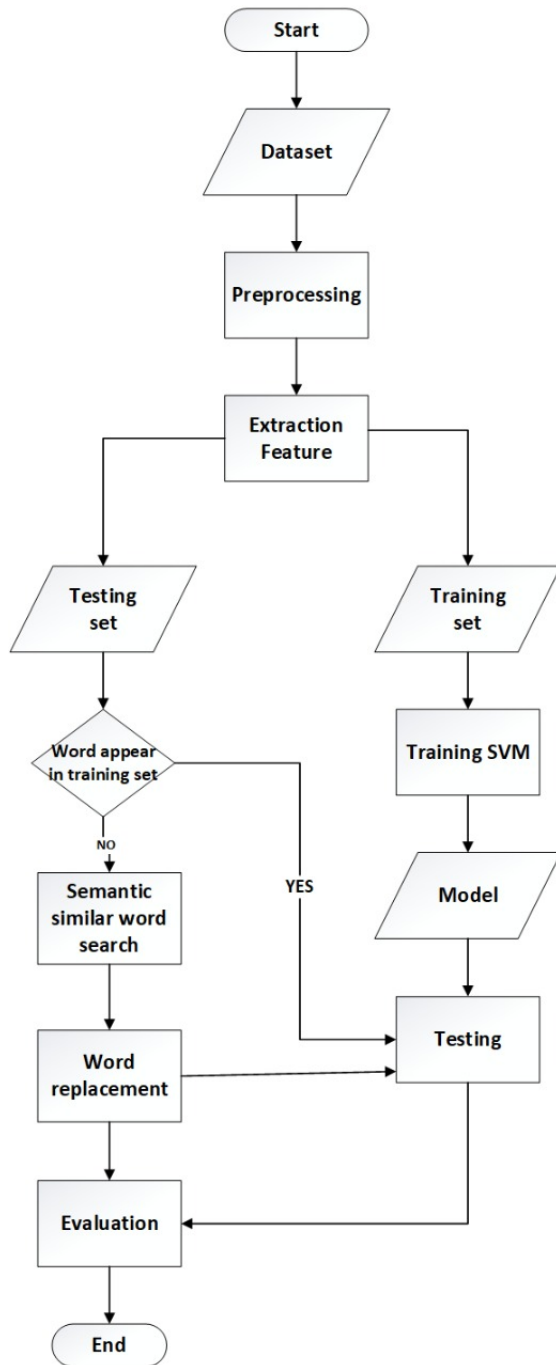
Persamaan (5) digunakan untuk menghitung probabilitas kemunculan kata (W_t). (W_t) adalah kata yang terdapat dalam kamus, k merupakan ukuran jendela konteks ketetanggaan. Sedangkan prinsip kerja *skipgram* memprediksi kata-kata disekitarnya dengan sebuah kata input, sehingga perumusan skip gram digunakan untuk menghitung probabilitas dari kosakata dalam jendela kata k dengan probabilita seperti pada Persamaan 6.

$$P(W_{t-k} | \tau(W_{t-k+1}, W_{t+k-1}, \dots, W_{t+k} | W_t)) \quad (6)$$

III. METODE PENELITIAN

Bagian sistem ditunjukkan pada Gambar 1. Tahap *preprocessing* terdiri dari: *case folding*, *data cleaning*, tokenisasi, *stopword removal* *postagging*, *lematisasi*. Setelah data dibersihkan tahap selanjutnya pembobotan kata dengan menggunakan TFIDF. Data hasil pada tahap TFIDF berupa vektor yang akan melalui tahap *training* ataupun *testing*. Tahap *training* menggunakan algoritma SVM untuk membentuk model sistem. Model tersebut akan digunakan pada tahap *testing* yang dilakukan proses penerapan model pada data latihan dengan penambahan pekerjaan dalam mencari kemiripan kata pada data uji dengan pendekatan perhitungan semantik. Pencarian kemiripan kata menggunakan basis pengetahuan korpus dengan membangun kamus *word2vec* yang memberikan

representasi data berupa vektor sehingga dapat menghitung kedekatan antar kata, tahapan yang dilakukan ditunjukkan pada Gambar 3. Kamus *word2vec* hanya dapat dibangun menggunakan data yang besar, maka pada penelitian ini menggunakan 62.088 komentar instagram tanpa label untuk membangun kamus *word2vec*. Tahap selanjutnya adalah klasifikasi menggunakan metode SVM dengan hasil kelas *cyberbullying* dan *non-cyberbullying*.



Gambar 1. Gambaran Umum Sistem

A. Dataset

Penelitian ini menggunakan data berupa kalimat yang diambil pada kumpulan komentar pada instagram. Jumlah dataset yang digunakan adalah 595 komentar sebagai data latih dan 90 komentar sebagai data uji. Dataset yang digunakan diambil dari penelitian dengan topik yang sama pada penelitian "Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network"[13]. Penelitian tersebut menggunakan data gambar serta komentar dengan pelabelan secara *multilabel* yaitu *cyberaggression* dan *cyberbullying*. Kontribusi penelitian ini adalah melakukan pelabelan ulang yang dilakukan oleh tiga responden dengan kualifikasi: memiliki kemampuan berbahasa Inggris, mengetahui *cyberbullying*, dan sebagai pengguna instagram. Tabel I merupakan contoh dataset yang digunakan:

Tabel I. DATA LATIH SEBAGAI DATA INPUT PELATIHAN SEBELUM PREPROCESSING

No.	Kalimat Komentar
1.	kylerrhollyfieldSuch an inspiration (created_at: 2013-04-29 21:37:09)
2.	lougutz22But keep doing what ur doing u looking great mama(created_at: 2013-12-26 03:54:24)
3.	mademan2311Lol someone really say they'd drink your bath water?>@paigehathaway lol You're sexy but helllllll nawwwwwwwww to that one (created_at: 2014-02-27 18:08:51)
4.	maro0ha_a Harry's pure voice makes me melt!!! (created_at:2014-04-26 13:40:30)
5.	djkrunc I fux wit it (created_at: 2012-09-24 20:17:54)

Pelabelan yang dilakukan mengacu pada sepuluh kategori *cyberbullying*, yaitu: (1) ancaman kekerasan fisik melalui teks. (2) penghinaan terkait kebencian. (3) pemanggilan nama (termasuk homophobia). (4) ancaman kematian. (5) berakhirnya hubungan platonis. (6) pelecehan seksual melalui teks. (7) penuntutan/instruksi. (8) mengancam untuk merusak suatu hubungan. (9) ancaman yang ditujukan untuk keluarga. (10) ancaman melalui pesan[6]. Setiap komentar yang memenuhi salah satu dari sepuluh kategori tersebut diidentifikasi sebagai komentar yang mengandung *cyberbullying*[6]. Kategori tersebut dihasilkan dari perilaku pengguna perangkat komunikasi seperti email, blog, halaman web, dan situs jejaring sosial. Berdasar kategori tersebut dirumuskan kategori yang sesuai dengan penelitian ini karena menggunakan data yang diambil dari komentar instagram, maka terdapat tiga kategori komentar yang termasuk ke dalam *cyberbullying* yaitu: (1) Mengancam target dengan cara menimbulkan rasa takut, (2) Menghina target dengan menggunakan kata yang menyudutkan, memalukan, dan mengutarakan kebencian, (3) Pelecehan seksual.

Label yang digunakan ketika penelitian merupakan hasil dari *vote* ketiga responden dalam melakukan pelat-

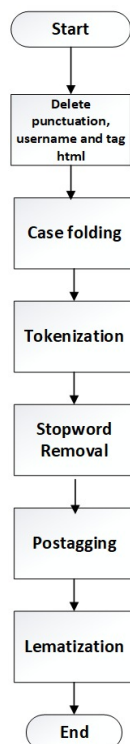
belan Bagian ini berisi dua sub-bagian, yaitu Hasil Pengujian dan Analisis Hasil Pengujian. Pengujian dan analisis yang dilakukan selaras dengan tujuan TA sebagaimana dinyatakan dalam Pendahuluan. Beberapa contoh data hasil pelabelan dapat dilihat pada tabel II dibawah:

Tabel II. HASIL PELABELAN DATA

Komentar	Label
<i>I hate one direction soooo mutch amd you and the fans seid bitches und nooottt nice and i kill the band</i>	1
<i>Perfect body, beyond terrible boobs</i>	1
<i>Bigboy loser fuckface</i>	1
<i>Fuck Yessss! Amazing!!!!</i>	1
<i>If u don't stop reading this u wil die at 11:01 this is a curse repost on seven pics</i>	1
<i>Nice to meet you man! Keep killing it!</i>	0
<i>Shiiet, ya the most humble and crazy mofo I ever met</i>	0
<i>You were funny on that tattoo show. Best by far!</i>	0
<i>Congratulations Jenni!! You guys are going to be great parents</i>	0
<i>Fucking impressive!!</i>	0

B. Preprocessing

Kalimat komentar yang terdapat pada dataset masih mengandung banyak *noise*. *Noise* dapat mempengaruhi hasil akhir pemodelan data, sehingga perlu untuk dihilangkan agar kalimat yang akan diproses merupakan kalimat yang berkualitas. Gambar 2 merupakan skema *preprocessing* pada penelitian ini.



Gambar 2. Gambaran *Preprocessing*

Tahapan awal dalam pemrosesan data adalah tahap

preprocessing. *Raw data* tersebut akan melalui *preprocessing* dengan tahapan seperti berikut:

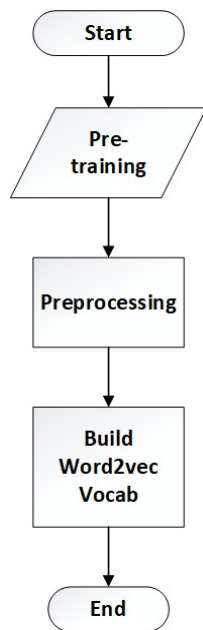
- 1) Menghilangkan simbol-simbol, tanda baca, user-name, dan *tag html*. Contoh:
 - a) "`maro0ha_a Harry's pure voice makes me melt!!!`" menjadi "Harrys pure voice makes me melt".
 - b) "`adrian_d_t Just be happy people are showing you love`" menjadi "Just be happy people are showing you love".
- 2) Tahapan *case folding* memiliki tujuan dalam mengubah huruf kapital menjadi huruf kecil (*lowercase*).
 - a) "Harrys pure voice makes me melt" menjadi "harrys pure voice makes me melt".
 - b) "Just be happy people are showing you love" menjadi "just be happy people are showing you love".
- 3) Tokenisasi sebagai tahapan mengubah setiap kata menjadi token.
 - a) "harrys pure voice makes me melt" menjadi "'harrys', 'pure', 'voice', 'makes', 'me', 'melt'".
 - b) "just be happy people are showing you love" menjadi "'just', 'be', 'happy', 'people', 'are', 'showing', 'you', 'love'".
- 4) *Stopword Removal* digunakan dalam menghapus token kata yang tidak penting. Token yang terdapat dalam list *stopword* akan dihapus.
 - a) "'harrys', 'pure', 'voice', 'makes', 'me', 'melt'" menjadi "'harrys', 'pure', 'voice', 'makes', 'me', 'melt'".
 - b) "'just', 'happy', 'people', 'showing', 'you', 'love'" menjadi "'just', 'happy', 'people', 'showing', 'you', 'love'".
- 5) Penambahan *postagging* berguna untuk membantu ketepatan proses selanjutnya dalam penyederhanaan fitur kata.
 - a) "'harrys', 'pure', 'voice', 'makes', 'me', 'melt'" menjadi "['harrys', 'NNS'], ['pure', 'JJ'], ['voice', 'NN'], ['makes', 'VBZ'], ['me', 'PRP'], ['melt', 'VB']".
 - b) "'just', 'happy', 'people', 'showing', 'you', 'love'" menjadi "['just', 'RB'], ['happy', 'JJ'], ['people', 'NNS'], ['showing', 'VBG'], ['you', 'PRP'], ['love', 'VBP']".
- 6) Lematisasi sebagai tahap penyederhanaan bentuk kata menjadi kata dasar bertujuan dalam memperkecil ukuran fitur.
 - a) "'harrys', 'pure', 'voice', 'makes', 'me', 'melt'" menjadi "'harry', 'pure', 'voice', 'make', 'me', 'melt'".
 - b) "'just', 'happy', 'people', 'showing', 'you', 'love'" menjadi "'just', 'happy', 'people', 'show', 'you', 'love'".

C. TFIDF

Tahapan *preprocessing* menghasilkan data yang siap untuk diolah pada proses pembobotan setiap kata (term). Bertujuan untuk mencari fitur-fitur kata dan menghitung frekuensi dari setiap kata (TF) serta dihitung nilai bobot frekuensi pada setiap kata (W_{tf}). Kemudian akan dilakukan perhitungan dokumen yang mengandung sejumlah kata (DF) dari frekuensi kata yang telah ditemukan dan melakukan inverse nilai DF ($idft$). Nilai pembobotan TF-IDF didapatkan dari hasil perkalian proses perhitungan term frekuensi dan inverse dari DF. Implementasi TFIDF dapat menggunakan *library*, terdapat dua *library* yaitu *Tfidftransformer* dan *Tfidfvectorizer*. Pada penelitian ini menggunakan *Tfidfvectorizer* dalam membantu menyelesaikannya.

D. Word2vec

Word2vec digunakan dalam memetakan setiap fitur kata ke dalam bentuk vektor. Vektor yang dimiliki setiap fitur kata memiliki tujuan untuk mencari kemiripan semantik antar kata dengan melihat kedekatan antar vektor. Gambaran sistem yang membangun kamus *word2vec* ditunjukkan pada Gambar 3:



Gambar 3. Membangun Kamus Word2vec

E. Penambahan informasi semantik

Penambahan informasi semantik dilakukan pada tahap testing berdasarkan wawasan kamus *word2vec* yang telah dibangun. Setiap fitur hasil ekstraksi dari data uji dicocokkan dengan fitur ekstraksi pada data latih. Apabila terdapat fitur ekstraksi data uji yang tidak terdapat dalam data latih maka dicari penggantinya dari kamus *word2vec* menggunakan fungsi *most_similar* dengan mengambil sepuluh fitur kata teratas yang secara *default* menerapkan perhitungan *cosine-similarity*. List kata yang terpilih dicari pada data latih. Apabila kata tersebut ditemukan

dalam daftar data latih, maka fitur kata yang tidak ada diganti menggunakan fitur kata yang mirip dengan kata yang tidak ada tersebut, yang terdapat pada data latih. Berdasarkan analisa pada *pre-training* dataset, setiap kalimat komentar pada data uji memiliki setidaknya satu kata yang tidak terdapat dalam fitur kata pada data latih. Secara umum penambahan informasi tersebut dituliskan dalam Algoritma 1. Penambahan informasi ini bertujuan untuk membuat data uji yang belum dikenali oleh model, agar dapat diklasifikasikan dengan tepat.

Algorithm 1 Mencari sinonim kata pada data uji dengan pendekatan perhitungan semantik

```

1: procedure SEMANTIC SIMILARITY
2:   for every feature ∈ sentence do
3:     if feature ∉ TFIDF vocabulary then
4:       search top ten similar words ∈ word2vec
       dictionary()
5:     for each word ∈ top ten words do
6:       if word ∈ TFIDF vocabulary then
7:         old features ← new word ∈ TFIDF
8:       end if
9:     end for
10:  end if
11: end for
12: end procedure
    
```

F. Klasifikasi Menggunakan Support Vector Machine

Tahap Klasifikasi dilakukan dengan pendekatan supervised learning menggunakan metode Support Vector Machine. Data input dalam proses klasifikasi didapatkan dari hasil ekstraksi fitur menggunakan TFIDF yang sudah berbentuk vektor. Klasifikasi yang akan dilakukan menggunakan bantuan library Support Vector Machine (*LibSVM*). Library *LibSVM* adalah perangkat lunak terpadu yang digunakan untuk mendukung klasifikasi vektor. Jenis kernel yang digunakan pada penelitian ini adalah kernel *polynomial*.

G. Evaluasi

Evaluasi yang dilakukan dalam penelitian ini menggunakan akurasi yang dapat diukur dari hasil prediksi model dalam klasifikasi menggunakan *confusion-matrix*. *Confusion-matrix* yang digunakan dapat dilihat pada tabel III sehingga memiliki persamaan seperti terlihat pada Persamaan (7) [8].

Tabel III. CONFUSION-MATRIX

		Prediction		Total
		(Negative)	(Positive)	
Actual	(Negative)	p	q	$p+q$
	(Positive)	u	v	$u+v$
Total		$p+u$	$q+v$	m

$$Accuracy = \frac{p+v}{p+q+u+v} = \frac{p+v}{mv} \quad (7)$$

Keterangan:

- p : Hasil prediksi dan hasil sebenarnya *non-cyberbullying*.
- q : Hasil prediksi *cyberbullying*, tetapi hasil sebenarnya *non-cyberbullying*.
- u : Hasil prediksi *non-cyberbullying*, tetapi hasil sebenarnya *cyberbullying*.
- v : Hasil prediksi *cyberbullying*, tetapi hasil sebenarnya *cyberbullying*.

IV. HASIL DAN PEMBAHASAN

Secara umum penelitian ini bertujuan mengetahui pengaruh penambahan informasi semantik. Penambahan informasi semantik dapat membuat setiap kata pada data uji tercakup dalam fitur data latih. Tabel IV merupakan daftar beberapa kata data uji yang diganti oleh fitur pada data latih yang memiliki kemiripan domain. *Word2vec* menggunakan *context words* untuk mencari *similarity* sehingga kata-kata yang memiliki kemiripan domain akan memiliki skor yang hampir sama. Kata-kata yang dipilih sebagai pengganti kata yang tidak terdapat dalam fitur pada data latih didapatkan dari sepuluh kata berdasarkan referensi kamus *word2vec* dengan kata terpilih merupakan kata yang terdaftar dalam fitur pada data latih.

Tabel IV. DAFTAR KATA DENGAN KEMIRIPAN SEMANTIK.

No.	Kata yang diganti	Kata Pengganti
1.	<i>badass</i>	<i>scary</i>
2.	<i>ignore</i>	<i>answer</i>
3.	<i>party</i>	<i>gym</i>
4.	<i>perfection</i>	<i>talented</i>
5.	<i>blood</i>	<i>injured</i>
6.	<i>trophy</i>	<i>winner</i>
7.	<i>sing</i>	<i>lyric</i>
8.	<i>vagina</i>	<i>bone</i>
9.	<i>typical</i>	<i>female</i>
10.	<i>sweetest</i>	<i>gift</i>
5.	<i>reply</i>	<i>answer</i>
6.	<i>murderer</i>	<i>fitness</i>
7.	<i>doctor</i>	<i>public</i>
8.	<i>fear</i>	<i>low</i>
9.	<i>visit</i>	<i>finish</i>
10.	<i>whoever</i>	<i>besides</i>

Tabel V menunjukkan beberapa kata yang diganti pada data uji dan memberikan hasil prediksi benar ketika SVM memberikan hasil yang salah. Salah satu contoh kalimat komentarnya adalah “*a little party never killed nobody*”, kata “*party*” tidak terdapat dalam data latih yang diganti dengan kata “*gym*”. Kedua kata tersebut sama-sama termasuk kelompok aktivitas.

Tabel V. DAFTAR KATA DENGAN KEMIRIPAN SEMANTIK YANG MEMBERIKAN HASIL PREDIKSI BENAR KETIKA SVM MEMBERIKAN HASIL YANG SALAH

No.	Kata yang diganti	Kata Pengganti
1.	<i>tatto</i>	<i>pervert</i>
2.	<i>reply</i>	<i>answer</i>
3.	<i>idk</i>	<i>bother</i>
4.	<i>party</i>	<i>gym</i>
5.	<i>beater</i>	<i>grader</i>

Tabel VI merupakan beberapa daftar kata yang diganti pada data uji dan memberikan hasil prediksi salah ketika SVM memberikan hasil yang benar. Salah satu contoh kalimat komentarnya adalah “*u look fuked birdman get rid of the tatto man u get the doe to get rid of them*”, kata “*u, fuked, rid, tatto, doe*” tidak terdapat dalam data latih yang diganti dengan kata *guy, tool, blow, mix*. Kata “*fuked*” digantikan dengan kata “*tool*” yang tidak memiliki kedekatan makna, selain itu kata “*doe*” yang mempunyai arti “kelinci betina” digantikan dengan “*mix*” yang merupakan sebuah kata kerja.

Tabel VI. DAFTAR KATA DENGAN KEMIRIPAN SEMANTIK YANG MEMBERIKAN HASIL PREDIKSI SALAH KETIKA SVM MEMBERIKAN HASIL YANG BENAR

No.	Kata yang diganti	Kata Pengganti
1.	<i>u</i>	<i>guy</i>
2.	<i>tatto</i>	<i>blow</i>
3.	<i>prey</i>	<i>politics</i>
4.	<i>fuked</i>	<i>tool</i>
5.	<i>born</i>	<i>prove</i>

Kata-kata yang diganti dapat berupa kata yang mirip berdasarkan bentuk katanya, kedekatan makna, memiliki konteks yang sama, bahkan suatu makna yang berlawanan. Misalnya kata *perfection* yang artinya adalah kesempurnaan diganti dengan kata *talented* karena keduanya merupakan kata sifat dengan makna yang positif. Berdasar sepuluh kata dengan *similarity score* tertinggi pada Gambar 4, kata yang terpilih dari kamus *word2vec*, kata *talented* merupakan salah satu yang memiliki makna terdekat dan memenuhi kondisi fitur yang terdapat pada data latih. Contoh lainnya kata *trophy* diganti dengan *winner* karena masih dalam suatu konteks yang sama. Beberapa kata diganti oleh lawan kata seperti kata *ignore* diganti dengan *answer* yang memiliki makna yang berlawanan.

```
[('talented', 0.9981449842453003),
('ikr', 0.9977477192878723),
('ridiculous', 0.997543454170227),
('angie', 0.9973456859588623),
('sing', 0.9973452687263489),
('yay', 0.9973209500312805),
('ugh', 0.9972549080848694),
('hahah', 0.9971303939819336),
('especially', 0.9969629049301147),
('hahahahaha', 0.9969158172607422)]
```

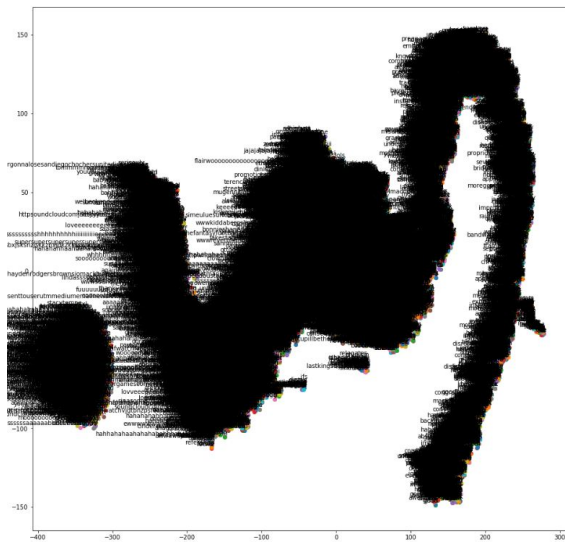
Gambar 4. Daftar Kata Terpilih pada Kamus *Word2vec*

Kata-kata yang digunakan pada kalimat komentar mengandung banyak tipe penulisan yang menyebabkan kata yang sebenarnya terdapat dalam pengujian harus diganti karena perbedaan ejaan dalam kata. Contohnya seperti kata *fuck* memiliki tipe penulisan lebih dari satu yaitu *fuckk, fuked* dan lain sebagainya yang mengulang tiap karakter tetapi sebenarnya memiliki kemiripan makna.

Perbedaan tersebut juga menyebabkan penggantian kata yang tidak sesuai.

Peningkatan akurasi yang diperoleh sangat bergantung pada kamus yang dibangun. Hal tersebut disebabkan oleh vektor yang dipetakan sesuai dengan kekayaan topik bahasan korpus yang mempengaruhi sepuluh daftar kata teratas yang diperoleh dari *word2vec*. Pada penelitian ini hanya terdapat sekitar 700 topik pembahasan pada 62.088 komentar, yang artinya masih sempit dalam menemukan makna yang sesuai dengan topik pembahasan dalam data uji. Selain itu, panjang kalimat komentar yang digunakan dalam membangun kamus *word2vec* tidak seragam, bahkan banyak terdapat query yang hanya terdiri atas dua atau tiga kata setelah melalui tahap praproses. Sedangkan nilai vektor kata didapatkan dari perhitungan ketetanggaan kata disekitarnya, hal ini ditunjukkan dengan nilai *window* yang digunakan.

Perhitungan yang digunakan untuk dapat mengetahui pengaruh penambahan informasi semantik pada penelitian ini yaitu menggunakan akurasi. Akurasi menunjukkan kenaikan 7% dengan menggunakan penambahan informasi semantik. Pemodelan klasifikasi pada kedua skenario tersebut menggunakan SVM dengan parameter optimal dengan jenis kernel yang digunakan *polynomial* dengan nilai degree adalah 2, sedangkan pemodelan *word2vec* juga menggunakan parameter optimal dengan nilai *minimal count* 1 dan *window* 3. Gambar 5 merupakan hasil visualisasi vektor *word2vec*.

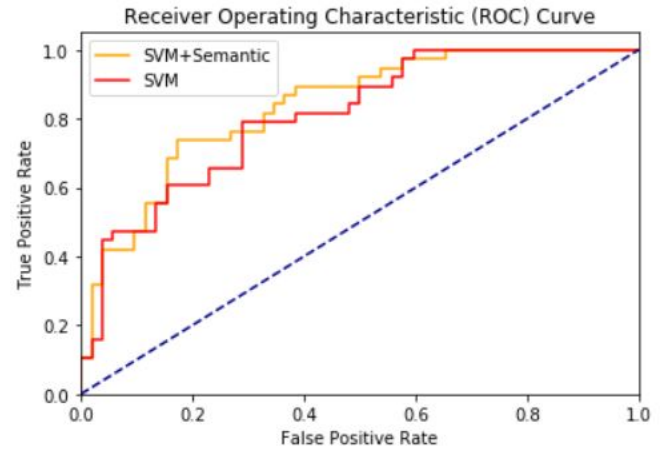


Gambar 5. Visualisasi Vektor *Word2vec*

Tabel VII menunjukkan hasil perbandingan tersebut. Perbandingan performansi juga dapat dilihat pada Gambar 6 *Receiver Operating Characteristic Curve* (Kurva ROC).

Tabel VII. PERBANDINGAN PERFORMANSI

No.	Metode	Akurasi
1.	SVM	67%
2.	SVM + Semantic Similarity	74%



Gambar 6. Kurva *Receiver Operating Characteristic*

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Kualitas kamus yang dibangun dalam melakukan pemetaan setiap fitur kata berpengaruh pada hasil pengujian. Kamus yang dibangun oleh *word2vec* menggunakan 62.088 komentar instagram. Berdasar *dataset* tersebut terdapat sekitar 700 topik pembahasan. Jumlah tersebut membuat keragaman kata yang digunakan masih sangat sempit sehingga makna yang terbentuk juga tidak bervariasi. Hal tersebut menyebabkan tidak semua kata pengganti memiliki kesamaan makna pada kata yang diganti. Ketidaksamaan makna ini terjadi karena *word2vec* adalah *domain similarity*. *Pre-training* data menggunakan *dependency-based word embeddings* mungkin memberikan hasil lebih baik.

B. Saran

Penelitian selanjutnya diharapkan dapat meningkatkan performansi dengan memilih korpus yang digunakan untuk membangun kamus dengan variasi yang lebih luas. Melakukan tahap praproses sesuai dengan kondisi data dengan menormalisasi kata sehingga mengurangi jumlah kata yang tidak terdapat dalam data training (*out of vocabulary*). Dengan demikian, penggantian fitur hanya pada kata-kata yang benar-benar tidak terdapat pada data latih.

REFERENCES

- [1] L. Hackett. *Ditch The Label : The Annual Cyberbullying Survey 2017*. 2017.
- [2] Patricia W. Agatston Robin M. Kowalski, Susan P. Limber. *Cyberbullying: Bullying in the Digital Age*. 2012.
- [3] Hani Nurrahmi and Dade Nurjanah. Indonesian twitter cyberbullying detection using text classification and user credibility. *In 2018 International Conference on Information and Communications Technology (ICOIACT)*, pages 543–548, 2018.
- [4] Geetika Gautam and Divakar Yadav. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. *7th International Conference on Contemporary Computing*, pages 437–442, 2014.

- [5] Xin Rong. word2vec parameter learning explained. arXiv:1411.2738 [cs.CL], 2016. Online; Accessed 9 November 2019.
- [6] Ian Riversa and Nathalie Noretb. 'i h8 u': findings from a five-year study of text and email bullying. *British Educational Research Journal*, 36:643–671, 2010.
- [7] Jiawei Han Micheline Kamber Jian Pei. *Data Mining: Concepts and Techniques 3rd Edition*. 2011.
- [8] Carlo Vercellis. *Business Intelligence: Data Mining and Optimization for Decision Making*. 2004.
- [9] M. Rajasekhara Babu P. Venkata Krishna and Ezendu Ariwa. *Global Trends in Information Systems and Software Applications: 4th International Conference*. 2011.
- [10] Sylvie Ranwez Sébastien Harispe, Stefan Janaqi and Jacky Montmain. *Semantic Similarity from Natural Language and Ontology Analysis*. 2015.
- [11] Martin Warin. Using wordnet and semantic similarity to disambiguate an ontology. 2004.
- [12] Greg Corrado Tomas Mikolov, Kai Chen and Jeffrey Dean. Efficient estimation of word representations in vector space[.]. *Conference: Proceedings of the International Conference on Learning Representations (ICLR 2013)*, 1:28–36, 2013.
- [13] Rahat Ibn Rafiq Sabrina Arredondo Mattson, Richard Han and Qin Lv. Analyzing labeled cyberbullying incidents on the instagram social network. *International Conference on Social Informatics*, pages 49–66, 2015.