Analisis Sentimen *Hate Speech* Pada Portal Berita *Online* Menggunakan *Support Vector Machine* (SVM)

Aniq Noviciatie Ulfah*[1], M. Khairul Anam[2]
[1,2]STMIK Amik Riau; Jl. Purwodadi Ujung Km. 10, telp/fax(0761)589561
[3]Jurusan Teknik Informatika, STMIK Amik Riau, Pekanbaru
e-mail: *[1]aniqnoviciatieulfah@sar.ac.id, [2]khairulanam@sar.ac.id

Abstrak

Salah satu bentuk tindak kriminal yang bisa dijerat dengan undang-undang ITE adalah Hate speech. Namun saat ini netizen di Indonesia masih banyak menggunakan kata-kata Hate Speech dalam mengomentari berita di media online. Dampaknya adalah banyak netizen saat ini yang diperkarakan ke kepolisian oleh pihak yang merasa dirugikan. Hal ini terjadi karena kurangnya informasi dari netizen mengenai hate speech. Penelitian yang akan dilakukan terkait pendeteksian kata-kata yang mengadung Hate Speech pada porta berita online. Pendekatan yang digunakan untuk melakukan pendeteksian kata-kata Hate Speech menggunakan Neural Language Processing dengan menggunakan metode Support Vector Machine (SVM). Untuk mengukur tingkat keakuratan dilakukan beberapa skenario uji coba sehingga tingkat keakuratannya menjadi lebih baik. Komentar pada penelitian ini didapat pada sebuah portal berita online terpopuler di Indonesia. Algoritma SVM dapat diterapkan dalam menganalisa komentar terkait isu politik yang mengngandung Hate Speech dengan nilai akurasi yang bisa sebesar 53. 88% serta nilai Recall adalah 49,69%, Precision adalah 48,77%, Classification error adalah 46,12% dan fmeasure adalah 49.23%.. Dengan adanya penelitian yang akan dilakukan ini bisa menjadi rujukan portal berita untuk menerapkan sistem filtering sehingga kedepannya kasus-kasus mengenai Hate Speech ini dapat diminimalisir.

Kata kunci—Portal berita, *Hate Speech, Support Vector Machine*

Abstract

Hate speech is a form of crime in which the violator threatened with punishment by ITE law. But now netizens in Indonesia still use many of the words of Hate Speech in commenting on news in the online media. The impact of this situation is many netizens currently being sued by the police for those who feel disadvantaged. It happens because of the lack of information from netizens about hate speech. The research to be conducted is related to the detection of words that contain Hate Speech in the online news port. The approach used to detect Hate Speech words uses Neural Language Processing using the Support Vector Machine (SVM) method. Several trial scenarios were carried out to measure the level of accuracy so that the level of accuracy was better. Comments on this study obtain on the most popular online news portal in Indonesia. SVM algorithm applied in analyzing comments related to political issues that contain Hate Speech with an accuracy value that can be as big as 53. 88% and Recall value is 49.69%, Precision is 48.77%, Classification error is 46.12%, and feature is 49.23%. With this research to be conducted, it can become a news portal reference for implementing a filtering system to reduce the possibility of Hate Speech cases.

Keywords—News Portal, *Hate Speech, Support Vector Machine*

1. PENDAHULUAN

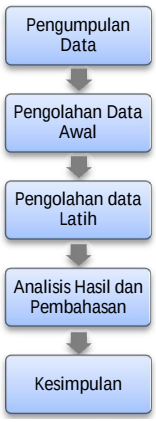
Indonesia merupakan salah satu negara dengan pengguna internet terbesar di dunia, Hal ini terbukti bahwa Indonesia menduduki peringkat ke 4 dunia berdasarkan dengan waktu penggunaan internet perhari [1]. Internet banyak dimanfaatkan oleh pengguna salah satunya yaitu pemanfaatan internet di bidang sosial – politik.Pengguna internet ini biasa disebut dengan *netizen*. *Netizen* atau masyarakat *online* yang mengakses bidang ini sekitar 50,46% dari total pengguna internet di Indonesia yakni sebesar 143,26 juta jiwa [2]. Netizen tidak hanya mengakses hal terkait sosial-politik juga memberikan pendapat atau komentar terhadap isu politik tersebut. Para *netizen* ingin memberikan partipasi politik mereka pada media *online*, aktivitas seperti ini biasa disebut dengan politik *online*. Politik *online* merupakan kegiatan partisipasi politik yang dilakukan oleh *netizen* untuk melakukan aktivitas politik secara *online*. Aktivitas tersebut bertujuan untuk memberikan opini mereka terhadap isu politik pada portal web berita ataupun berinteraksi kepada sesama *netizen* untuk membahas mengenai isu politik [3]. Komentar yang mengandung kata-kata *Hate Speech* masih bisa ditemui diberbagi portal berita, Karena pada portal berita tidak ada fitur untuk melakukan filter pada setiap komentar yang masuk. Sampai saat ini portal berita hanya terdapat fitur lapor agar admin dapat menghilangkan kata-kata tersebut. Hal ini bisa menimbuklan banyak masalah bila tidak cepat ditanggapi oleh admin, komenter tersebut bisa dilaporkan oleh *netizen* lain kepihak yang berwajib.

Beberapa peneliti telah melakukan riset mengenai komentar *netizen*, diantaranya adalah penelitian mengenai sentimen analisa pada situs-situs akomodasi, tempat belanja dan kuliner yang ada dikota kupang menggunakan algoritma *naive bayes* dari penelitian ini akurasi yang dihasilkan untuk menentukan komentar positif, negatif dan netral adalah 66,22% [4]. Penelitian lain melakukan klasifikasi sentimen pada komentar di *facebook*, penelitian ini menggunakan juga menggunakan algoritma *naive bayes* pada komentar negatif dan positif dengan tingkat keakurasian sebesar 83% [5]. Selanjutnya penelitian menggunakan metode *K-Nearest Neighbor* (KNN) untuk mengklasifikasi komentar positif dan negatif untuk mengetahui sikap seseorang dalam konteks dokumen pada sosial media *Facebook*, tingkat keakurasian dari penelitian yang dilakukan cukup baik, yaitu sebesar 79,21% [6]. Kemudian penelitian untuk melakukan analisa sentimen di twitter mengenai program televise [7]. Setelah dilakukan analisa kemudian peneliti melakukan klasifikasi kedalam kelas positif dan negatif menggunakan metode *lexicon-based* dan *Support Vector Machine*. Tingkat keakurasian yang dihasilkan pada setiap program berbeda mulai dari 50%-80%.

Penelitian yang dilakukan menggunakan metode SVM yang digunakan untuk mendeteksi kata-kata *Hate Speech* pada komentar di media *online*. Nantinya penelitian bisa dilihat tingkat keakurasiannya dalam mengklasifikasi *hate speech*. Untuk mengukur tingkat keakuratan dalam klasifikasi nantinya akan menggunakan *rapidminer* dan akan dilakukan beberapa skenario uji coba dalam mengukur tingkat keakuratan sehingga memperoleh hasil yang lebih baik. Harapan dari penelitian ini adalah bisa digunakan sebagai rujukan media *online* dalam melakukan filtering komentar-komentar dari *netizen* Indonesia. Sehingga kedepannya kasus-kasus yang terkait dengan *Hate Speech* di media *online* dapat diminimalisir.

2. METODE PENELITIAN

Adapun kerangka kerja dalam penelitian ini sebanyak 5 tahapan yang digambarkan berikut:



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja diatas, maka masing–masing tahapan tersebut dapat dijelaskan sebagai berikut:

2.1. Pengumpulan Data

Proses pengumpulan data awal dilakukan untuk mengidentifikasi masalah yang ada. Metode yang dilakukan yaitu dengan mengumpulkan data terkait isu politik di Indonesia melalui portal berita detik.com. Dokumen yang diperoleh merupakan dokumen mentah. Dokumen mentah ini mengandung bagian-bagian yang tidak berarti bagi proses klasifikasi *sentiment analysis*, misalnya *stopword* (kata penghubung) dalam bahasa Indonesia. Agar dokumen mentah tersebut dapat diubah menjadi suatu representasi/dokumen dengan format yang sesuai untuk algoritma *learning* yang digunakan dalam proses klasifikasi dari suatu opini, maka perlu dilakukan proses pengolahan data set agar siap digunakan sebagai inputan pada tahap *document preprocessing* [8].

1. Penyaringan dokumen (*document filtering*)
   Pemfilteran dokumen, merupakan suatu proses untuk menghilangkan bagian-bagian dari dokumen mentah yang tidak mempunyai relevansi atau arti bagi proses klasifikasi [9]. Contoh, tanggal yang mungkin terdapat dalam dokumen mentah, label, topik, atau elemen-elemen klasifikasi lainnya yang dsertakan dalam dokumen, akan dihilangkan karena elemen-elemen tersebut menspesifikasikan nilai atau kategori yang sebenarnya didapatkan melalui cara lain, yaitu algoritma *learning.*
2. *Case folding*
   Merupakan proses penyamaan *case* dalam sebuah dokumen. Ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf capital. Oleh karena itu peran pada tahap ini dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standard (huruf kecil).
3. Tokenisasi
   Merupakan proses untuk membagi teks yang berasal dari kalimat atau paragraph menjadi bagian-bagian tertentu [10]. Contoh tokenisasi dari kalimat “saya suka membaca buku” menghasilkan empat token yaitu “saya”, “suka”, “membaca”, “buku”. Biasanya yang menjadi acuan pemisah antar token adalah spasi dan tanda baca. Tokenisasi seringkali dipakai dalam ilmu linguistic dan hasil tokenisasi berguna untuk analisis teks lebih lanjut.

4. Penghapusan kata-kata penghubung (*stopword elimination*)
   Kata penghubung (*stopword)* didefinisikan sebagai sebuah kata yang sangat sering muncul dalam suatu dokumen teks yang kurang memberikan arti penting terhadap isi dokumen [11]. Kata depan dan konjungsi merupakan kandidat besar dari kata penghubung yang harus dihilangkan. Contoh "yang", "di", "dan", "itu", dan lain sebagainya. Langkah ini bermanfaat untuk mengurangi jumlah *feature* yang akan digunakan.

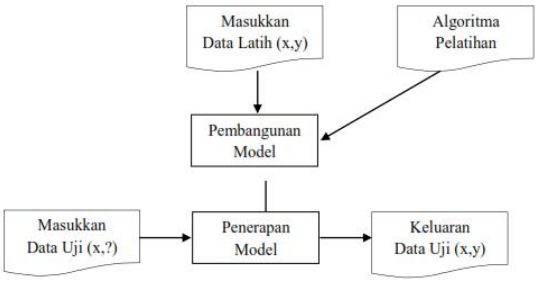
2.2. Pengolahan Data Awal

Merupakan tahapan dalam mempersiapkan data yang diperoleh (data mentah) dirubah menjadi data siap diolah disimpan dalam format .xlsx (Ms. Excel). Tahap awal dalam pengolahan data yaitu komentar pada portal dirubah menjadi format CSV/Excel sehingga bisa di proses untuk tahap selajutnya. Pada tahap kedua yaitu *document filtering* dan *tokenization* yang bertujuan untuk mengurangi ukuran dari sentimen dengan menentukan *ranking* dari *term* yang paling berpengaruh pada proses klasifikasi [8]. Tahap ketiga yaitu menghilangkan *stopword* (yang, di, ke, dari, dsb) karena dinilai akan mengurangi makna dari kalimat [12]. Tahap selanjutnya adalah *feature selection* yaitu memberikan kelas pada kata dengan metode POS *tagger* bagasa indonesia yang terdiri dari 4 kelas kata bahasa indoensia yaitu [13]: kata sifat, kata keterangan, kata benda, kata kerja.Tahap selanjutnya yaitu transormasi data, dimana data pada tahap awal diberi *coding* (simbol). Hal ini untuk mempermudah dalam menerapkan algoritma klasifikasi (SVM) yang digunakan nantinya. Sebelum di proses ke tahap selanjutnya, data di pecah menjadi dua bagian yaitu data latih dan data uji yang juga disimpan dalam file CSV/XLXs, dimana data latih digunakan untuk pembentukan model, kemudian model yang terbentuk harus diujikan kembali menggunakan data uji.

2.3. Pengolahan Data Latih

Pada tahap ini dilakukan proses implementasi algoritma klasifikasi menggunakan Matlab. Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dua pekerjaan utama klasifikasi: (1) pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya [14].

Model dalam klasifikasi menerima masukkan dan mampu melakukan pelatihan terhadap masukkan dan memberikan jawaban sebagai keluaran dari hasil pelatihannya.Kerangka kerja klasifikasi ditunjukkan pada Gambar 1 dibawah ini.

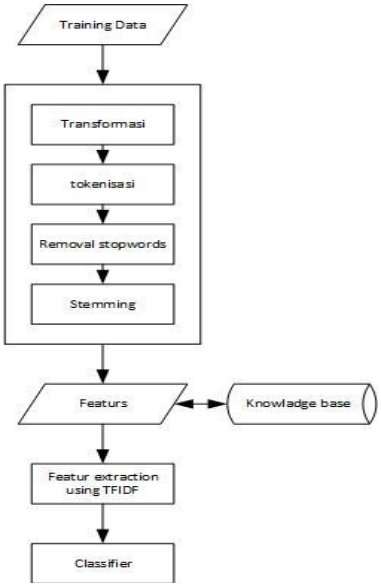


Gambar 2. Proses Pekerjaan Klasifikasi [8]

Model klasifikasi yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yang belum diketahui. Dalam pembangunan model selama proses pelatihan tersebut diperlukan suatu algoritma untuk membangunnya. Pada Gambar 2 terdapat dua langkah proses yaitu induksi merupakan langkah untuk membangun

model klasifikasi dari data latih yang diberikan (proses pelatihan) dan deduksi merupakan langkah untuk menerapkan model tersebut pada data uji sehingga kelas yang sesungguhnya dari data uji dapat diketahui (proses prediksi).

Implementasi algoritma menjelaskan proses penerapan dan pengujian algoritma SVM dalam klasifikasi kalimat yang mengandung *Hate Speech* pada porta berita yang terkait isu politik.Dibawah ini dapat dilihat alur proses implementasi algoritma:



Gambar 3. Alur Proses SVM [15]

2.4. Analisa Hasil dan Pembahasan

Merupakan tahapan dimana dilakukannya proses pengujian akurasi, *recall*, *precision*, *f measure*, *classification error*. Pengujian akurasi, *recall*, *precision*, *f measure*, *classification error* digunakan untuk mengetahui nilai benar dari hasil keluaran sistem yang diuji dengan aplikasi Python.Evaluasi performasi dari SVM dengan menggunakan kategorisasi opini yang dibuat.

3. HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan uraian tentang hasil dan pembahasan yang bertujuan untuk mendapatkan jawaban tentang semua permasalahan dari topik penelitian yang diangkat yaitu Analisis Sentimen (*Hate Speech*) Menggunakan SVM Pada Portal Berita *Online*.

Pengolahan Data Awal

Data yang dikumpulkan merupakan komentar dari situs detik.com, kemudian komentar-komentar tersebut dianalisa berdasarkan kategori *Hate Speech* dan tidak. Berikut ini adalah proses pengolahan data.

Pengambilan Data

Data diambil dari portal berita detik.com selama lima hari mulai dari tanggal 1 Agustus sampai 5 Agustus 2019 dengan jumlah 33 berita. Semua komentar yang ada di berita diambil untuk dijadikan data pada penelitian ini. Berikut Gambar 4 adalah contoh dari komentar yang

telah diambil. Dari sekian banyak komentar hanya 207 komentar yang digunakan untuk data pada penelitian. Berikut ini adalah contoh dari komentar yang telah diambil dapat dilihat pada Gambar 4.

NO	Komentar
1	Si prabowo gak penting, nyapres kalah melulu, nyuruh jongosnya aja gak digubris, ngelus elus kucing sj di hambalang sono...
2	Biarin aja lah... orang ini ambekan. Nanti di ajak kerjasama, Jadi tambah ambekan. Kayak anak kecil mereka. Justru Yang harus di lakukan adalah membuat jera agar 5 tahun lagi nggak mencalonkan diri.
3	Biarin aja,, kalah maksa menang,,udah kerja aja tanpa dia juga 5 thn berhasil,,
4	Tidak perlu.....
5	Buat apa membuang Tenaga memikirkan menanti sebuah jawaban yg kita semua sudah tahu jawabanya....Maju terus Pak Jokowi mbah Yai untuk memajukan bangsa dan mensejaterahkan rakyat ....Udah GASPOL PAK JOKOWI MBAH YAII..
6	curang..... pakai maksa2 segala
7	Pak Jokowi, pls jaga wibawa.. apabila sdh tanya 1x sdh cukup jgn smp berulang2.. tnp rekonsiliasipun sy yakin kepemimpinan bapak pasti baik2 saja..🙏
8	yang penting sudah gugur kewajiban pak Jkw dan tim utk mengajak rekonsiliasi. kalo pak Prab gak mau ya sudah, gak usah dipaksa. manja..

Gambar 4. Contoh Komentar

Klasifikasi Sentiment Secara Manual

Menurut [16], pembentukan kontruksi kategorisasi teks secara otomatis membutuhkan ilmu dari seorang ahli domain. Hal ini berguna untuk melakukan perbandingan analisa sistem dengan analisa ahli Bahasa. Namun pada penelitian ini proses klasifikasi dilakukan secara manual menggunakan excel dan di proses menggunakan *Rapidminer* dengan cara melihat berbagai kasus yang sudah pernah terjadi di di Indonesia. Setelah mendapat data-data kasus yang pernah terjadi di Indonesia, barulah penulis akan mengidentifikasi sampel komentar berdasarkan kasus yang pernah terjadi untuk menentukan komentar tersebut *Hate Speech* atau tidak. Berikut adalah contoh dari klasifikasi secara manual.

Makanya pendukung lu jangan diboongin, ngaku2 menang ampe sujud nungging berkali2, kalah terus teriak dicurangin dah tau pendukungnya banyak yg bodoh yah pasti aja mereka gak mau terimalah. Untung orang ini ga mimpin karena ga cocok buat memimpin negara sebesar dan sekaya ini.	hate speech
Sudah saatnya tim Pak Jokowi dan Pak Prabowo saling membantu untuk menuju Indonesia maju dan juga saling mengingatkan kalo ada kesalahan. Kita negara yg besar, kepentingan rakyat lebih penting daripada kepentingan pribadi dan golongan	tidak

Gambar 5. Contoh Klasifikasi Data Secara Manual

Gambar diatas menunjukkan salah satu contoh klasifikasi manual yang berlandaskan kasus yang pernah terjadi di Indonesia. Kategori kelas akhir ditentukan dari banyaknya kata yang terdapat ada pada kasus yang pernah terjadi, jika terdapat kata yang sama lebih dari 2 kata maka komentar tersebut masuk kategori *hatespeech*. Selanjutnya rangkaian proses yang akan dilakukan dari pengolahan dataset sehingga perhitungan reputasi adalah sebagai berikut.

Pengolahan Data Latih

Sebelum melakukan analisis terhadap sentiment yang dikumpulkan, langkah yang perlu dilakukan adalah pembersihan data yang bertujuan mengurangi dimensi kata yang tidak tidak berpengaruh pada hasil pengolahan data. Sehingga dapat melakukan proses klasifikasi data lebih cepat dengan hasil yang lebih akurat [8]. Pada penelitian ini pemrosesan data menggunakan aplikasi *Rapirminer*. Agar data dapat diolah maka perlu adanya *preprocessing* terlebih dahulu [17]. Untuk langkah-langkah pengolahan data latih sebagai berikut. Gambar 6 dibawah ini merupakan tahapan dalam *preprocessing*.

Gambar 6. Tahapan *Preprocessing* menggunakan *RapidMiner*

1) Case folding

Teks dalam komentar biasanya memiliki beragam penulisan, salah satunya adalah menggunakan huruf besar dan kecil. Untuk mengatasi hal ini, teks akan diubah dalam huruf kecil melalui operator '*transform case'* di *rapidminer*.

2) Tokenization

Tokenisasi adalah proses untuk membagi teks yang berasa dari kalimat atau paragraph menjadi beberapa bagian tertentu [10]. Tahapan ini akan membagi teks dari kalimat berdasarkan spasi dan tanda baca untuk tahap analisa teks tahap selanjutnya.

3) Removal stopwords

Pada proses ini mengunakan kamus data yang dapat di lihat pada Lampiran B. *Stop words* adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopwords* untuk bahasa Inggris diantaranya "of", "the". Sedangkan untuk bahasa Indonesia diantaranya "yang", "di", "ke".

4) Stemming

Pentingnya stemming dalam proses pembuatan sistem temu kembali yakni dimana saat menghilangkan imbuhan pada sebuah kata menjadi hal yang perlu diperhatikan. Karena dalam proses stemming yang penting yakni terlebih untuk menghilangkan imbuhan pada awalan setelah itu akhiran. Apabila yang dilakukan adalah sebaliknya maka tidak akan ditemukan kata dasar yang tepat dan sesuai dengan kamus bahasa. Dimana dari hasil proses tersebut akan didapatkan sebuah informasi mengenai banyaknya term yang muncul dalam sebuah dokumen setelah dilakukan perhitungan *term frequency*.

5) Generate TFIDF

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Metode ini akan menghitung bobot setiap token t di dokumen d dengan rumus:

Wdt = tfdt * IDFt

Dimana :

- d : dokumen ke-d
- t : kata ke-t dari kata kunci
- W : bobot dokumen ke-d terhadap kata ke-t
- tf : banyaknya kata yang dicari pada sebuah dokumen
- IDF : Inversed Document Frequency

Nilai IDF didapatkan dari IDF : log2 (D/df) dimana :

- D : total dokumen
- df : banyak dokumen yang mengandung kata yang dicari

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

Pembentukan Data Uji

Pada penelitian ini, peneliti membuat 2 dataset, proses pembentukan data dan training menggunakan x *validation*. Teknik ini digunkan untuk pembentukan model yang fit dari data training sebaik mungkin. Pada perangkat lunak *Rapidmaner*, hal ini dapat dilakukan dengan operator 'X-*Validation*' dengan parameter *number of validations* (jumlah validasi) dan sampling type (tipe pengambilan sampel).

Metode sampel *stratified sampling* dapat mewakili ciri-ciri populasi yang heterogen, hal ini cocok untuk pada data yang mengadung banyak variasi bahasa dan beragam topik seperti komentar. Jumlah validasi x yang diujicobakan pada penelitian ini sebanyak 1-10 kali. Hal ini bertujuan untuk melihat nilai performa terbaik yang diperoleh masing-masing algortima klasifikasi.

Analisis Hasil dan Pembahasan

Dalam melakukan analisa menggunakan algoritma SVM. Algortima dipilih karena kemampuan meminimalkan *error* dalam data training serta meminimalkan *error* yang dipengaruhi oleh dimensi. Strategi yang digunakan dinamakan *Structur Risk Minimization* (SRM). Faktor yang kedua dipilinya SVM karena merupakan salah satu metode yang dapat digunakan dalam sebuah masalah berdimensi tinggi namun jumlah sampel data terbatas. Yang terakhir adalah kemudahan implementasi SVM pada data yang telah memiliki *library*.

Nilai parameter yang dipilih operator SVM merupakan nilai standar uang *rapidminer* berikan. Perubahan dilakukan pada parameter faktor pinalti C yang merupakan nilai toleransi *misclassification* dari SVM. Nilai yang terlalu tinggi dapat menyebabkan *overfitting* yaitu model yang terbentuk cocok untuk kelompok data tertentu, sebaiknya nilai yang terendah menyebabkan klasifikasi yang terlalu umum. Oleh karena itu, penelitian ini memilih angka inputan untuk C adalah 1. Angka ini memberikan akurasi yang jauh lebih baik dibandingkan nilai standar yang diberikan oleh sistem (0). Menurut [18] untuk mendapatkan nilai parameter yang tepat untuk nilai optimum SVM dapat menggunakan *cross validation*. *Cross validation* juga dinilai menghasilkan performa yang baik untuk pembagian data. Hal inilah yang melandasi penelitian ini menggunakan *cross validation* sebagai metode untuk melakuan pembentukan data testing dan data training. Nilai parameter lainnya merupakan nilai standar yang diberikan oleh sistem dan tidak berpengaruh signifikan pada optimasi akurasi klasifikasi.

Proses klasifikasi

Setelah melakukan *preprocessing* yang dikembangkan dengan mengubah huruf menjadi seragam, tokenisasi, *stopwords*, stemming, dan TFIDF. Proses selanjutnya ialah mengkonversi data menjadi document dengan operator '*process document from data'*. Kemudian menenetukan *field* mana yang akan dijadikan sebagai *class* dengan menggunakan operator '*set role'*. Selanjutnya proses klasifikasi sentiment dilakukan dalam operator'x-*validation*'. Data masukan yang akan diklasifikasi berisi kolom tambahan *class* yang kemudian akan diisi dengan kategori *Hate Speech* dan tidak berdasarkan proses perhitungan dari SVM. Model yang terbentuk dari data training diimplementasikan ke data testing melalui operator '*apply model'*. Tahap terakhir

dari klasifikasi adalah menghitung performansi. Hasil keluaran dalam proses klasifikasi adalah skor probabilitas sentiment *Hate Speech* dan tidak dari masing-masing komentar.

Hasil

Dari proses klasifikasi kita pada melihat pada kategori *Hate Speech* sistem menemukan *Hate Speech* sebanyak 87 komentar sementara untuk kategori tidak *Hate Speech* terdapat 105 komentar dengan rata-rata akurasinya adalah 53,88 %. *Recall* pada penelitian ini adalah 49,69%, *Precision* adalah 48,77%, *Classification error* adalah 46,12% dan *fmeasure* adalah 49.23%.

4. KESIMPULAN

Berdasarkan hasil dan proses analisis yang dilakukan dalam penelitian ini, maka dapat ditarik simpulan yaitu Algoritma SVM dapat diterapkan dalam menganalisa komentar terkait isu politik yang mengandung *Hate Speech* dengan nilai akurasi yang bisa sebesar 53. 88%. Sedangkan nilai *recall* adalah 49,69%, *Precision* adalah 48,77%, *Classification error* adalah 46,12% dan *fmeasure* adalah 49.23%.

5. SARAN

Berdasarkan keterbatasan yang muncul dalam penelitian, maka saran untuk pengembangan penelitian selanjutnya adalah proporsi jumlah data antar sentimen yang digunakan diusahakan seimbang karena dapat mempengaruhi peningkatan nilai akurasi, hal ini dapat dilihat pada nilai akurasi yang dihasilkan ketika penggunaan semua data dan hanya data yang terkait kinerja pelayanan publik.

DAFTAR PUSTAKA

[1] S. Kemp., 2019. "*Digital 2019: Global Internet Use Accelerates*," [Online]. Available: https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates. [Accessed: 04-Apr-2019].

[2] APJI 2017, "*Penetrasi & Perilaku Pengguna Internet Indonesia*," *Apjii*, p. Hasil Survey, https://apjii.or.id/content/read/39/342/Hasil-Survei-Penetrasi-dan-Perilaku-Pengguna-Internet-Indonesia-2017

[3] M. K. Anam, 2017 "*Analisis Respons Netizen Terhadap Berita Politik Di Media Online*," *Jurnal Ilmiah Ilmu Komputer*, Vol. 3, No.1, pp. 14–21, Universitas Islam Indonesia, Yogyakarta.

[4] P. Aliandu, 2015, "*Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City*," *Procedia Computer. Science.*, Vol. 72, pp. 300–305, Widya Mandira Catholoc University, Kupang.

[5] A. R. C and Y. Lukito, 2016 "*Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes*," Vol. 2, No. 2, pp. 26–34, Jurnal Informatika dan Sistem Informasi, Program Studi Teknik Informatika Universitas Ciputra, Surabaya.

[6] A. Salam, J. Zeniarja, and R. S. U. Khasanah, 2018, "*Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa*," in *Prosiding SINTAK*, pp. 480–486, Fakultas Teknologi Informasi Universitas Stikubank (UNISBANK), Semarang.

[7] Tiara, M. K. Sabariah, and V. Effendy, 2015, "*Analisis Sentimen pada Twitter untuk Menilai Performansi Program Televisi dengan Kombinasi Metode Lexicon-Based dan Support Vector Machine*," *e-Proceeding Eng.*, Vol. 2, No. 1, pp. 1237–1247.

[8] Han, J.W., Kamber, M. and Pei, J. 2012, *Data Mining Concepts and Techniques, 3rd Edition*, Morgan Kaufmann Publishers, Waltham.

[9] K. Dave, I. Way, S. Lawrence, and D. M. Pennock, 2003. "*Mining the Peanut Gallery : Opinion Extraction and Semantic Classification of Product Reviews,*" Proceedings of the 12th International Conference, Pages 519–528, ACM in our Didital Library.

[10] C. D. Manning, P. Raghavan, and H. Schütze, 2009 *Introduction to Information Retrieval*, Cambridge University Press Cambridge, England.

[11] B. Patel and D. Shah, 2013 "*Significance of Stop Word Elimination in Meta Search Engine*,".IEEE, pages 52-55, New York.

[12] A. Pak and P. Paroubek, 2010 "*Twitter as a Corpus for Sentiment Analysis and Opinion Mining," pp. 1320–1326,*. Proceedings of the International Conference on Language Resources and Evaluation, LREC.

[13] B. Liu, 2010 "*Sentiment Analysis and Subjectivity*," pp. 1–38, Department of Computer Science University of Illinois at Chicago.

[14] E. Prasetyo, 2012, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, Andi Offset, Yogyakarta.

[15] N. Naw, 2018, "*Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers*," Vol. 8, No. 10, pp. 407–411, IJSRP Publications.

[16] F. Sebastiani, 2002, "*Machine Learning in Automated Text Categorization*, ACM Computing Surveys" Vol. 34, No. 1, pp. 1–47.

[17] S. Mujilahwati, 2016. "*Pre-Processing Text Mining Pada Data Twitter*," Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016) Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan, Yogyakarta.

[18] Akthar, F. and Hahne, C. 2012, *Rapid Miner 5 Operator Reference*, Rapid-I GmbH.