

Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme *Support Vector Machine* (SVM)

Arsya Monica Pravina¹, Imam Cholissodin², Putra Pandu Adikara³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹arsyamonica@yahoo.com, ²imamcs@ub.ac.id, ³adikara.putra@ub.ac.id

Abstrak

Dengan meningkatnya penggunaan Twitter, media sosial yang bekerja secara *real-time* untuk masyarakat dapat menyampaikan keluhan maupun apresiasinya terhadap maskapai-maskapai penerbangan, perlu dibuat sebuah sistem yang dapat melakukan klasifikasi suatu *tweet* yang berisikan opini termasuk ke dalam kelas apa, dalam penelitian ini terdapat kelas positif dan negatif. Hal tersebut dilakukan agar dapat membantu perusahaan maskapai penerbangan dalam hal evaluasi peningkatan pelayanan serta dapat membantu masyarakat dalam memilih maskapai penerbangan dengan tepat. Sehingga dilakukan klasifikasi sentimen dengan fitur *Lexicon Based* yang dapat menerima opini berbahasa lain selain Bahasa Indonesia (dalam penelitian ini digunakan Bahasa Inggris) untuk melakukan analisis sentimen. Digunakan algoritme *support vector machine* untuk melakukan klasifikasi. Hasil dari penelitian ini menunjukkan parameter optimal dan pengaruh penggunaan *Lexicon Based Features*. Dengan digunakan parameter C bernilai 10 dan *learning rate* bernilai 0,03 serta digunakan *Lexicon Based Features* dengan iterasi sebanyak 50 kali memberikan hasil *accuracy* sebesar 40%, *precision* 40%, *100% recall*, dan *f-measure* sebesar 57,14%.

Kata kunci: analisis sentimen, opini maskapai penerbangan, twitter, *support vector machine*, *lexicon based features*

Abstract

With the increasing use of Twitter, social media that works in real-time for the public can convey complaints and appreciation to airlines, it is necessary to create a system that can classify a tweet containing opinions including what is the best class, in this study there are positive and negative classes. This is done so that it can help airline companies in terms of evaluating service improvements and can help people choose the right airline. Thus a sentiment classification with *Lexicon Based features* which is able to receive information in languages other than Indonesian (in this study used in English) is done to conduct sentiment analysis. Use the *support vector machine* algorithm to classify. The results of this study show optimal parameters and the effect of using *Lexicon Based Features*. By using parameter C is 10 and the learning rate is 0.03 also used *Lexicon Based Features* with an iteration of 50 times giving accuracy 40%, precision 40%, recall 100%, and f-measure 57,14%.

Keywords: sentiment analysis, airlines opinion, twitter, *support vector machine*, *lexicon based features*

1. PENDAHULUAN

Twitter merupakan media sosial yang bekerja secara *real-time*, yang memungkinkan pengguna untuk mengekspresikan opini dan perasaan mereka mengenai banyak isu atau permasalahan (Hamdan, Bellot & Bechet, 2015). Twitter memberikan banyak keuntungan bagi masyarakat. Misalnya dalam dunia bisnis, masyarakat dapat mengetahui apakah sebuah layanan, produk, atau lain sebagainya dinilai

baik atau tidak melalui opini masyarakat yang dituliskan di Twitter. Atau keuntungan bagi perusahaan penyedia layanan atau produk tersebut juga dapat menggunakan opini masyarakat tersebut sebagai bahan evaluasi agar dapat meningkatkan kualitas maupun pelayanannya. Disebutkan oleh Direktur Jenderal Sumber Daya Perangkat Pos dan Informatika (SDPP) Kementerian Komunikasi dan Informatika, tercatat sebanyak 19,5 juta masyarakat Indonesia adalah pengguna Twitter dan Indonesia tercatat sebagai negara ke lima

yang paling banyak dan aktif dalam penggunaan Twitter.

Beberapa perusahaan maskapai penerbangan, seperti Garuda Indonesia maskapai bintang lima yang paling dikenal oleh masyarakat, diikuti dengan empat maskapai penerbangan lain yang tidak kalah dikenalanya seperti Citilink anak perusahaan dari Garuda Indonesia dengan harga lebih terjangkau, Batik Air anggota dari Lion Group yang cukup diminati oleh masyarakat, Sriwijaya Air maskapai penerbangan yang cukup berkembang dan kini baru saja menjadi bagian dari perusahaan Garuda Indonesia, dan Lion Air yang juga merupakan anggota dari Lion Group seperti Batik Air yang sangat dikenal oleh masyarakat yang memiliki harga sangat terjangkau. Kelima maskapai penerbangan tersebut, menggunakan media sosial Twitter sebagai salah satu media komunikasi antara perusahaan dengan pelanggannya. Tidak hanya untuk melakukan promosi, perusahaan maskapai tersebut juga tentunya menerima banyak pertanyaan, masukan, kritik, saran, hingga apresiasi dari masyarakat melalui *tweets* yang *diposting* oleh masyarakat tersebut. Hal-hal atau opini yang masyarakat sampaikan pun belum tentu seluruhnya positif atau negatif. Tidak hanya maskapai penerbangan yang membutuhkan *feedback* dari masyarakat, masyarakat juga membutuhkan pengetahuan lebih mengenai apa yang sedang sering terjadi dalam dunia penerbangan, khususnya di Indonesia. Biasanya, masyarakat juga membutuhkan pendapat atau opini orang lain dalam memilih suatu produk ataupun pelayanan, sama halnya dengan memilih maskapai penerbangan yang tepat untuk keperluan pribadi, organisasi, maupun perusahaan.

Opini masyarakat tentang maskapai penerbangan pada dokumen Twitter tersebut perlu dikaji dalam sebagai pemrosesan teks. Analisis sentimen merupakan proses yang sangat dibutuhkan dalam menyaring opini-opini masyarakat dan diklasifikasikan ke dalam kelas positif dan negatif. Sehingga dengan diperolehnya hasil klasifikasi tersebut, dapat membantu kebutuhan perusahaan maupun masyarakat.

Metode *Support Vector Machine* (SVM) akan digunakan dalam proses klasifikasi opini-opini tersebut. Berdasarkan penelitian-penelitian sebelumnya, ditemukan bahwa metode *support vector machine* dapat menghasilkan akurasi yang cukup tinggi dalam melakukan analisis

sentimen. Pada penelitian yang dilakukan oleh Ike Pertiwi Windasari, Fajar Nurul Uzzi, Kodrat Iman Satoto Makalah ditemukan bahwa metode *support vector machine* yang dikolaborasikan dengan ekstraksi fitur TF-IDF menghasilkan akurasi sebesar 86% lebih unggul dibandingkan dengan metode Naïve Bayes yang juga dikolaborasikan dengan TF-IDF ketika melakukan analisis opini masyarakat mengenai Gojek pada Twitter yang dibagi menjadi dua kelas yaitu positif dan negatif (Windasari, Uzzi & Satoto, 2017). TF-IDF atau *Term Frequency-Inverse Document Frequency* ditemukan dapat membantu dalam peningkatan akurasi apabila dikolaborasikan dengan metode SVM dibandingkan dengan ekstraksi fitur lainnya seperti *Ratio* dan *N-Gram* (Juniawan, 2017). Selain itu, dalam penelitian sebelumnya, ditemukan sebuah fitur yang sering digunakan dalam proses analisis sentimen dan membantu dalam mempermudah menganalisis sehingga menghasilkan hasil akurasi yang lebih baik. Fitur tersebut adalah *Lexicon Based Features*, yang pada beberapa penelitian sebelumnya menghasilkan nilai evaluasi *confusion matrix* yang baik yaitu hasil akurasi sebesar 79%, hasil presisi sebesar 65%, hasil *recall* sebesar 97%, dan *f-measure* sebesar 78% (Rofiqoh, 2017).

Dari berbagai referensi penelitian yang ditemukan, Metode *Support Vector Machine* merupakan pilihan metode yang baik dibandingkan metode klasifikasi lainnya yang akan peneliti gunakan dalam analisis sentimen opini terhadap maskapai penerbangan pada dokumen Twitter.

2. TINJAUAN PUSTAKA

2.1. Maskapai Penerbangan

Maskapai penerbangan merupakan organisasi atau jasa layanan transportasi yang tentunya sangat dibutuhkan masyarakat untuk berpindah tempat dari satu kota ke kota yang lain maupun satu negara ke negara lain dengan waktu yang cepat.

2.2. Media Sosial

Media sosial adalah media yang terdiri atas tiga bagian, yaitu infrastruktur informasi beserta alat yang digunakan untuk memproduksi dan mendistribusikan isi media, isi media yang berupa pesan-pesan pribadi, berita, dan, gagasan, serta yang terakhir produk-produk budaya yang berbentuk digital. Kemudian yang memproduksi

dan mengkonsumsi isi media dalam bentuk digital adalah individu, organisasi, dan industri (Howard & Parks, 2012).

2.2.1. Twitter

Twitter merupakan media sosial yang memungkinkan pengguna untuk mengekspresikan opini dan perasaan mereka mengenai banyak isu atau permasalahan (Hamdan, Bellot & Bechet, 2015). Berbeda dengan media sosial yang lain yang harus menjadi teman terlebih dahulu baru dapat berinteraksi, Twitter memungkinkan antarpengguna tetap terhubung walaupun mereka tidak saling berteman (Windasari, Uzzi & Satoto, 2017).

2.3. Analisis Sentimen

Analisis sentimen biasanya melakukan penelitian dalam bentuk analisis, salah satunya mengenai opini dan emosi banyak orang terhadap suatu entitas sebagai contoh permasalahan, topik atau layanan. Dengan penjelasan lain, analisis sentimen merupakan proses ekstraksi emosi atau opini dari sebuah teks atau bacaan (Kaur & Mangat, 2015)

2.4. Text Mining

Text mining merupakan subyek penelitian yang sangat baru dan mulai diminati banyak orang. Dalam penyelesaian masalah, *text mining* biasa digabungkan dengan beberapa subyek lain seperti *Data Mining*, *Natural Language Processing*, dan lain-lain. Dalam *text mining*, terdapat tahap seperti ekstraksi teks menggunakan teknik tertentu, pemrosesan teks atau yang biasa disebut *pre-processing text*, pembobotan atau pemberian indeks pada teks, maupun analisis suatu teks.

Text mining merupakan sebuah proses penemuan informasi, relasi, dan fakta yang tersembunyi di dalam teks ketika dilakukan pemrosesan dan analisis data dalam jumlah besar, struktur teks yang kompleks dan tidak lengkap, dimensi tinggi, serta data yang *noise*. Sangat banyak kegunaan *text mining* yang dibutuhkan dalam kehidupan sehari-hari (Adiwijaya, 2006).

2.4.1. Pre-Processing

Pre-processing atau pemrosesan teks merupakan langkah awal untuk data yang akan diolah masuk pada proses klasifikasi terutama

berfokus pada pembersihan data yang bertujuan untuk menghilangkan dan mengatasi *noisy data*, termasuk mengatasi informasi yang hilang atau tidak lengkap (Adiwijaya, 2006). Tahap ini bertujuan agar nantinya hasil perhitungan akan optimal. Berikut adalah langkah-langkah yang akan dilakukan dalam pemrosesan teks: *case folding*, *cleaning*, *translation*, *stemming*, *stopword removal* dan tokenisasi.

2.4.1.1. Case Folding

Pada tahap ini, dilakukan perubahan seluruh huruf kapital atau huruf besar menjadi huruf kecil (Indraloka & Santosa, 2017).

2.4.1.2. Cleaning

Pada tahap ini, dilakukan penghapusan karakter-karakter selain yang ditentukan seperti huruf atau karakter di luar dari alfabet a-z (termasuk tanda baca), menghapuskan URL atau *link*, menghapuskan *hashtag*, menghapuskan *username*. Biasanya, aturan dalam tiap penelitian berbeda-beda pada tahap *cleaning*. Pada penelitian ini, dilakukan juga pergantian kalimat “batik air” menjadi “batikair”, “garuda Indonesia” menjadi “garudaindonesia”, “sriwijaya air” menjadi “sriwijayaair”.

2.4.1.3. Translation

Pada tahap ini, akan diterjemahkan suatu kalimat berbahasa Inggris ke Bahasa Indonesia untuk setiap katanya. Analisis sentimen saat ini sangat dinamis dalam bidang linguistik komputasi. Selain Bahasa Indonesia, terdapat berbagai bahasa yang digunakan dalam hal berkomunikasi di media sosial khususnya Twitter. Dari berbagai macam Bahasa, penulis memilih Bahasa Inggris yang akan diproses dengan diterjemahkan ke dalam Bahasa Indonesia dalam sistem yang dibuat. Sebelumnya, sempat ditemukan bahwa dengan memerhatikan multibahasa pada sistem akan meningkatkan hasil dari klasifikasi sentimen (Turchi, 2013).

2.4.1.4. Stemming

Stemming akan merubah kata-kata dalam dokumen menjadi kata akar atau dasarnya (*root word*). Proses *stemming* pada dokumen Bahasa Indonesia cukup kompleks, karena harus dilakukan penghilangan seluruh imbuhan pada kata-kata yang terdapat pada *tweets*. Digunakan *library* Sastrawi *Stemming* berbahasa Indonesia

yang berbasis algoritme Nazief dan Adriani (Afuan, 2013).

2.4.1.5. Stopword Removal

Pada tahap ini akan dilakukan penyaringan kata-kata yang sering maupun jarang muncul, biasa disebut dengan *stopword*. Proses ini disebut "*stopword removal*". Dengan menghapus kata-kata yang jarang muncul tersebut, tampaknya akan menjadi optimal untuk mempertahankan kinerja klasifikasi sekaligus mengurangi data sparsial dan menyusutkan ruang fitur secara substansial. (Saif et al., 2014).

2.4.1.6. Tokenisasi

Pada tahap ini, dilakukan pemisahan setiap kata dalam suatu kalimat dalam dokumen. Memisahkan kata biasanya menggunakan spasi. Sebenarnya penulisan dapat berbeda-beda, namun tujuan utamanya adalah yaitu memotong kalimat berdasarkan tiap kata yang menyusun kalimat tersebut (Indraloka & Santosa, 2017).

2.5. Pembobotan Term Frequency-Inverse Document Frequency

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu proses dari teknik ekstraksi fitur dengan proses memberikan nilai pada masing-masing kata yang ada pada *tweets* latih (data latih). Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, akan dilakukan pembobotan atau perhitungan. Pemberian skor dalam TF-IDF berdasarkan frekuensi munculnya kata dalam dokumen. Nilai TF-IDF dapat ditemukan dengan menggunakan Persamaan (1), (2), dan (3).

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

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

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

Keterangan:

$Wtf_{t,d}$ = bobot kata dalam setiap dokumen

$tf_{t,d}$ = jumlah kemunculan kata t dalam dokumen d

N = jumlah seluruh dokumen

df = jumlah dokumen yang mengandung term

idf_t = bobot *inverse* dari nilai df

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

2.6. Lexicon Based Features

Lexicon Based Features adalah metode atau fitur yang digunakan pada sistem ini untuk mencocokkan kata-kata di data latih dengan kamus sentimen berisi kata positif dan kata negatif untuk diketahui tingkat polaritas tiap kata (Peng, 2011), sehingga dapat berfungsi sebagai pengujian. *Lexicon* merupakan suatu himpunan yang telah diketahui sentimennya (Desai & Mehta, 2016). Penerapan *Lexicon Based Features* akan meningkatkan akurasi dengan membantu penambahan bobot dari dokumen yang bersentimen positif maupun negatif.

2.7. Normalisasi Min-Max

Normalisasi data bertujuan untuk mengurangi kesalahan yang terdapat dalam proses *data mining* (Wirawan & Eksistyanto, 2015). Menurut penelitian yang dilakukan sebelumnya, digunakan nilai *newmax* sebesar 0,9 dan *newmin* sebesar 0,1 untuk pembobotan *Lexicon* (Rofiqoh, 2017). Perhitungan dilakukan dengan rumus pada Persamaan (4).

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

Keterangan:

vi' = hasil dari proses normalisasi data ke-i

vi = data ke-i yang dilakukan normalisasi

\min_a = data minimum dari seluruh data a

\max_a = data maksimum dari seluruh data a

newmax = nilai maksimum dari normalisasi

newmin = nilai minimum dari normalisasi

2.8. Support Vector Machine

Support Vector Machine (SVM) termasuk *machine learning (supervised learning)* yang dapat memprediksi kelas berdasarkan dari hasil proses pelatihan. Dengan melakukan pelatihan menggunakan data masukan dalam bentuk numerik dan hasil dari ekstraksi fitur didapatkan sebuah pola yang nantinya akan digunakan dalam proses pelabelan. Nilai atau pola yang dihasilkan dari Metode *Support Vector Machine* sebenarnya adalah sebuah garis pemisah yang disebut dengan *hyperplane*, yang mana garis

tersebut berperan dalam memisahkan **tweet** dengan sentimen positif (berlabel 1) dengan *tweet* yang memiliki sentimen negatif (berlabel 0). Dalam mengambil keputusan dengan metode SVM, digunakan fungsi kernel $K(x_i, x_d)$. Pada penelitian ini akan digunakan persamaan kernel polinomial yang ditunjukkan pada Persamaan (5).

$$K(x_i, x_d) = (X_i^T X_{j+1})^d, \gamma > 0 \quad (5)$$

Pada penelitian ini, digunakan *Sequential Learning* yang merupakan algoritme sederhana untuk memproses data latih dari SVM yang digunakan dengan waktu singkat dibandingkan beberapa algoritme lainnya (Vijayakumar, 1999). Di bawah ini adalah langkah-langkah pelatihan dalam *Support Vector Machine* menggunakan *Sequential Learning*:

1. Inisialisasi parameter: a_i , γ , C , dan ϵ

Keterangan:

α = alfa

γ = konstanta gamma.

C = variabel *slack*.

E = epsilon

2. Hitung matriks Hessian, persamaan ditunjukkan pada (6).

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

Dengan nilai i dan $j=1,2,3,\dots,n$

Keterangan:

x_i = data ke- i

x_j = data ke- j

y_i = kelas data ke- i

y_j = kelas data ke- j

λ = lamda

d = *degree* atau derajat kernel polynomial

$K(x_i x_j)$ = fungsi kernel.

3. Melakukan 3 perhitungan sebagai berikut (sampai batas interaksi):

$$a) E_i = \sum_{j=1}^i a_j D_{ij} \quad (7)$$

Keterangan:

a_j = alfa ke- j

D_{ij} = matriks Hessian

E_i = *error rate*

$$b) \vartheta a_i = \min(\max[\gamma(1 - E_i), a_i], C - a_i) \quad (8)$$

Keterangan:

α_i = alfa ke- i

γ = konstanta gamma

E_i = *error rate*

C = variabel *slack*

$$c) a_i = a_i + \vartheta a_i \quad (9)$$

α_i = alfa ke- i

ϑa_i = delta alfa ke- i

4. Akan didapatkan nilai $SV=(\alpha_i > thresholdSV)$, kemudian melakukan perhitungan nilai bias yang ditunjukkan pada Persamaan (10).

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

Keterangan:

α_i = alfa ke- i

y_i = kelas data ke- i

$K(x_i, x^-)$ = fungsi kernel data negatif.

$K(x_i, x^+)$ = fungsi kernel data positif.

5. Melakukan perhitungan fungsi $f(x)$,

$$f(x) = \sum_{i=1}^m a_i y_i K(x_i, x) + b \quad (11)$$

Keterangan:

b = bias

α_i = alfa ke- i

y_i = kelas data ke- i

$K(x_i, x)$ = fungsi kernel.

2.9. Confusion Matrix

Confusion Matrix merupakan teknik yang digunakan untuk mengevaluasi klasifikasi model untuk memperkirakan objek yang benar atau salah. Sebuah matriks dari prediksi akan dibandingkan dengan kelas asli yang berisi informasi aktual dan prediksi nilai klasifikasi. Setelah sistem berhasil mengklasifikasikan *tweet*, dibutuhkan ukuran untuk menentukan seberapa valid atau tepat klasifikasi telah dibuat oleh sistem. Tabel 1 menunjukkan *confusion matrix* yang digunakan untuk membantu dalam perhitungan sistem evaluasi (Tiara, Sabariah, & Effendy, 2015). Pengujian akurasi dalam

pengujian ini menggunakan *confusion matrix* empat kondisi sebagai berikut: *True Positive (TP)*, *True Negatif (TN)*, *False Positif (FP)*, dan *False Negative (FN)*.

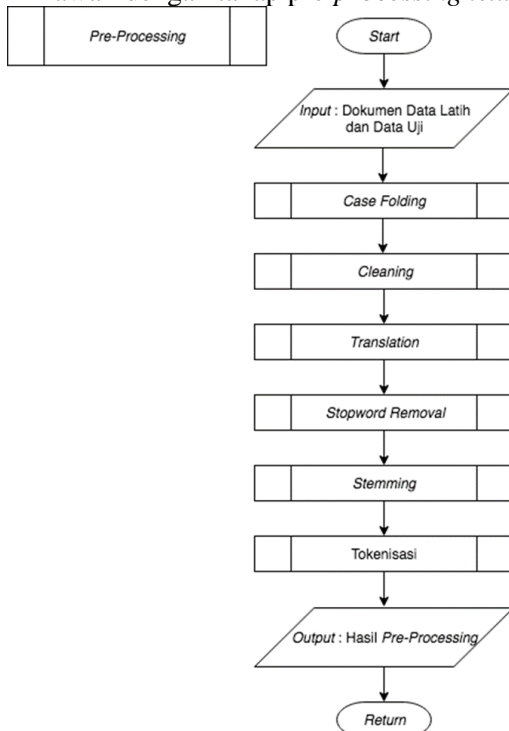
Tabel 1. *Confusion Matrix*

<i>Classification</i>	<i>Predicted Positives</i>	<i>Predicted Negatives</i>
<i>Actual Positive Cases</i>	<i>Number of True Positive Cases (TP)</i>	<i>Number of False Negative Cases (FN)</i>
<i>Actual Negative Cases</i>	<i>Number of False Positive Cases (FP)</i>	<i>Number of True Negative Cases (TN)</i>

3. METODE USULAN

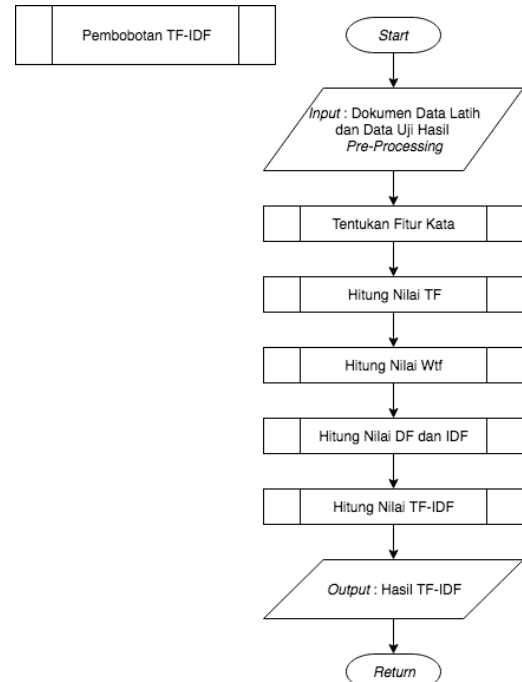
3.1. Alur Proses Sistem

Diawali dengan tahap *pre-processing text*.



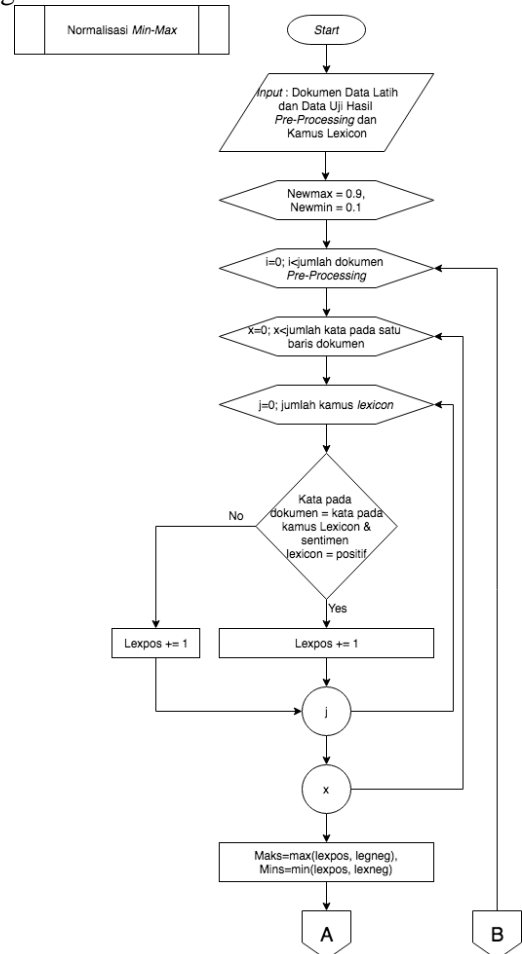
Gambar 1. Alur tahap *pre-processing*

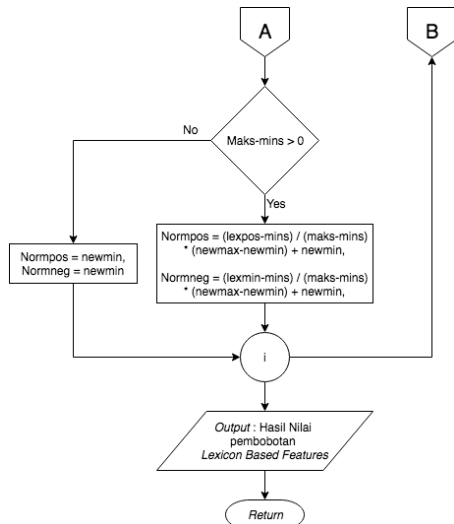
Dilanjutkan dengan perhitungan bobot TF-IDF.



Gambar 2. Alur pembobotan TF-IDF

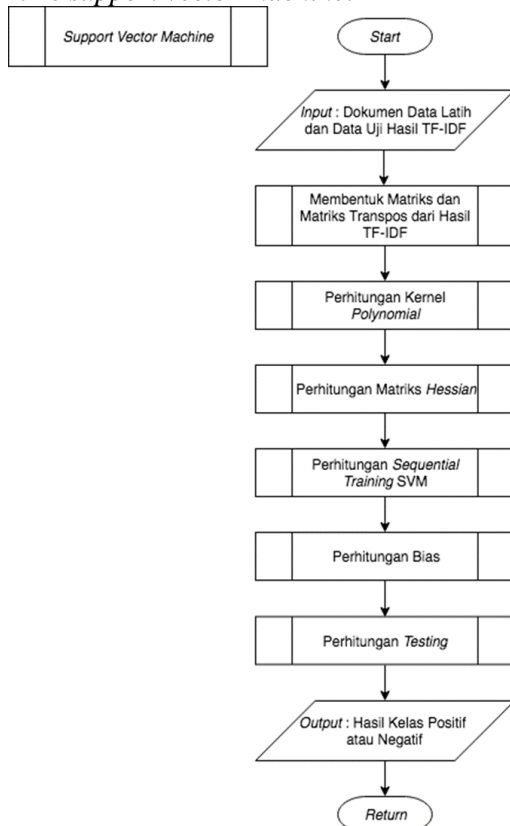
Tahap selanjutnya adalah pembobotan dengan *Lexicon Based Features*.





Gambar 3. Alur tahap *Lexicon Based Features* dengan Normalisasi *Min-Max*

Tahap akhir yaitu klasifikasi dengan algoritme *support vector machine*.

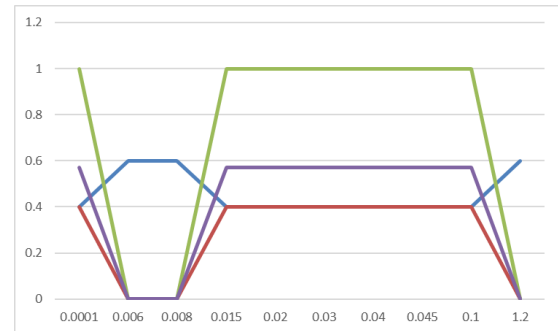


Gambar 3. Alur pembobotan SVM

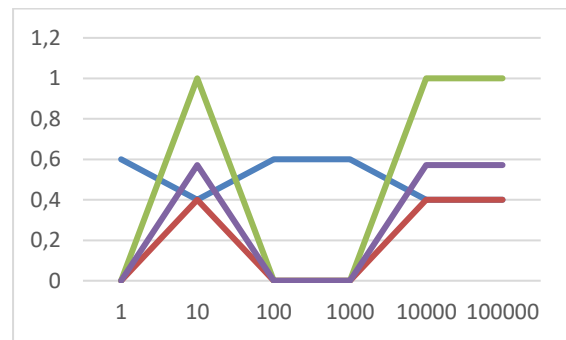
4. PENGUJIAN DAN ANALISIS

Pengujian yang telah dilakukan adalah pengujian terhadap parameter SVM yaitu nilai *learning rate* (γ), nilai C, iterasi maksimum, dan pengaruh implementasi *Lexicon Based Features* dengan digunakan data latih sebanyak 200 data dan data uji sebanyak 50 data. Pada Gambar 4 dan Gambar 5, diketahui warna

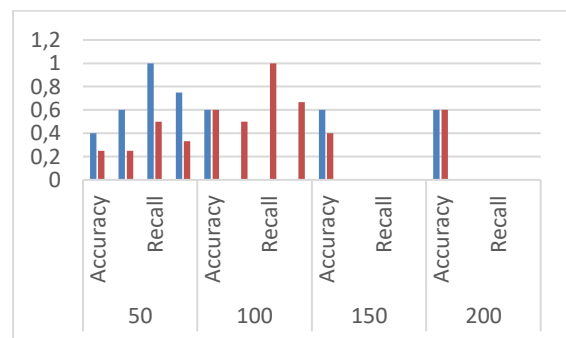
biru menunjukkan akurasi, warna merah menunjukkan presisi, warna hijau menunjukkan *recall*, dan warna ungu menunjukkan *f-measure*. Sedangkan pada Gambar 6, diketahui warna biru menunjukkan implementasi *Lexicon Based Features* dan warna merah menunjukkan tanpa implementasi *Lexicon Based Features*.



Gambar 4. Grafik hasil pengujian nilai *learning rate* ketika iterasi sebanyak 50 kali



Gambar 5. Grafik hasil pengujian nilai C ketika iterasi sebanyak 100 kali



Gambar 6. Grafik hasil pengujian pengaruh implementasi *lexicon based features* ketika iterasi sebanyak 50 kali

Pada Gambar 4, iterasi sebanyak 50 kali, nilai *learning rate* yang terlihat optimal dan stabil ketika digunakan *learning rate* dalam rentang nilai 0,015 sampai dengan 0,1. Dapat dilihat ketika menggunakan nilai 1.2 grafik langsung turun drastis. Pada grafik-grafik hasil pengujian *learning rate* yang lain dengan jumlah iterasi yang bervariasi, ditemukan bahwa nilai

learning rate 0,03 merupakan nilai yang paling optimal dan stabil. Ketika nilai lain sempat turun naik, nilai 0,03 tetap stabil atau bahkan meningkat. Iterasi sebanyak 50 kali terlihat optimal dan stabil karena ketika terjadi penurunan dapat terjadi peningkatan lagi dan menjadi stabil.

Pada Gambar 5, iterasi sebanyak 100 kali, nilai *c* terlihat optimal ketika digunakan nilai *c* sebesar 10 dan 10000. Dapat dilihat terjadi peningkatan drastis pada kedua nilai tersebut dari nilai-nilai *c* sebelumnya, terutama pada nilai *f-measure*-nya. Namun, jika dilihat dari iterasi sebelumnya, nilai 10000 tidak termasuk dalam nilai yang optimal, bahkan pada iterasi sebelumnya ketika nilai *c* adalah 10000 terjadi penurunan drastis. Sehingga dipilihlah *c* = 10, karena di grafik-grafik hasil pengujian nilai *c* di iterasi lainnya, nilai 10 selalu meningkat atau stabil, tidak pernah turun. Pada iterasi sebanyak 100 kali, terjadi peningkatan dan penurunan tetapi kemudian menjadi stabil.

Pada Gambar 6, dapat dilihat perbandingannya secara jelas ketika digunakan *lexicon* dan tidak. Terutama pada iterasi sebanyak 50 kali, terlihat bahwa ketika *lexicon* digunakan akan lebih tinggi akurasi, presisi, dan *recall*-nya dibandingkan ketika tidak menggunakan *lexicon*. Namun, untuk nilai *f-measure* terlihat seimbang. Pada jumlah iterasi lainnya, tidak seluruhnya mendapatkan nilai, banyak yang bernilai 0. Sehingga iterasi paling optimal adalah iterasi sebanyak 50 kali.

Sehingga didapatkan akurasi paling baik sebesar 40%, *precision* sebesar 40%, *recall* sebesar 100%, dan *f-measure* sebesar 57,14%. Tingkat akurasi tersebut didapatkan dengan jumlah iterasi maksimum sebanyak 50 kali dengan diimplementasikannya fitur *lexicon based*. Dalam perhitungan akurasi digunakan parameter-parameter optimal yang telah disebutkan sebelumnya. Hasil klasifikasi dipengaruhi parameter yang optimal dan fitur yang digunakan.

5. KESIMPULAN DAN SARAN

Metode klasifikasi *Support Vector Machine* dengan fitur *Lexicon Based* dapat digunakan dalam menganalisis sentimen opini maskapai penerbangan pada dokumen Twitter dengan optimal. Didapatkan nilai parameter *learning rate* (γ) sebesar 0,03 dan nilai *C* sebesar 10 sebagai nilai parameter paling optimal. Didapatkan tingkat akurasi paling baik sebesar

40%, *precision* sebesar 40%, *recall* sebesar 100%, dan *f-measure* sebesar 57,14%. Tingkat akurasi tersebut didapatkan dengan jumlah iterasi maksimum sebanyak 50 kali dengan diimplementasikannya fitur *lexicon based*. Dalam perhitungan akurasi digunakan parameter-parameter optimal yang telah disebutkan sebelumnya. Hasil klasifikasi dipengaruhi parameter yang optimal dan fitur yang digunakan.

Saran untuk penelitian selanjutnya adalah data dapat diklasifikasikan menjadi tiga kelas, yaitu sentimen positif, negatif, dan netral, dapat diimplementasikan suatu metode tambahan seperti metode optimasi agar dapat lebih mudah dalam mengetahui sebuah teks berdasarkan makna tiap kata di dalam kalimat, dapat menambah dan mengurangi tahap pada pre-processing, misal: menggunakan formalisasi (sebagai contoh, seperti pujangga) karena banyak kata yang tidak baku atau tidak formal. Selain itu, dapat mencoba tanpa melakukan tahap *stopword removal*, yang di mana saat ini sedang diperdebatkan penggunaannya terhadap analisis sentiment, dapat dilakukan inovasi pada sistem, misalnya selain sistem dapat menganalisis sentimen dari sebuah opini, sistem juga dapat merangkum opini dari keseluruhan opini yang ada.

6. DAFTAR PUSTAKA

- Adiwijaya, I., 2006. Text Mining dan Knowledge Discovery. Kolokium bersama komunitas datamining Indonesia & soft-computing Indonesia, [online] pp.1–9. Available at: <http://web.ipb.ac.id/~ir-lab/pdf/tm_text_summarization.pdf>.
- Afuan, L., 2013. Stemming Dokumen Teks Bahasa Indonesia. *Telematika*, 6(2), pp.34–40.
- Twitter Using Multilingual Machine Translated Data. *IEEE Intelligent Systems*, [online] 18(1), pp.12–13. Available at: <<http://ieeexplore.ieee.org/document/1179188/>>.
- Hamdan, H., Bellot, P. and Bechet, F., 2015. Lsislif: Feature extraction and label weighting for sentiment analysis in Twitter. *Proceedings of International Workshop on Semantic Evaluation (SemEval-2015)*, (SemEval), pp.568–
- Kaur, H. and Mangat, V., 2017. A Survey of

- Sentiment Analysis techniques. pp.921–925.
- Tiara, Sabariah, M.K. and Effendy, V., 2015. Sentiment analysis on Twitter using the combination of lexicon-based and support vector machine for assessing the performance of a television program. 2015 3rd International Conference on Information and Communication Technology, ICoICT 2015, pp.386–390.
- Windasari, I.P., Uzzi, F.N. and Satoto, K.I., 2017. Sentiment Analysis on Twitter Posts : An analysis of Positive or Negative Opinion on GoJek. pp.266–269.
- Desai, M. and Mehta, M.A., 2017. Techniques for sentiment analysis of Twitter data: A comprehensive survey. Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016, (March), pp.149–154.
- Rofiqoh, U., Perdana, R.S. and Fauzi, M.A., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, [online] 1(12), pp.1725–1732. Available at: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>>.
- Peng, W., 2011. Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization. s.l.:s.n.
- Vijayakumar, S., 1999. Sequential Support Vector Classi. Proc. International Conference on Soft Computing (SOCO'99). pp 610-619.
- Howard, P.N. and Parks, M.R., 2012. Social Media and Political Change: Capacity, Constraint, and Consequence. Journal of Communication, 62(2), pp.359–362.
- Saif, H., Fernandez, M., He, Y. and Alani, H., 2014. SentiCircles for contextual and conceptual semantic sentiment analysis of Twitter. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8465 LNCS, pp.83–98.