

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/320234928>

Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features

Article · August 2017

CITATION
1

READS
627

3 authors, including:



Rizal Setya Perdana
Brawijaya University
39 PUBLICATIONS 10 CITATIONS

SEE PROFILE



Mochammad Ali Fauzi
Brawijaya University
83 PUBLICATIONS 43 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Artificial Neural Network Applications [View project](#)



Twitter Sentiment Analysis [View project](#)

Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan *Lexicon Based Features*

Umi Rofiqoh¹, Rizal Setya Perdana², M. Ali Fauzi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹umi.roffi@gmail.com, ²rizalespe@ub.ac.id, ³moch.ali.fauzi@ub.ac.id

Abstrak

Analisis sentimen adalah salah satu cabang penelitian dari *Text Mining* yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen. Dokumen teks yang digunakan dalam penelitian berasal Twitter tentang opini masyarakat mengenai penyedia layanan telekomunikasi seluler. Metode yang digunakan adalah Support Vector Machine dengan menggunakan *Lexicon Based Features* sebagai pembaharuan fitur nya selain memakai fitur TF-IDF. Data yang digunakan pada penelitian ini sebanyak 300 data yang dibagi menjadi dua jenis data dengan perbandingan 70% untuk data latih dan 30% untuk data uji. Hasil akurasi sistem yang diperoleh dari analisis sentimen dengan metode Support Vector Machine dan *Lexicon Based Features* sebesar 79% menggunakan nilai *degree* sebesar 2, nilai konstanta *learning rate* 0,0001, serta jumlah iterasi maksimum sebanyak 50 kali. Sedangkan sistem analisis sentimen tanpa menggunakan *Lexicon Based Features* menghasilkan tingkat akurasi sebesar 84% dengan nilai parameter yang sama.

Kata kunci: *analisis sentimen, Twitter, Support Vector Machine, Lexicon Based Features*

Abstract

Sentiment analysis is a part of research from Text Mining which is usefull to classify text documents contained opinion based on sentiment. Text document that is used in research comes from Twitter from people's opinion about cellular telecommunication service provider. The used method is Support Vector Machine with using Lexicon Based Features as its feature renewal instead of using TF-IDF features. The used data in this research is 300 data which divided into two types of data with ratio 70% for training data and 30% for testing data. The result of system accuracy that is obtained from sentiment analysis using Support Vector Machine and Lexicon Based Features method is 79% using degree value 2, constant learning rate value 0.0001, and maximum iteration is 50 times. While sentiment analysis system without using Lexicon Based Features is resulting accuracy at 84% with the same parameter values.

Keywords: *sentiment analysis, Twitter, Support Vector Machine, Lexicon Based Features*

1. PENDAHULUAN

Teknologi telekomunikasi semakin berkembang dari tahun ke tahun, terutama teknologi telepon seluler. Untuk mendukung kelancaran komunikasi dengan orang lain, masyarakat membutuhkan penyedia layanan telekomunikasi seluler yang dapat menunjang jalannya telepon seluler mereka dan sesuai dengan kebutuhan. Perusahaan-perusahaan penyedia layanan telekomunikasi seluler berlomba-lomba untuk meningkatkan inovasi dan mutu pada layanan produk mereka, misalnya seperti biaya bertelepon yang semakin murah, bonus layanan pesan singkat yang

semakin banyak, paket layanan data internet dengan kuota yang banyak dan kecepatan koneksi yang tinggi, serta sinyal yang dapat menjangkau ke semua wilayah. Tercatat tiga penyedia layanan telekomunikasi seluler yang paling populer di Indonesia yaitu XL Axiata, Telkomsel, dan Indosat versi Socialbakers (Socialbakers, 2017). Setiap penyedia layanan telekomunikasi seluler mempunyai keunggulan tersendiri yang mampu memuaskan kebutuhan penggunaanya sekaligus mempunyai kekurangan yang membuat masyarakat mengeluh dan berpindah ke penyedia layanan telekomunikasi seluler lain. Sehingga setiap perusahaan penyedia layanan telekomunikasi seluler pun

perlu memperhatikan apa yang menjadi keunggulan dan kelemahan produk layanannya sendiri atau perusahaan lain agar tidak terjadi penurunan layanan dan berpindahnya konsumen ke penyedia jasa yang lain.

Agar dapat mengetahui bagaimana tingkat kepuasan dari pengguna terhadap suatu penyedia layanan telekomunikasi seluler, maka diperlukan analisis sentimen pada Twitter pengguna yang menyangkut penyedia layanan telekomunikasi seluler tersebut. Analisis sentimen atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen, dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks (Liu, 2012). Analisis sentimen ini dapat mengelompokkan polaritas dari teks dalam kalimat atau dokumen untuk mengetahui apakah opini pada kalimat atau dokumen tersebut apakah termasuk positif atau negatif.

Metode yang digunakan adalah Support Vector Machine. Metode tersebut dipilih karena pada penelitian Vidya, Fanany, dan Budi tentang analisis sentimen pada tingkat reputasi merk penyedia layanan telekomunikasi seluler, juga membuktikan bahwa metode Support Vector Machine memiliki akurasi yang lebih tinggi dibandingkan Naive Bayes dan Decision Tree yaitu mencapai 82,40% (Vidya, Fanany, & Budi, 2015). Untuk fiturnya menggunakan *Lexicon Based Features*. Pada penelitian Siddiqua, Ahsan, dan Chy (2016) telah membuktikan bahwa *Lexicon based features* merupakan fitur pada analisis sentimen pada microblog yang paling penting peranannya setelah diteliti menggunakan seleksi fitur statistik chi-square (χ^2) dan *information gain* (Siddiqua, Ahsan, & Chy, 2016).

Dengan penelitian ini diharapkan agar para perusahaan penyedia layanan telekomunikasi seluler dapat melihat bagaimana tingkat kepuasan masyarakat terhadap produk layanan mereka sehingga mereka akan terpacu untuk meningkatkan layanan dan mutu pada produk layanan mereka.

2. METODE USULAN

Tahapan proses pada sistem analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter dengan metode Support Vector Machine dan *Lexicon Based Features* antara lain praproses teks, pembobotan kata, pembobotan *Lexicon Based Features*, dan

klasifikasi dengan Support Vector Machine. Praproses teks terdiri dari tokenisasi, *data cleaning*, *case folding*, filterisasi, dan *stemming*. Hasil dari praproses teks akan dilakukan pembobotan. Pertama, pembobotan kata yang terdiri dari perhitungan TF, W_{tf} , DF, IDF, dan TF-IDF. Kemudian pembobotan *Lexicon Based Features* dengan cara mengecek kata bersentimen positif atau negatif berdasarkan kamus atau *lexicon* yang ada di *sentiment dictionaries* (data yang berisi kata-kata bersentimen) dan menghitung frekuensi kemunculannya pada dokumen teks. Kemudian frekuensi kata tersebut dinormalisasi dengan metode *Min-max* dan menjadi bobot untuk *Lexicon Based Features*. Setelah fitur pembobotan kata dan *Lexicon Based Features* diperoleh, maka data-data tersebut akan diklasifikasi dengan metode Support Vector Machine. Data latih yang mengandung fitur bobot TF-IDF dan *Lexicon Based Features* akan diolah dalam proses pelatihan hingga terbentuk suatu *hyperplane* pemisah antar kelas data latih. Setelah terbentuk *hyperplane* tersebut, maka data uji dapat dimasukkan ke dalam fungsi $\text{sign}(h(x))$ untuk mendapatkan hasil klasifikasi analisis sentimen berupa sentimen positif atau negatif. Alur dari proses sistem dapat ditunjukkan pada Gambar 1.

Data yang digunakan pada penelitian ini sebanyak 300 data yang dibagi menjadi dua jenis data dengan rasio yaitu 70% untuk data latih 30% untuk data uji. Pembagian rasio data ini berdasarkan pada penelitian yang telah dilakukan oleh Ariadi dan Fithriasari (Ariadi & Fithriasari, 2015).

2.1. Pembobotan Kata

Pembobotan kata bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. Fitur kata yang telah diberi bobot dapat digunakan untuk proses klasifikasi. Tahapan pembobotan kata yaitu:

1. Term Frequency (tf)

Term frequency atau tf merupakan jumlah kemunculan atau frekuensi kata pada suatu dokumen. Sementara W_{tf} adalah jumlah bobot dari tf yang telah dihitung dengan logaritma. Persamaan dari *Term Frequency*:

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

2. Document Frequency (df)

Document Frequency (df) merupakan frekuensi atau jumlah dokumen yang

mengandung suatu kata.

3. Inverse Document Frequency (idf)

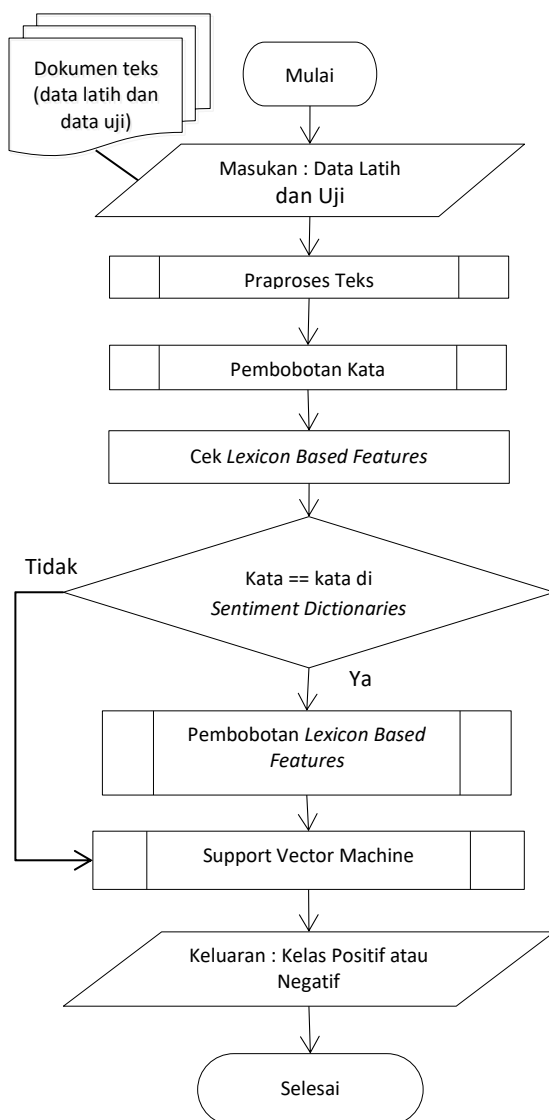
Inverse Document Frequency (idf) adalah bobot kebalikan dari bobot *document frequency*. Kata yang jarang muncul di banyak dokumen mempunyai bobot *Inverse Document Frequency* yang tinggi. Persamaan dari *Inverse Document Frequency* (idf):

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

Keterangan:

N : jumlah dokumen teks.

df_t : jumlah dokumen yang mengandung suatu kata t .



Gambar 1. Diagram Alur Sistem

4. Term Frequency-Inverse Document Frequency (tf-idf)

Pembobotan ini adalah hasil perkalian dari pembobotan *term frequency* dan *inverse document frequency* dari suatu *term*.

Persamaannya:

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

Keterangan :

$w_{tf_{t,d}}$: *Term Frequency*.

idf_t : *Inverse Document Frequency*.

2.2. Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Support Vector Machine masuk kelas *supervised learning*, dimana dalam implementasinya perlu adanya tahap pelatihan menggunakan *sequential training SVM* dan disusul tahap pengujian (Santosa, 2015).

Konsep klasifikasi dengan Support Vector Machine adalah mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data. Support Vector Machine mampu bekerja pada dataset yang berdimensi tinggi dengan menggunakan kernel trik. Support Vector Machine hanya menggunakan beberapa titik data terpilih yang berkontribusi (*support vector*) untuk membentuk model yang akan digunakan dalam proses klasifikasi. Ilustrasi metode Support Vector Machine ditunjukkan pada Gambar 2.

Persamaan Support Vector Machine:

$$f(x) = w \cdot x + b \quad (4)$$

atau

$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (5)$$

Keterangan :

w : parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik *support vector*)

x : titik data masukan Support Vector Machine

a_i : nilai bobot setiap titik data

$K(x, x_i)$: fungsi *kernel*

b : parameter *hyperplane* yang dicari (nilai *bias*)

Untuk penelitian ini menggunakan kernel *polynomial*. Persamaannya:

$$K(x, y) = (x \cdot y + c)^d \quad (6)$$

Keterangan:

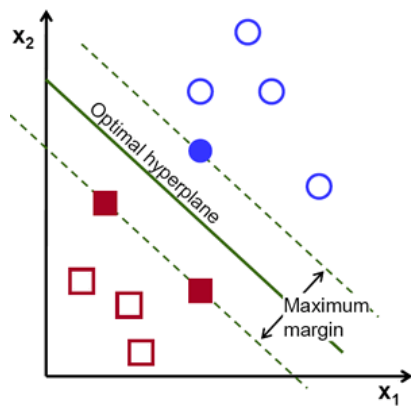
$K(x, y)$: Nilai kernel dari data x dan data y

x : fitur data 1

y : fitur data 2

c : nilai konstanta

d : nilai *degree*



Gambar 2. Ilustrasi Metode Support Vector Machine

2.3. Pembobotan *Lexicon Based Features*

Lexicon Based Features merupakan fitur kata yang memiliki sentimen positif atau negatif berdasarkan kamus atau *lexicon*. *Lexicon* adalah kumpulan kata sentimen yang telah diketahui dan terhimpun (Desai & Mehta, 2016). Untuk proses pembobotan pada fitur ini, dibutuhkan kamus atau *lexicon* yang berisi kata-kata yang mengandung sentimen yang disebut dengan *sentiment dictionaries* (Buntoro, Adji, & Purnamasari, 2014, Cho, et al., 2014). Sentimen yang digunakan berupa positif dan negatif. Terdapat dua macam fitur yang digunakan pada penelitian ini yaitu fitur jumlah kata positif dan fitur jumlah kata negatif dalam dokumen. Agar bobot *Lexicon Based Features* seimbang dengan bobot tf-idf, maka fitur jumlah kata positif dan kata negatif perlu dinormalisasikan dengan metode Min-max.

2.4. Normalisasi Min-max

Normalisasi Min-max merupakan proses transformasi yang mana merubah nilai suatu variabel diskalakan menjadi berukuran lebih kecil dengan jangkauan antara -1 sampai 1 atau 0,0 sampai 0,1 (Junaedi, 2011). Normalisasi digunakan untuk mengurangi kesalahan pada proses data mining (Wirawan, 2015).

Metode normalisasi ini merupakan metode yang paling sederhana dengan melakukan transformasi linier terhadap data asli dan memiliki kelebihan yaitu terdapat keseimbangan nilai perbandingan antara nilai data sebelum dinormalisasi dengan nilai data yang telah dinormalisasi. Persamaan normalisasi Min-max:

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (7)$$

Keterangan :

v'_i : Nilai data baru hasil normalisasi min-max.

v_i : Nilai data yang akan di normalisasi.

\min_A : Nilai minimum data.

\max_A : Nilai maksimum data.

new_min_A : Nilai minimum yang diharapkan dari proses normalisasi.

new_max_A : Nilai maksimum yang diharapkan dari proses normalisasi.

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Pengaruh Parameter Support Vector Machine

Pengujian pengaruh parameter Support Vector Machine dilakukan untuk mengetahui nilai-nilai parameter SVM yang optimal untuk proses analisis sentimen. Pada pengujian ini, terdapat tiga parameter yang diuji yaitu nilai *degree* untuk perhitungan kernel serta nilai konstanta *learning rate* dan jumlah iterasi maksimum untuk perhitungan *sequential training SVM*.

Pada pengujian pengaruh nilai *degree*, nilai konstanta *learning rate* yang digunakan adalah 0,0001 dan jumlah iterasi maksimum adalah 50. Pengujian ini dilakukan sebanyak lima kali dengan nilai *degree* yang bervariasi. Hasil dari pengujian pengaruh nilai *degree* ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pengujian Pengaruh Nilai *degree*

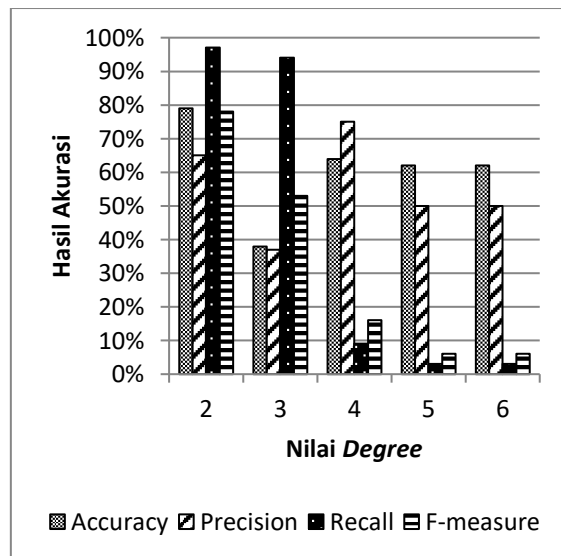
	Kernel Polynomial (Degree)				
	2	3	4	5	6
Accuracy	79%	38%	64%	62%	62%
Precision	65%	37%	75%	50%	50%
Recall	97%	94%	9%	3%	3%
F-measure	78%	53%	16%	6%	6%

Dari hasil pengujian tersebut dapat diketahui pada grafik pengujian pengaruh nilai *degree* yang ditunjukkan pada Gambar 3.

Berdasarkan pada Gambar 3 dari hasil pengujian pengaruh nilai *degree* menunjukkan bahwa tingkat akurasi tertinggi diperoleh ketika nilai *degree* adalah 2. Ketika nilai *degree* semakin diperbesar, maka tingkat akurasi akan berubah dan tidak stabil lalu semakin menurun.

Sementara untuk pengujian nilai konstanta *learning rate* dan jumlah iterasi maksimum, nilai *degree* yang digunakan adalah 2.

Pengujian nilai konstanta *learning rate* sebanyak lima kali dan jumlah iterasi maksimum sebanyak empat kali. Hasil dari pengujian pengaruh nilai *learning rate* dan jumlah iterasi maksimum ditunjukkan pada Tabel 2.

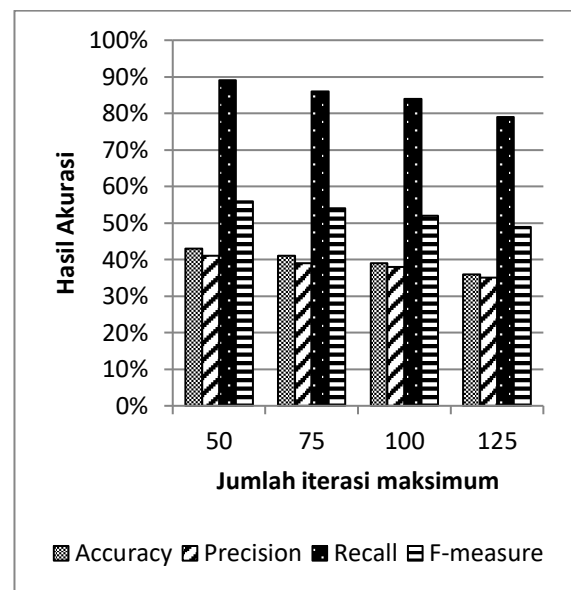


Gambar 3. Grafik Pengujian Pengaruh Nilai *degree*

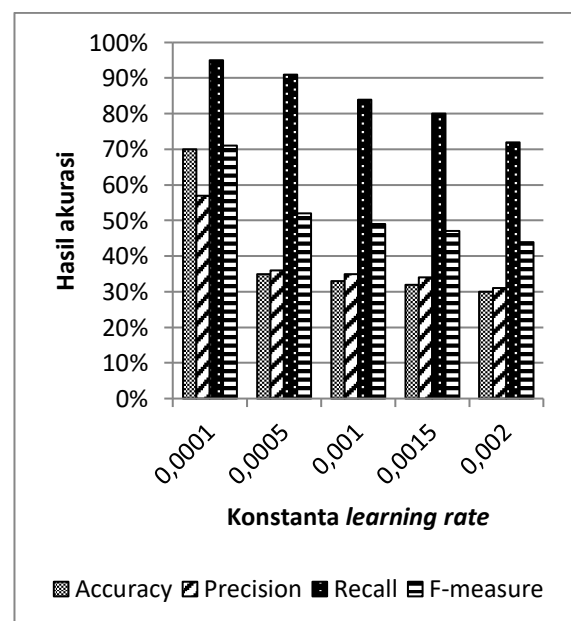
Tabel 2. Hasil Pengujian Nilai Konstanta *Learning Rate* dan Jumlah Iterasi Maksimum

		Konstanta γ					Rata-rata
		0,0001	0,0005	0,001	0,0015	0,002	
50	Accuracy	79%	36%	34%	32%	34%	43%
	Precision	65%	36%	36%	34%	35%	41%
	Recall	97%	91%	91%	82%	85%	89%
	F-measure	78%	52%	51%	48%	50%	56%
75	Accuracy	74%	34%	32%	33%	32%	41%
	Precision	60%	36%	34%	34%	33%	39%
	Recall	94%	91%	82%	82%	79%	86%
	F-measure	74%	51%	48%	48%	47%	54%
100	Accuracy	68%	34%	34%	32%	29%	39%
	Precision	54%	36%	35%	33%	31%	38%
	Recall	94%	91%	85%	79%	71%	84%
	F-measure	69%	51%	50%	47%	43%	52%
125	Accuracy	60%	35%	32%	31%	23%	36%
	Precision	48%	36%	33%	33%	25%	35%
	Recall	94%	91%	79%	76%	53%	79%
	F-measure	64%	52%	47%	46%	34%	49%
Rata-rata	Accuracy	70%	35%	33%	32%	30%	
	Precision	57%	36%	35%	34%	31%	
	Recall	95%	91%	84%	80%	72%	
	F-measure	71%	52%	49%	47%	44%	

Dari hasil pengujian tersebut dapat dilihat pada grafik pengujian pengaruh jumlah iterasi maksimum dan nilai konstanta *learning rate* yang ditunjukkan pada Gambar 4 dan Gambar 5.



Gambar 4. Grafik Pengujian Pengaruh Jumlah Iterasi Maksimum



Gambar 5. Grafik Pengujian Pengaruh Nilai Konstanta *Learning Rate*

Berdasarkan pada Gambar 4 dan 5 dari hasil pengujian nilai konstanta *learning rate* dan jumlah iterasi maksimum menunjukkan bahwa nilai konstanta *learning rate* yang optimal adalah 0,0001. Sedangkan jumlah iterasi maksimum optimal yang digunakan adalah 50. Apabila nilai konstanta *learning rate* dan jumlah iterasi maksimum diperbesar maka tingkat akurasi akan semakin menurun. Hal ini dikarenakan apabila nilai konstanta *learning rate* berubah atau jumlah iterasi maksimum bertambah, maka *support vector* yang

ditentukan juga akan ikut berubah dan berpengaruh pada *hyperplane* garis pembatas antar kelas data latih.

3.2. Pengujian Pengaruh *Lexicon Based Features* dan Akurasi Sistem

Pada pengujian pengaruh *Lexicon Based Features* terhadap analisis sentimen menggunakan Support Vector Machine ini akan dilakukan perbandingan hasil akurasi sistem yang menggunakan *Lexicon Based Features* dan tanpa menggunakan *Lexicon Based Features*. Nilai-nilai parameter Support Vector Machine yang digunakan dalam pengujian ini adalah nilai parameter Support Vector Machine optimal berdasarkan hasil pengujian sebelumnya yaitu nilai *degree* = 2, nilai konstanta *learning rate* = 0,0001, dan jumlah iterasi maksimum = 50 kali. Sementara pada pengujian akurasi sistem akan ditunjukkan berapa tingkat akurasi yang dapat dicapai oleh sistem analisis sentimen menggunakan metode Support Vector Machine dengan menggunakan *Lexicon Based Features* ataupun tanpa menggunakan *Lexicon Based Features*. Hasil pengujian pengaruh *Lexicon Based Features* dan akurasi sistem ditunjukkan pada Tabel 3.

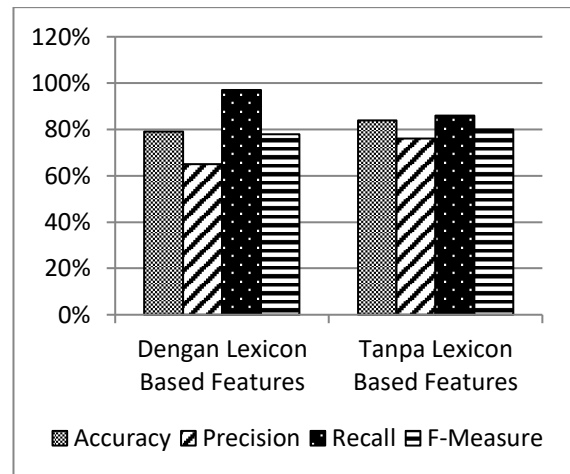
Tabel 3. Hasil Pengujian Pengaruh *Lexicon Based Features* dan Akurasi Sistem

		Dengan <i>Lexicon Based Features</i>		Tanpa <i>Lexicon Based Features</i>	
		Hasil Sebenarnya		Hasil Sebenarnya	
		Positif	Negatif	Positif	Negatif
Hasil Prediksi	Positif	33	18	26	7
	Negatif	1	38	8	48
<i>Accuracy</i>		79%		84%	
<i>Precision</i>		65%		76%	
<i>Recall</i>		97%		86%	
<i>F-Measure</i>		78%		80%	

Dari hasil pengujian tersebut dapat dilihat pada grafik pengujian pengaruh *Lexicon Based Features* dan akurasi sistem yang ditunjukkan pada Gambar 6.

Berdasarkan pada Gambar 6 dari hasil pengujian pengaruh *Lexicon Based Features* terhadap sistem analisis sentimen menunjukkan bahwa tingkat akurasi sistem analisis sentimen dengan menggunakan *Lexicon Based features* lebih rendah daripada sistem analisis sentimen yang tanpa menggunakan *Lexicon Based Features*. Hal ini dikarenakan terdapat kata-kata

bersentimen positif di dalam data uji yang seharusnya bersentimen negatif dan sebaliknya.



Gambar 6. Grafik Pengujian Pengaruh Nilai Konstanta *Learning Rate*

Untuk hasil akurasi sistem dengan menggunakan *Lexicon Based Features* menghasilkan *accuracy* sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%. Hal ini menunjukkan bahwa sistem yang menggunakan *Lexicon Based Features* baik digunakan dalam analisis sentimen, walaupun tingkat akurasinya lebih rendah daripada sistem analisis sentimen yang tidak menggunakan *Lexicon Based Features* yaitu nilai *accuracy* sebesar 86%, *precision* sebesar 84%, *recall* sebesar 76%, dan *f-measure* sebesar 80%. Namun perbandingan tingkat akurasi tersebut tidak menunjukkan perbedaan yang terlalu jauh atau signifikan. Untuk nilai *recall* sistem analisis sentimen dengan menggunakan *Lexicon Based Features* lebih besar yaitu 97% dibandingkan dengan nilai *recall* sistem analisis sentimen yang tanpa menggunakan *Lexicon Based Features* sebesar 76%. Hal ini dikarenakan sistem analisis sentimen dengan menggunakan *Lexicon Based Features* lebih dapat mengenali sentimen data uji melalui pengecekan kata bersentimen pada saat pembobotan *Lexicon Based Features* apabila digabungkan dengan sistem analisis sentimen tanpa menggunakan *Lexicon Based Features*.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan sebelumnya, maka dapat diambil kesimpulan mengenai analisis sentimen terhadap tingkat kepuasan pengguna penyedia

layanan telekomunikasi seluler pada Twitter dengan metode Support Vector Machine dan *Lexicon Based Features* yaitu sebagai berikut.

1. Metode Support Vector Machine dapat diimplementasikan pada analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler dengan hasil klasifikasi berupa sentimen positif atau negatif. Data yang digunakan untuk penelitian ini diambil dari Twitter berupa data teks yang merupakan *tweet-tweet* dari pengguna tiga penyedia layanan telekomunikasi seluler paling populer menurut Socialbakers yaitu Indosat, Telkomsel, dan XL Axiata. Data tersebut berjumlah 300 yang dibagi menjadi dua jenis dengan perbandingan 70%:30% yaitu data latih sebanyak 210 dan data uji sebanyak 90. Data latih terdiri dari 35 data teks bersentimen positif dan 35 dokumen teks bersentimen negatif dari masing-masing jenis penyedia layanan telekomunikasi seluler. Data latih dan uji diproses dalam analisis sentimen melalui tahapan praproses, pembobotan kata berupa tf-idf, pembobotan *Lexicon Based Features* berupa perhitungan frekuensi kemunculan kata yang mengandung sentimen positif atau negatif pada tiap dokumen kemudian dinormalisasi dengan metode Min-max, dan klasifikasi Support Vector Machine untuk mendapatkan hasil analisis sentimen dokumen uji berupa sentimen positif atau negatif.
2. Untuk penelitian analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter dengan metode Support Vector Machine dan *Lexicon Based Features*, didapatkan nilai-nilai parameter Support Vector Machine yang optimal yaitu nilai *degree* untuk perhitungan kernel polynomial adalah 2 serta nilai konstanta learning rate dan jumlah iterasi maksimum untuk perhitungan *sequential training SVM* adalah 0,0001 dan 50. Apabila nilai *degree* diperbesar maka hasil akurasi akan mengalami kemunduran. Sementara nilai konstanta *learning rate* dan jumlah iterasi maksimum semakin diperbesar akan merubah *support vector* yang ditentukan sebagai titik data yang paling dekat dengan garis fungsi pemisah antar kelas data dan menyebabkan tingkat akurasi menurun.
3. Penambahan *Lexicon Based Features*

memiliki pengaruh untuk analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter, namun tingkat akurasi lebih rendah dibandingkan dengan analisis sentimen yang tidak menggunakan *Lexicon Based Features*. Hal ini dikarenakan terdapat kalimat sindiran atau muncul kata-kata bersentimen positif pada data uji yang seharusnya bersentimen negatif dan sebaliknya.

4. Pada penelitian analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada Twitter dengan metode Support Vector Machine dan *Lexicon Based Features* menghasilkan nilai *accuracy* sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%. Nilai *recall* yang sangat besar tersebut dipengaruhi oleh banyaknya data uji yang sebenarnya positif dan terdeteksi sebagai positif oleh sistem. Hal ini membuktikan bahwa analisis sentimen dengan metode Support Vector Machine dan *Lexicon Based Features* baik digunakan untuk klasifikasi analisis sentimen.

5. DAFTAR PUSTAKA

- Ariadi, D., & Fithriasari, K., 2015. Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. JURNAL SAINS DAN SENI ITS Vol. 4, No.2, [e-journal] 248(6). Tersedia melalui: Digital Library Institut Teknologi Sepuluh Nopember <<http://digilib.its.ac.id>> [Diakses 22 Juli 2017]
- Buntoro, G. A., Adji, T. B., & Purnamasari, A. E., 2014. Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. CITEE 2014, [e-journal]. Tersedia melalui: ResearchGate <<https://www.researchgate.net/publication/270794878>> [Diakses 2 Maret 2017]
- Cho, H., Kim, S., Lee, J., & Lee, J., 2014. Data-driven Integration of Multiple Sentiment Dictionaries for Lexicon-Based Sentiment Classification of Product Reviews. Knowledge-Based

- Systems 71 (2014) 61–71, [e-journal]. Tersedia melalui: Sciencedirect <<http://www.sciencedirect.com/>> [Diakses 2 Maret 2017]
- Desai, M. & Mehta, M.A., 2016. Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey. International Conference on Computing, Communication and Automation, [e-journal]. Tersedia melalui: IEEE <<http://iee.org>> [Diakses 28 Februari 2017]
- Junaedi, H., Budianto, H., Maryati, I., & Melani, Y., 2011. Data Transformation Pada Data Mining. Prosiding Konferensi Nasional “Inovasi dalam Desain dan Teknologi” - IDEaTech 2011, [e-journal] 93(7). Tersedia melalui: Ideatech <<http://ideatech.stts.edu/>> [Diakses 25 Maret 2017]
- Liu, B., 2012. Sentiment Analysis and Opinion Mining. [e-book]. Chicago: Morgan & Claypool Publisher. Tersedia di: <<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>> [Diakses 2 Maret 2017]
- Santosa, B., 2015. Tutorial Support Vector Machine. [e-book] Surabaya: Institut Teknologi Sepuluh Nopember. Tersedia di: Google Cendekia <<https://scholar.google.com/>> [Diakses 28 Februari 2017]
- Siddiqua, U.A., Ahsan, T., & Chy, A.N., 2016. Combining a Rule-based Classifier with Ensemble of Feature Sets and Machine Learning Techniques for Sentiment Analysis on Microblog. 19th International Conference on Computer and Information Technology, [e-journal] 304(6). Tersedia melalui: IEEE <<http://iee.org>> [Diakses 25 Februari 2017]
- Socialbakers, 2017. Twitter statistics in Indonesia – Telecom. [online] Socialbakers. Tersedia di : <<https://www.socialbakers.com/statistics/twitter/profiles/indonesia/brands/telecom/>> [Diakses 7 Maret 2017]
- Vidya, N.A., Fanany, M.I., & Budi, I., 2015. Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. The Third Information Systems International Conference, [e-journal] 519(8). Tersedia melalui: ScienceDirect <<http://www.sciencedirect.com>> [Diakses 22 Februari 2017]
- Wirawan, I. N. T., & Eksistyanto, I., 2015. Penerapan Naive Bayes pada Intrusion Detection System dengan Diskritisasi Variabel. JUTI: Jurnal Ilmiah Teknologi Informasi - Volume 13 Nomor 2 Juli 2015, [e-journal] 189(8). Tersedia melalui: JUTI: Jurnal Ilmiah Teknologi Informasi <<http://juti.if.its.ac.id/>> [Diakses 30 Maret 2017]