

Analisis Sentimen Pengguna *Twitter* Terhadap Layanan Internet Provider Menggunakan Algoritma *Support Vector Machine*

Sentiment Analysis of Twitter Users on Internet Service Providers Using Support Vector Machine Algorithm

Fadhilah Dwi Ananda, Yoga Pristyanto

Universitas AMIKOM Yogyakarta, Indonesia

Article Info

Article history:

Received, 25 Maret 2021
Revised, 14 April 2021
Accepted, 5 Mei 2021

Kata Kunci:

Sentiment Analysis
Twitter Data
Classification
Support Vector Machine
Internet Provider

ABSTRAK

Media sosial saat ini merupakan media komunikasi yang sering digunakan oleh kalangan masyarakat Indonesia dalam menyampaikan sebuah opini. Salah satu media yang sering digunakan masyarakat adalah twitter. Twitter merupakan media sosial yang memberikan banyak informasi melalui tweet, dari informasi yang ditulis tersebut terdapat data yang dapat diolah. Penelitian ini menggunakan teknik text mining dengan menerapkan algoritma Support Vector Machine dipergunakan untuk klasifikasi sentimen pengguna twitter terhadap layanan internet Biznet. Kernel yang digunakan adalah kernel Linear dan kernel RBF. Pengujian dilakukan dengan 3 skenario, pada skenario 1 menggunakan 800 data, skenario 2 menggunakan 900 data dan skenario 3 menggunakan 1000 data, untuk pembagiannya yaitu 90% data *training* dan 10% data *testing* dari masing-masing skenario. Berdasarkan hasil pengujian yang dilakukan menggunakan kernel linear dan kernel RBF dapat diambil kesimpulan sebagai berikut. Algoritma SVM menggunakan dengan kernel linear maupun kernel RBF memiliki hasil kinerja evaluasi baik dari sisi akurasi, presisi dan recall yang relatif sama. Sehingga dapat dikatakan bahwa algoritma SVM baik dengan kernel RBF maupun Linear sama sama dapat digunakan dengan baik dalam menentukan sentimen pengguna internet Biznet. Selain itu dengan 3 skenario pengujian dengan jumlah data yang berbeda algoritma SVM baik dengan kernel RBF maupun Linear sama sama konsisten kinerjanya.

ABSTRACT

Social media is currently a communication medium that is often used by Indonesians in conveying an opinion. One of the media that is often used by the public is twitter. Twitter is a social media that provides a lot of information via tweets, from the written information, there is data that can be processed. This study uses text mining techniques by applying the Support Vector Machine algorithm to classify Twitter users' sentiments on Biznet internet services. The kernels used are the Linear kernel and the RBF kernel. The test is carried out with 3 scenarios, in scenario 1 using 800 data, scenario 2 using 900 data, and scenario 3 using 1000 data, for the division that is 90% training data and 10% testing data from each scenario. Based on the results of tests carried out using the linear kernel and the RBF kernel, the following conclusions can be drawn. The SVM algorithm using linear kernels and RBF kernels has relatively the same performance evaluation results in terms of accuracy, precision, and recall. So it can be said that the SVM algorithm with both the RBF and Linear kernels can be used properly in determining the sentiment of Biznet internet users. In addition, with 3 test scenarios with different amounts of data, the SVM algorithm both with the RBF and Linear kernels has the same consistent performance.

Keywords:

Sentiment Analysis
Twitter Data
Classification
Support Vector Machine
Internet Provider

This is an open access article under the [CC BY-SA](#) license.



Penulis Korespondensi:

Yoga Pristyanto,
Program Studi Sistem Informasi,
Universitas AMIKOM Yogyakarta,
Email: yoga.pristyanto@amikom.ac.id

1. PENDAHULUAN

Pada saat ini penggunaan internet di Indonesia berkembang secara luar biasa cepatnya dan internet memegang peran penting di dalam aspek kehidupan manusia. Berdasarkan data dari Asosiasi Penyelenggara Jasa Internet Indonesia pada tahun 2018 pengguna internet mencapai angka 171,17 juta. Penggunaan internet yang menempati urutan pertama paling sering digunakan adalah layanan untuk berkomunikasi lewat pesan yang mencapai 24,7 % dan yang menempatkan urutan kedua yaitu sosial media yang mencapai 18,9 %, media sosial yang sering digunakan yaitu Facebook, Twitter, Instagram dan lain-lain [1]. Beberapa tahun terakhir telah terjadi gelombang kepentingan dalam metode komputasi yang memiliki pengaruh terhadap opinion mining, untuk deteksi subjektivitas, analisis sentimen dan emosi [2]. Sosial media seperti Twitter merupakan media komunikasi yang populer di kalangan pengguna internet di Indonesia. Menurut data yang telah dirilis pada tahun 2012 Indonesia menjadi negara peringkat kelima pengguna Twitter terbesar di dunia [3]. Twitter sering digunakan untuk mengungkapkan emosi mengenai suatu hal, baik dalam hal memuji ataupun mencela dengan bentuk emosi. Emosi yang timbul itu sendiri dapat dikategorikan menjadi lima emosi dasar yaitu senang, cinta, sedih, takut dan marah. Emosi senang dan cinta termasuk ke dalam emosi positif. Emosi sedih, takut dan marah termasuk ke dalam emosi negatif [4]. Tweet yang dipublikasikan pengguna melalui Twitter mengandung sebuah opini mengenai sebuah objek, objek tersebut dapat berupa kejadian disekitar masyarakat seperti suatu produk atau layanan. Hal ini membuat perusahaan menggunakan Twitter sebagai media untuk menyebarkan informasi [5]. Studi komputasional dari opini-opini pengguna internet, sentimen dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks disebut dengan *Sentiment Analysis* atau *Opinion Mining* [6].

Analisis sentimen atau opinion mining merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap dan emosi seseorang terhadap sebuah produk, layanan, organisasi, individu, masalah, peristiwa atau topik. *Opinion mining* digunakan untuk melihat pendapat terhadap sebuah produk dan jasa, peristiwa sosial dan politik serta kegiatan tertentu lainnya. Analisis sentimen atau *opinion mining* berfokus untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah atau objek oleh seseorang, apakah beropini positif atau negatif [7]. Pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian untuk analisis sentimen berkembang pesat. Salah satu dari kegunaannya adalah untuk membantu mengetahui pendapat dari masyarakat atau seseorang terhadap suatu produk atau jasa dengan menggunakannya sebagai alat bantu untuk melihat pendapat konsumen terhadap produk atau jasa tersebut. Di Amerika kurang lebih 20-30 perusahaan yang berfokus pada layanan analisis sentimen [8]. Analisis sentimen sangat bermanfaat dalam dunia usaha seperti melakukan analisis terhadap suatu produk atau jasa untuk melihat respon konsumen terhadap produk atau jasa tersebut, sehingga dapat membuat strategi-strategi baru pada tahapan-tahapan berikutnya.

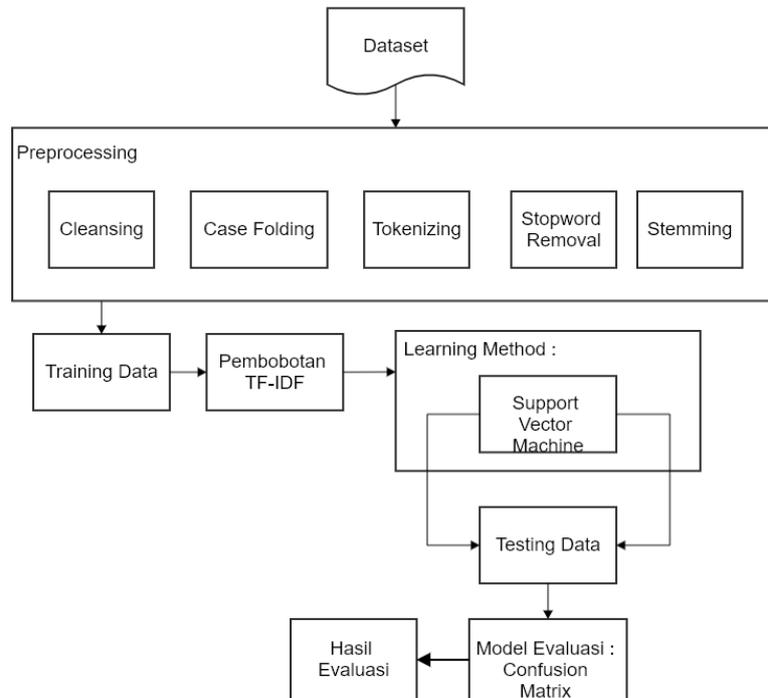
Text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi teks pada analisis sentimen. Klasifikasi teks dapat didefinisikan sebagai proses untuk menentukan suatu dokumen teks ke dalam suatu kelas tertentu. Ada beberapa algoritma yang dapat digunakan diantaranya *Support Vector Machine (SVM)*, *Naïve Bayes Classifier*, *k-Nearest Neighbor (KNN)*, *Artificial Neural Network (ANN)* dan *Decision Tree* [9]. Dalam [10], ditunjukkan bahwa algoritma *Support Vector Machine* memiliki ketahanan dan kemampuan generalisasi yang lebih tinggi serta akurasi klasifikasi yang lebih stabil dibandingkan dengan algoritma yang lain. Penelitian yang dilakukan [11] dengan membandingkan tiga teknik *machine learning* untuk klasifikasi sentiment. Teknik tersebut adalah *Naïve Bayes Classifier (NBC)*, *Maximum Entropy* dan *Support Vector Machine (SVM)*. Hasil yang didapat dari penelitian tersebut *Support Vector Machine (SVM)* memberikan akurasi tertinggi sebesar 82,7%. Penelitian tentang analisis sentimen *review* produk menggunakan algoritma *Support Vector Machine (SVM)*. Hasil implementasi klasifikasi *Support Vector Machine (SVM)* dapat memperoleh nilai akurasi yang lebih tinggi jika dibandingkan dengan model klasifikasi lainnya. Hal ini dikarenakan adanya pengaruh pada data yang diolah yaitu data *review*. Jika semakin banyak data tersebut, maka semakin berpengaruh pada akurasi yang diperoleh [12]. Penelitian yang menggabungkan dua metode juga pernah dilakukan oleh [13]. Penelitian yang dilakukan adalah klasifikasi lagu daerah Indonesia berdasarkan lirik menggunakan metode TF-IDF dan *Naïve Bayes*. Klasifikasi ini diharapkan menjadi salah satu cara untuk mengenal dan memetakan lagu-lagu daerah yang ada di Indonesia agar bangsa Indonesia bisa mengenalnya budayanya sendiri. Jumlah data yang digunakan dalam penelitian ini adalah 90 lagu dari berbagai daerah. Lagu daerah akan diklasifikasikan menurut wilayah. Penelitian ini akan menguji metode ekstraksi ciri *term frequency-inverse document* frekuensi (TF-IDF) dan *Naïve Bayes* sebagai metode klasifikasinya. Penelitian ini membuktikan ekstraksi TF-IDF Metode dan klasifikasi dengan *Naïve Bayes* dapat digunakan untuk mengelompokkan lirik lagu daerah berdasarkan daerahnya lagu dengan akurasi 73,4% di set Indonesia Barat dan Indonesia Timur [13]. Penelitian yang menggunakan algoritma SVM dengan membandingkan hasil dari dua kernel yaitu *kernel Linear* dan *kernel RBF (Radial Basis Function)* sudah pernah dilakukan oleh [14]. Penelitian yang dilakukan dengan memodifikasi algoritma SVM untuk klasifikasi tujuh kelas diantaranya positif film, netral film, negatif film, positif bukan film, netral bukan film, negatif bukan film, dan spam / diluar topik. Komentar dengan hasil klasifikasi positif film dan negatif film yang digunakan dalam menentukan rating sentimen. Jumlah *like* pada komentar juga ikut menentukan *rating* sentimen. Klasifikasi komentar menggunakan STRUCT-SVM. Hasil dari STRUCT-SVM menunjukkan akurasi mencapai 70% untuk *linear kernel* dan 71% untuk *RBF kernel*.

Mayoritas penelitian yang dilakukan menggunakan algoritma *naïve bayes* dan SVM, namun pada umumnya SVM lebih unggul dibandingkan *naïve bayes*. Oleh karena itu penelitian ini akan dilakukan dengan menerapkan metode *Support Vector Machine* untuk klasifikasi sentimen pengguna Twitter terhadap layanan internet Biznet dengan skenario membandingkan antar kernel pada SVM. Hal ini dilakukan untuk menguji kernel manakah yang dapat menghasilkan kinerja terbaik pada algoritme

SVM khususnya dalam klasifikasi sentiment tweet pada layanan internet Biznet. Selain itu juga dilakukan skenario pengujian dengan jumlah komposisi data tertentu untuk menguji apakah kinerja SVM konsisten terhadap perubahan jumlah data.

2. METODE PENELITIAN

Sistem klasifikasi yang dilakukan untuk analisis sentimen ini memiliki suatu rancangan bagaimana alur sistem ini akan berjalan. Gambaran umum sistem yang akan di buat pada Gambar 1.

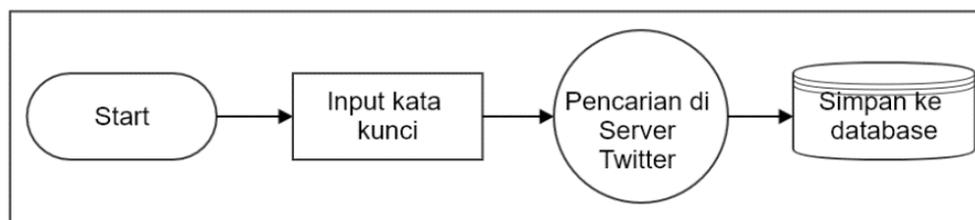


Gambar 1. Tahapan Penelitian

2.1 Akuisisi Dataset

Data yang digunakan dalam penelitian ini adalah data tweet hasil dari penarikan pada server twitter, data tweet tersebut didapat dengan memanfaatkan fitur API (Application Interface) yang telah disediakan oleh Twitter. API digunakan untuk mengambil data tweet dari server tweet kemudian data tersebut dikumpulkan dalam sebuah file dengan format csv.

Pada proses pengumpulan data tweet, peneliti menggunakan kata kunci dan hastag dengan nama perusahaan Internet Service Provider (ISP) yaitu biznet dan #biznet. Kemudian proses pengambilan data tweet tersebut diambil dari pengguna twitter yang memiliki tweet yang sesuai dengan kata kunci menggunakan API twitter. Kemudian dari server twitter akan mengambil data tweet yang sesuai kata kunci, selanjutnya data tweet yang sudah didapatkan dari server twitter di simpan kedalam file yang berformat csv. Gambaran proses pengumpulan data pada Gambar 2.



Gambar 2. Proses Pengambilan Data

Pada penelitian ini Total data *tweet* yang diambil sebanyak 1000 data *tweet*. *Tweet* yang telah di ambil kemudian di simpan untuk di lakukan pelabelan dengan menjadikan dua kelas yaitu positif dan negatif, seperti yang ditunjukkan pada tabel 1.

Tabel 1. Proses Pemberian Label

Tweet	Opini
Emang biznet lebih aman dan nyaman. Internet kenceng dan stabil, harga jg murah banget. :D	Positif
Biznet home lama-lama gue kepret ya. Sebulan 3 kali mati tu internet dan bisa berhari-hari.	Negatif
begini lagi penting mau meeting online karena perbedaan waktu Biznet mati putus Sulit mengandalkan kamu	Negatif

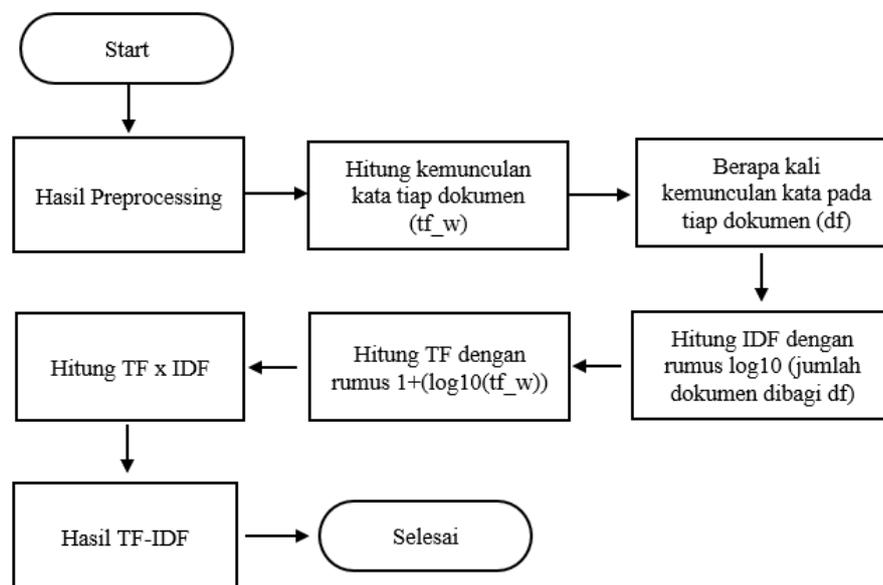
2.2 Preprocessing

Proses ini digunakan untuk membersihkan data dari *noise* dan siap untuk digunakan pada proses selanjutnya, proses *preprocessing* memiliki beberapa tahapan berikut merupakan tahap-tahap *preprocessing* [15].

1. *Cleansing*: Pada proses *cleansing* kalimat dibersihkan dari tanda baca, *URL*, *username*, *hashtag* (#) dan *emoticon*.
2. *Case folding*: Merupakan proses pengubahan huruf dari huruf yang bercampur (*lowercase* dan *uppercase*) menjadi semua huruf kecil.
3. *Tokenization*: Merupakan proses untuk merubah kalimat menjadi token-token atau potongan kata tunggal.
4. *Stopword Removal*: Merupakan penghilangan kata yang tidak berbobot meskipun kata tersebut sering kali hadir di dalam kalimat.
5. *Stemming*: Merupakan proses untuk mengubah kata-kata yang terdapat dalam suatu dokumen ke dalam kata-kata dasar.

2.3 Pembobotan Kata TF-IDF

Setelah melewati proses *preprocessing*, langkah selanjutnya adalah tahap pembobotan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Tahap pembobotan ini bertujuan untuk memberikan nilai frekuensi suatu kata sebagai bobot yang nantinya dapat di proses pada *Support Vector Machine*. Langkah pertama adalah menghitung nilai *term frequency* tiap kata. Langkah kedua adalah menghitung nilai *document frequency* tiap kata. Langkah ketiga adalah menghitung *inverse document frequency* dan langkah terakhir yaitu menghitung bobot atau *weight* dari hasil perkalian *term frequency* dikalikan dengan *inverse document frequency* [16]. Alur dari pembobotan TF-IDF dilihat pada Gambar 3.



Gambar 3. Proses Pembobotan TF-IDF

2.4 Klasifikasi

Merupakan suatu proses pemberian nilai objek data ke salah satu kelas tertentu dari beberapa kelas yang sudah didefinisikan menggunakan algoritma *Support Vector Machine* yang ditujukan untuk membuat model klasifikasi yang nyatakan dalam Tabel 2.

Tabel 2. Algoritma SVM [17]

Algoritma Support Vector Machine	
1.	Kelas data : $y_i \{-1, +1\}$, untuk $i = 1, 2, \dots, i$, yang mana i adalah banyaknya data.
2.	Diperoleh <i>decision function</i> berikut : $f(x) = \text{sign}(w \cdot x + b)$
3.	Berdasarkan pada <i>decision function</i> , diasumsikan kedua kelas dapat terpisah oleh <i>hyperplane</i> berdimensi d : <ul style="list-style-type: none"> - $\vec{w} \cdot \vec{x} + b = 0$ - $\vec{w} \cdot \vec{x} + b \leq -1$ - $\vec{w} \cdot \vec{x} + b \geq +1$ Dengan \vec{w} adalah <i>vector</i> bobot, \vec{x} adalah <i>vektor</i> data (<i>input</i>) dan b adalah <i>bias</i> .
4.	Fungsi keputusan klasifikasi $\text{sign}(f(x))$: $f(x) = w \cdot x + b$ atau $f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b$ Dengan a_i (nilai bobot setiap titik data), $K(x, x_i)$ (fungsi <i>kernel</i>).
5.	Persamaan <i>kernel</i> : <ul style="list-style-type: none"> - <i>Kernel Linear</i> $K(x, y) = x \cdot y$ - <i>Kernel RBF</i> $K(x, y) = \exp\left(-\frac{\ x - y\ ^2}{2\sigma^2}\right)$

2.5 Evaluasi

Metode evaluasi yang digunakan yaitu *confusion matrix* yang menghasilkan nilai *accuracy*, *precision* dan *recall* serta Kurva ROC untuk mengukur nilai AUC. Dengan semakin besar *area under curva* (AUC), semakin baik hasil prediksi. Berikut Tabel 3 merupakan *tabel confusion matrix* [18].

Tabel 3. Confusion Matrix [18]		
Aktual	Prediksi	
	True	False
True	TP	FN
False	FP	TN

Perhitungan dari *confusion matrix* dinyatakan dalam Persamaan (1) sampai dengan Persamaan (3). Parameter TP (*True Positive*) menunjukkan jumlah prediksi positif dari aktual kelas positif. FP (*False Positive*) menyatakan jumlah prediksi positif dari kelas aktual negatif. TN (*True Negative*) menyatakan jumlah prediksi negatif dari kelas aktual negatif. FN (*False Negative*) menunjukkan jumlah prediksi negatif dari kelas aktual positif [19].

$$\text{Accuracy (A)} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (1)$$

$$\text{Precision (P)} = \frac{(TP)}{(TP+FP)} \quad (2)$$

$$\text{Recall (R)} = \frac{(TP)}{(TP+FN)} \quad (3)$$

3. HASIL DAN ANALISIS

3.1 Preprocessing

Proses ini digunakan untuk membersihkan data dari *noise* dan siap untuk digunakan pada proses selanjutnya, proses *preprocessing* memiliki beberapa tahapan berikut merupakan tahap-tahap *preprocessing*.

1. Cleansing

Proses cleansing dalam sistem berguna untuk membersihkan data tweet seperti angka, tanda baca, link, hastag, metion, emoticon dan menghasilkan kata yang akan diolah pada proses selanjutnya. Hasil proses cleansing adalah menghilangkan kata dari karakter yang tidak ada artinya dan akan diganti oleh karakter spasi. Berikut Tabel 4 merupakan hasil dari proses *cleansing*.

Tabel 4. Hasil Proses Cleansing

<i>Text</i>	<i>Cleansing</i>
@BiznetHome maaf biznet internet error lama banget udah setengah hari mohon di proses https://t.co/jFpNKsj8Q3	maaf biznet internet error lama banget udah setengah hari mohon di proses
@.biznet kita baikan aja yuk, foto di twitter loadnya lama bgt nih%÷¼•ü □ udahan marahnya%÷¼•ü □	biznet kita baikan aja yuk foto di twitter loadnya lama bgt nih udahan marahnya

2. Case Folding

Case folding dilakukan untuk mengubah seluruh ukuran huruf pada kata menjadi suatu bentuk ukuran huruf yang sama. Karena tidak semua tweet konsisten dalam penggunaan ukuran huruf. Fungsi case folding ini akan mengubah semua huruf menjadi lower case atau huruf kecil. Berikut Tabel 5 merupakan hasil dari proses *case folding*.

Tabel 5. Hasil Proses Case Folding

<i>Text</i>	<i>Case Folding</i>
@BiznetHome maaf biznet internet error lama banget udah setengah hari mohon di proses https://t.co/jFpNKsj8Q3	maaf biznet internet error lama banget udah setengah hari mohon di proses
@.biznet kita baikan aja yuk, foto di twitter loadnya lama bgt nih%÷¼•ü □ udahan marahnya%÷¼•ü □	biznet kita baikan aja yuk foto di twitter loadnya lama bgt nih udahan marahnya

3. Tokenization

Proses *tokenizing* dilakukan untuk memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal. Pada saat bersamaan, *tokenizing* juga membuang karakter selain huruf seperti tanda baca. Berikut Tabel 6 merupakan hasil dari proses *tokenization*.

Tabel 6. Hasil Proses Tokenization

<i>Text</i>	<i>Tokenization</i>
@BiznetHome maaf biznet internet error lama banget udah setengah hari mohon di proses https://t.co/jFpNKsj8Q3	'maaf', 'biznet', 'internet', 'error', 'lama', 'banget', 'udah', 'setengah', 'hari', 'mohon', 'di', 'proses'
@.biznet kita baikan aja yuk, foto di twitter loadnya lama bgt nih%÷¼•ü □ udahan marahnya%÷¼•ü □	'biznet', 'kita', 'baikan', 'aja', 'yuk', 'foto', 'di', 'twitter', 'loadnya', 'lama', 'bgt', 'nih', 'udahan', 'marahnya'

4. Stopword Removal

Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data *tweet* dibandingkan dengan kata yang terdapat dalam database *stopword*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada database. Berikut Tabel 7 merupakan hasil dari proses *stopword removal*.

Tabel 7. Hasil Proses Stopword Removal

<i>Text</i>	<i>Stopword Removal</i>
@BiznetHome maaf biznet internet error lama banget udah setengah hari mohon di proses https://t.co/jFpNKsj8Q3	maaf biznet internet error banget udah mohon proses
@.biznet kita baikan aja yuk, foto di twitter loadnya lama bgt nih%÷¼•ü □ udahan marahnya%÷¼•ü □	biznet baikan aja yuk foto twitter loadnya bgt nih udahan marahnya

5. Stemming

Proses terakhir dalam *preprocessing* adalah *stemming*, fungsi dari *stemming* untuk mengubah kata yang berlebihan menjadi kata dasar agar memudahkan dalam proses pembobotan nantinya. Fungsi ini menggunakan algoritma *Nazief* dan *Adriani* yaitu *library Sastrawi*. Berikut tabel 8 merupakan hasil dari proses *stemming*.

Tabel 8. Hasil Proses Stemming

<i>Text</i>	<i>Stopword Removal</i>
@BiznetHome maaf biznet internet error lama banget udah setengah hari mohon di proses https://t.co/jFpNKsj8Q3	maaf biznet internet error banget udah mohon proses
@.biznet kita baikan aja yuk, foto di twitter loadnya lama bgt nih%0÷¼•ü □ udahan marahnya%0÷¼•ü □	biznet baik aja yuk foto twitter loadnya bgt nih udahan marah

3.2 Pembobotan TF-IDF

Pada pembobotan *term* dilakukan proses pemberian nilai atau bobot terhadap setiap *term* yang ada pada setiap *tweet* yang telah melawati tahap *preprocessing*. Metode yang digunakan dalam pemberian bobot pada *term* yaitu TF-IDF. Pembobotan ini bertujuan untuk memberikan nilai kepada sebuah *term* dimana nilai dari *term* tersebut akan dijadikan sebagai input pada proses klasifikasi. Berikut Tabel 9 merupakan hasil dari proses pembobotan menggunakan TF-IDF.

Tabel 9. Hasil Pembobotan TF-IDF

No.	Kata	tf			df	idf $\log(\frac{N}{df})$	W = df * idf		
		D1	D2	D3			D1	D2	D3
1	Biznet	1	1	1	3	0	0	0	0
2	Aman	1			1	0.47712	0.47712	0	0
3	Nyaman	1			1	0.47712	0.47712	0	0
4	Internet	1			1	0.47712	0.47712	0	0
5	Kecang	1			1	0.47712	0.47712	0	0
6	Stabil	1			1	0.47712	0.47712	0	0
7	Harga	1			1	0.47712	0.47712	0	0
8	Murah	1			1	0.47712	0.47712	0	0
9	Banget	1			1	0.47712	0.47712	0	0
10	Home		1		1	0.47712	0	0.47712	0
11	Lama		1		1	0.47712	0	0.47712	0
12	Kepret		1		1	0.47712	0	0.47712	0
13	Ya		1		1	0.47712	0	0.47712	0
14	Bulan		1		1	0.47712	0	0.47712	0
15	Kali		1		1	0.47712	0	0.47712	0
16	Mati		1	1	2	0.17609	0	0.17609	0.17609
17	Internet		1		1	0.47712	0	0.47712	0
18	hari		1		1	0.47712	0	0.47712	0
19	meeting			1	1	0.47712	0	0	0.47712
20	online			1	1	0.47712	0	0	0.47712
21	beda			1	1	0.47712	0	0	0.47712
22	putus			1	1	0.47712	0	0	0.47712
23	sulit			1	1	0.47712	0	0	0.47712
24	andal			1	1	0.47712	0	0	0.47712
			Total				3.81696	3.99305	3.03881

No.	Kata	tf			df	idf	W = df * idf				
		D1	D2	D3			D1	D2	D3		
					$\log\left(\frac{N}{df}\right)$						
			y				1	-1	-1		

Pembobotan kata dilakukan untuk memberikan nilai bobot pada setiap term. Hasil dari nilai bobot tersebut kemudian akan digunakan dalam pembentukan vektor pada klasifikasi SVM

3.3 Klasifikasi dan Pengujian

Dalam proses klasifikasi Support Vector Machine dibutuhkan bobot setiap kata yang ada pada data training. Nilai bobot setiap term ini didapat dari proses pembobotan TF-IDF yang akan digunakan dalam pembuatan model dari algoritma Support Vector Machine. Berikut Gambar 4 adalah source code dari proses klasifikasi Support Vector Machine.

```
# Process of making models Klasifikasi SVM LINEAR
linear = svm.SVC(kernel="linear")

linear.fit(X_train,y_train)
linear = linear.predict(X_test)

# Process of making models Klasifikasi SVM RBF
rbf = svm.SVC(kernel="rbf", gamma=1, C=1)

rbf.fit(X_train,y_train)
rbf = rbf.predict(X_test)
```

Gambar 4. Source Code SVM

Pengujian dilakukan untuk mengetahui kinerja SVM dalam proses klasifikasi. Pengujian ini dilakukan dengan memasukkan data training dengan jumlah berdasarkan skenario yang telah ditentukan pada Tabel 10 dan untuk perbandingan data training dan data testingnya pada pengujian ini ialah 90% data training dan 10% data testing.

Tabel 10. Skenario Pembagian Data Pengujian

Skenario	Jumlah Data Training		Jumlah Data Testing (10% dari Total Data)	Total Data
	(90% dari Total Data)			
	Positif	Negatif		
1	360	360	80	800
2	405	405	90	900
3	450	450	100	1000

3.4 Evaluasi

Pengujian kinerja dari Support Vector Machine dilakukan dengan metode confusion matrix berdasarkan data testing yang akan di proses oleh model klasifikasi Support Vector Machine yang telah dibuat sebelum. Metode confusion matrix menghasilkan empat nilai yaitu True Positive, True Negative, False Positive dan False Negative selanjutnya dilakukan perhitungan mulai dari accuracy, precision dan recall. Berikut adalah hasil dari pengujian dengan data training 90% dan data testing 10%. Berikut ini Tabel 11 merupakan tabel confusion matrix hasil evaluasi terhadap algoritma SVM.

Tabel 11. Confusion Matrix Keseluruhan Skenario

	Skenario 1				Skenario 2				Skenario 3			
	TP	TN	FP	FN	TP	TN	FP	FN	TP	TN	FP	FN
SVMLinear Kernel	33	28	5	14	38	38	4	10	39	51	5	5
SVMRBF Kernel	33	28	5	14	39	38	4	9	39	49	7	5

Dari hasil confusion matrix dengan menggunakan persamaan (1), (2), dan (3) maka dapat dihitung nilai Akurasi, Presisi dan Recall. Berikut Tabel 12 hasil evaluasi kinerja algoritma SVM.

Tabel 12. Hasil Evaluasi Kinerja Algoritma SVM

Evaluasi	SVM Linear	SVM RBF
Skenario 1	Akurasi	76.25 %
	Presisi	86.84 %
	Recall	70.21 %
Skenario 2	Akurasi	84.44 %
	Presisi	90.47 %
	Recall	79.16 %
Skenario 3	Akurasi	90 %
	Presisi	88.63 %
	Recall	88.63 %

Berdasarkan keseluruhan skenario pengujian yang telah dilakukan menunjukkan bahwa hasil dari analisis sentimen dengan menggunakan algoritma SVM dengan *kernel linear* dan *kernel RBF* yang melalui 3 skenario pengujian memiliki akurasi untuk *Kernel linear* secara berurutan yaitu 76.25 %, 84.44 % dan 90 % yang terus meningkat sedangkan pada *kernel RBF* hasil akurasi yang didapat lebih kecil yaitu 76.25 %, 85.55 % dan 88 %. Untuk nilai presisi secara berurutan SVM *kernel linear* menghasilkan nilai 86.84%, 90.47% dan 88.63 %, sedangkan pada *kernel RBF* menghasilkan nilai presisi 86.84 %, 90.69 % dan 84.78. Presisi merupakan presentase dari keakuratan model hasil klasifikasi SVM oleh sistem untuk mengklasifikasikan kalimat yang bernilai positif, karena hasil klasifikasi terus menurun maka kemampuan untuk mengklasifikasi kalimat bernilai positif semakin berkurang. Sedangkan untuk nilai recall secara berurutan SVM *kernel linear* menghasilkan nilai 70.21 %, 79.16 % dan 88.63 % sedangkan pada *kernel RBF* menghasilkan nilai recall 70.21 %, 81.45% dan 88.63 %. Disisi lain pengujian sebanyak 3 skenario dilakukan untuk mengetahui konsistensi kinerja algoritme SVM baik *kernel RBF* maupun *kernel Linear* apabila jumlah data yang diujikan jumlahnya berbeda. Berdasarkan pengujian melalui 3 skenario yang telah dilakukan menunjukkan bahwa perbedaan jumlah data yang diproses sangat berpengaruh terhadap kinerja algoritme SVM baik *kernel RBF* maupun *kernel Linear*. Selain jumlah data, kinerja algoritme SVM dalam kasus klasifikasi sentiment ini ialah dengan adanya keterbatasan kamus kata. Hal ini disebabkan banyak kata-kata gaul atau kata singkat yang seringkali ditemukan pada suatu tweet disosial media twitter yang secara langsung akan mempengaruhi tahap pra-pemrosesan data.

4. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan menggunakan *kernel linear* dan *kernel RBF* dengan 3 skenario pengujian pada setiap kernel, dapat diambil kesimpulan sebagai berikut. Algoritma SVM menggunakan dengan *kernel linear* maupun *kernel RBF* memiliki hasil kinerja evaluasi baik dari sisi akurasi, presisi dan recall yang relatif sama. Sehingga dapat dikatakan bahwa algoritma SVM baik dengan *kernel RBF* maupun *Linear* sama-sama dapat digunakan dengan baik dalam menentukan sentimen pengguna internet Biznet. Selain itu dengan 3 skenario pengujian dengan jumlah data yang berbeda algoritma SVM baik dengan *kernel RBF* maupun *Linear* sama-sama konsisten kinerjanya. Untuk penelitian lebih lanjut hal yang perlu dilakukan ialah membuat kamus pada database kata gaul dan kata singkat, karena banyak tweet yang menggunakan bahasa kurang baku.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Universitas AMIKOM Yogyakarta yang telah memberikan dukungan penuh dalam penelitian ini.

REFERENSI

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia, "Penetrasi & Profil Perilaku Pengguna Internet Indonesia," *Apjii*, p. 51, 2018.
- [2] A. Balahur, R. Mihalcea, and A. Montoyo, "Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications," *Computer Speech and Language*, vol. 28, no. 1, pp. 1–6, 2014.
- [3] kominfo, "Indonesia Peringkat Lima Pengguna Twitter," <https://kominfo.go.id/>, 2012. .
- [4] P. R. Shaver, U. Murdaya, and R. C. Fraley, "Structure of the Indonesian emotion lexicon," *Asian Journal of Social Psychology*, vol. 4, no. 3, pp. 201–224, 2001.
- [5] R. Tiniges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, p. 650, 2020.
- [6] C. . Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," *Cambridge University Press*, 2008.
- [7] I. Rozi, S. Pramono, and E. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi," *Jurnal EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.

- [8] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press, 2015.
- [9] A. Fathan Hidayatullah, M. Rifqi Ma, and arif Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta Jl Ringroad Barat, "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) Agustus*, pp. 1907–5022, 2016.
- [10] H. Yu, X. Huang, X. Hu, and H. Cai, "A comparative study on data mining algorithms for individual credit risk evaluation," *Proceedings - 2010 International Conference on Management of e-Commerce and e-Government, ICMCG 2010*, pp. 35–38, 2010.
- [11] B. Pang and L. Lee, "Thumb Up? Sentiment Classification Using Machine Learning Techniques," *Proceeding of the Conference on Empirical Method in Natural Language Processing Empirical Method in Natural Language Processing*, pp. 79–86, 2002.
- [12] E. Tyagi and A. K. Sharma, "Sentiment Analysis of Product Reviews using Support Vector Machine Learning Algorithm," *Indian Journal of Science and Technology*, vol. 10, no. 35, pp. 1–9, 2017.
- [13] P. H. Saputro, M. Aristin, and Dy. L. Tyas, "Klasifikasi Lagu Daerah Indonesia Berdasarkan Lirik Menggunakan Metode Tf-Idf Dan Naïve Bayes," *Jurnal Teknoloi Informatika dan Terapan*, vol. 4, no. 1, pp. 45–50, 2017.
- [14] K. A. B. Permana, M. Sudarma, and W. G. Ariastina, "Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan STRUCT-SVM," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 113, 2019.
- [15] N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers," *Procedia Computer Science*, vol. 72, pp. 519–526, 2015.
- [16] A. D. Hartanto, A. Syaputra, and Y. Pristyanto, "Best parameter selection of rabin-Karp algorithm in detecting document similarity," 2019.
- [17] C.-W. Hsu, C. Chih-Chung, and C.-J. Lin, "A Practical Guide to Support Vector Classification," *BJU international*, vol. 101, no. 1, pp. 1396–400, 2008.
- [18] B. Max, *Principles of Data Mining*. London: Springer, 2007.
- [19] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2007.