

ALGORITMA SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI SIKAP POLITIK TERHADAP PARTAI POLITIK INDONESIA

Satrio Yudho Pangestu¹, Yuli Astuti², Lilis Dwi Farida³

¹Informatika,
²Manajemen Informatika,
²Sistem Informasi

Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, Jl. Ringroad Utara, Condong Catur,
Sleman, Yogyakarta

Email: ¹satrio.pangestu@students.amikom.ac.id, ²yuli@amikom.ac.id, ³rida@amikom.ac.id

Abstract

The use of social media that is increasingly easy and affordable becomes a new forum for Indonesian people to express their thoughts freely. Included in the preparation period for a democratic party held every five years. The public can freely believe through the social media they have, especially through twitter. People who come from different backgrounds often provide opinions that can lead to pros and cons. It can be used as feedback on political parties that carry presidential, vice-presidential, and successful team candidates so that they will be useful in potential assessments and can be used for better purposes. Sentiment analysis is done by sorting data from Twitter which is an opinion on political parties and the executive candidates they carry. The data is divided into 2 categories, positive and negative categories. The methods that will be used for sentiment analysis include preprocessing, word staining with TF-IDF, and making a classification model with the Support Vector Machine and K-Fold Cross Validation approach to test the accuracy of the model. The result of making a classification model is Support Vector Machine with training data of 900 to get 86% accuracy and testing using 10-Fold Cross Validation get an average accuracy rate of 71% with an error rate of 29%

Keywords: Sentiment analysis, Twitter, opinion mining, Support Vector Machine, politics

Abstrak

Penggunaan media sosial yang semakin mudah dan terjangkau menjadi wadah baru bagi masyarakat Indonesia untuk mengekspresikan pemikirannya secara bebas. Termasuk pada masa persiapan pesta demokrasi yang dilaksanakan setiap lima tahun sekali. Masyarakat dapat beropini dengan bebas melalui media sosial yang dimiliki, terutama melalui *twitter*. Masyarakat yang berasal dari latar belakang berbeda seringkali memberikan opini yang dapat menimbulkan pro dan kontra. Hal itu dapat dijadikan *feedback* terhadap partai politik yang mengusung pasangan calon presiden, wakil presiden, dan tim sukses sehingga akan berguna dalam penilaian potensial dan dapat digunakan untuk tujuan yang lebih baik. Analisis sentimen dilakukandengan cara memilah data dari twitter yang merupakan opini terhadap partai politik dan calon eksekutif yang diusungnya. Data tersebut dipilah menjadi 2 kategori yaitu kategori positif dan negatif. Metode-metode yang akan digunakan untuk analisis sentimen diantaranya *preprocessing*, pembedaan kata dengan TF-IDF, dan membuat model klasifikasi dengan pendekatan Support Vector Machine dan K-Fold Cross Validation untuk menguji tingkat akurasi model. Hasil dari pembuatan model klasifikasi *Support Vector Machine* dengan data latih sejumlah 900 mendapatkan *accuracy* sebesar 86% dan pengujian dengan menggunakan *10-Fold Cross Validation* mendapatkan tingkat rata – rata *accuracy* sebesar 71% dengan *error rates* sebesar 29%.

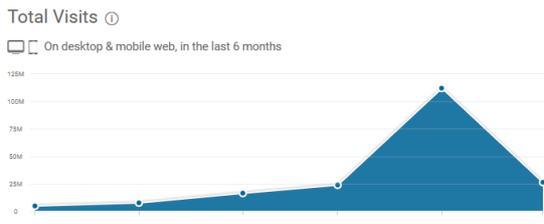
Kata Kunci: Analisis sentimen, Twitter, opinion mining, Support Vector Machine, politik

1. Pendahuluan

Pemilihan umum yang merupakan agenda lima tahunan selalu menjadi hal penting bagi masyarakat Indonesia. Komisi Pemilihan Umum (KPU) telah membuat jadwal tahapan pemilu yang dimulai sejak 17 Agustus 2017 untuk perencanaan program dan anggaran sampai dengan proses dan pelantikan yang akan dilaksanakan pada bulan Oktober 2019 [1].

Animo masyarakat terhadap pesta demokrasi tentu saja sangat besar. Data akses yang diperoleh melalui *similiarweb.com* mengenai tingkat pengaksesan halaman *kpu.go.id* menunjukkan bahwa website tersebut menempati peringkat 79 di Indonesia. Total pengunjung selama kurun waktu kurang lebih enam bulan semenjak bulan Desember 2018 yaitu 24.14M (76.68%), dengan grafik yang semakin meningkat di setiap

bulannya, dan mencapai grafik tertinggi pada bulan April 2019 (Gambar1) [2].



Gambar1. Trafik Akses Website KPU[2]

Akses internet yang memadai dan media sosial yang semakin beragam, menjadikan wadah perbincangan masyarakat Indonesia mengenai pemilihan umum menjadi semakin bebas dan meluas. Pembahasan pemilihan umum tentu saja tidak akan terlepas dari pro dan kontra. Opini masyarakat yang berasal dari berbagai latar belakang berbeda, dapat dijadikan *feedback* terhadap partai politik yang mengusung pasangan calon presiden, wakil presiden, dan tim sukses sehingga akan berguna dalam penilaian potensial dan dapat digunakan untuk tujuan yang lebih baik.

Salah satu media sosial yang digunakan adalah *twitter* yang berkembang menjadi sumber informasi yang kaya dan beragam. *Twitter* telah hampir memiliki 600 juta pengguna dan lebih dari 250 juta pesan per hari, dan dengan cepat akan menjadi tambang data bagi organisasi untuk mendapatkan informasi mengenai reputasi dan merk mereka dengan mengekstrasi dan menganalisis sentimen melalui *tweet* para pengguna *twitter*[3]. Analisis sentimen merupakan sebuah pemikiran bahwa nilai dari sebuah pemikiran menjadi hal yang penting dan dibutuhkan.

Terdapat beberapa penelitian yang telah dilakukan dalam klasifikasi sentimen pada *twitter*. Perbandingan beberapa algoritma klasifikasi seperti Naïve Bayes(NB), Support Vector Machine(SVM), dan Artificial Neural Network(ANN), diusulkan untuk analisis sentimen pada review film. Hasil komparasi menunjukkan bahwa SVM memiliki tingkat akurasi terbaik yaitu 81.1% dan AUC 0.904 [4].

Penelitian yang lain mengenai klasifikasi topik keluhan pelanggan berdasarkan *tweet* menggunakan gabungan *feature* hasil ekstraksi pada metode SVM memberikan kesimpulan pada metode *term frequency* (TF) menghasilkan akurasi paling baik sebesar 82.50% untuk klasifikasi bentuk keluhan. Sedangkan untuk jenis keluhan, mendapat akurasi sebesar 86.67% dengan metode *chi-square*[5].

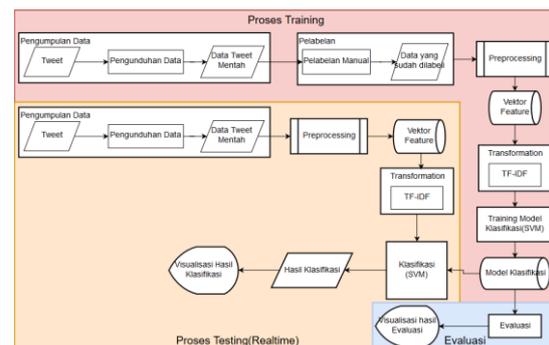
Sementara itu, metode Support Vector Machine (SVM) dengan menggunakan *Lexicon*

Based Features sebagai pembaruan fiturnya menghasilkan nilai akurasi sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97% dan *f-measure* sebesar 78%. Berdasarkan penelitian yang telah dilakukan sebelumnya, penelitian ini menggunakan algoritma Support Machine Machine (SVM) dan K-fold untuk mengetahui keberhasilan sistem. Klasifikasi sifat politik dari pengguna *twitter* Indonesia berdasarkan jumlah prosentase sentimen positif dan negatif dari *tweet* yang telah diunggah.

2. Metode Penelitian

Metode yang digunakan pada penelitian ini, seperti yang dijelaskan pada Gambar2, meliputi:

1. Pengumpulan Data
2. Pelabelan Data
3. *Preprocessing*:
 - a. *Tweet Cleaning*
 - b. *Tokenize*
 - c. *Normalisasi*
 - d. *Stopword Removal*
 - e. *Stemming*
 - f. *Vektor Feature*
4. Pembobotan
5. Pelatihan Model Klasifikasi SVM
6. Klasifikasi SVM
7. Evaluasi:
 - a. *K-fold Cross Validation*
 - b. *Confusion Matrix*



Gambar2. Tahapan Penelitian

Pengumpulan data *tweet* dilakukan dengan memanfaatkan *library* python yang telah tersedia yaitu *tweepy*. *Library* ini memudahkan dalam mengakses server *twitter* dengan menggunakan kode unik API *twitter*. Data yang diperoleh merupakan hasil dari *authentication* yang selanjutnya dilakukan pencarian berdasarkan kata kunci tertentu. *Tweet* diunduh pada kurun waktu antara tanggal 3-4 April 2019.

Proses selanjutnya adalah pelabelan 900 data *twitter* yang dilakukan secara manual. Untuk mengurangi penilaian secara subyektif, proses pelabelan dilakukan dengan bantuan dua suk

relawan. Pelabelan data diberi label positif, ditandai dengan angka "1", dan label negatif yang ditandai dengan angka "0".

Tahapan *preprocessing* merupakan tahapan awal untuk mempersiapkan data yang akan diolah. Tahapan *preprocessing* terbagi menjadi beberapa tahapan, yaitu:

1. *Cleaning Data* adalah tahapan pembersihan kata –kata yang tidak diperlukan seperti karakter hastag"# dan username
2. *Case Folding* merupakan tahapan untuk mengubah data tweet secara keseluruhan untuk menjadi lower case (huruf kecil)
3. *Normalisasi* kata merupakan proses untuk merubah kata – kata tidak baku menjadi baku seperti "msyrk" menjadi "masyarakat"
4. *Stopword removal* merupakan tahapan penghapusan kata – kata bantu seperti "saya", "dia", "aku", dan "mereka"
5. *Tokenizing* merupakan proses untuk mem bagi teks menjadi kata, seperti "aku sedang semangat belajar" menjadi "aku", "sedang", "semangat", "belajar".
6. *Stemming* merupakan tahapan untuk membersihkan kata – kata imbuhan awalan dan akhiran yang terdapat dalam teks seperti "berlari" menjadi "lari".

Hasil dari proses *cleaning data* seperti yang adapada Tabel 1.

TABEL 1. HASIL CLEANING DATA

Sebelum	Sesudah
Presiden ku @jokowi memang Mantap Betul Kinerjanya #KompakSATUJokowi#jokowi#01JokowiLagi https://twitter.com/Puspita25936554/status/1107461613181767680/photo/1 "Sampaikan ke @jokowi. Tunggu saya keluar dan rasakan pedasnya lidah saya," kata Bahar menebarkan ancaman. #baharbinsmith#jokowihttps://t.co/eFNXMT16AT	Presiden ku memang Mantap Betul Kinerjanya Sampaikan ke Tunggu saya keluar dan rasakan pedasnya lidah saya kata Bahar menebarkan ancaman

Selanjutnya adalah tahapan pembobotan dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Frekuensi kemunculan kata (TF) di dalam dokumen menunjukan seberapa penting suatu kata berada di dalam sebuah dokumen. Semakin banyak kata tersebut muncul, menunjukkan seberapa umum kata tersebut. Dan semakin banyak dokumen mengandung sebuah *term*, maka bobotnya akan semakin tinggi. Sementara itu, *Inverse Document Frequency* member bobot pada *term* yang memiliki frekuensi dokumen rendah atau pengukuran keunikan sebuah *term* dalam suatu dokumen pada sekumpulan dokumen (*corpus*) [6].

Rumus IDF seperti pada rumus (1):

$$idf = \log \frac{D}{df_i} \quad (1)$$

Dengan:

D = jumlah dokumen

df_i = kemunculan *term* dari D

Setelah mendapatkan IDF, maka dapat dilakukan perhitungan TF-IDF(w) untuk setiap *term* dengan persamaan (2).

$$W_{ij} = tf_{ij} + \left(\log \frac{D}{df_i} \right) \quad (2)$$

Dengan:

W_{ij} = bobot kata/*term*_{*t*} terhadap dokumend_{*i*}

tf_{ij} = jumlah kemunculan kata/*term*_{*t*} dalam dokumen d_{*i*}

SVM merupakan suatu teknik yang digunakan untuk melakukan prediksi pada kasus klasifikasi maupun regresi[7]. Konsep klasifikasi dengan SVM adalah mencari *hyperlane* terbaik yang memiliki fungsi sebagai pemisah dari dua kelas data, opini positif (+1) dan opini negatif(-1). Persamaan SVM seperti pada rumus (3).

$$w \cdot x + b = 0 \quad (3)$$

Dengan:

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

x = titik data masukan Support Vector Machine

b = parameter *hyperlane* yang dicari (nilai bias)

3. Hasil dan Pembahasan

Penelitian yang dilakukan menggunakan data *tweet* berbahasa Indonesia yang diperoleh melalui *twitter* pada tanggal 3-4 April 2019, setelah debat capres-cawapres yang keempat. Kata kunci yang digunakan antara lain adalah "Jokowi", "Prabowo", "PDIP", dan "Gerindra". Total data yang diperoleh sebanyak 900 data yang selanjutnya dilakukan pelabelan sentimen secara manual.

Pembobotan dengan TF-IDF dilakukan dengan menganalisis data *tweet* dengan kode T1-T6. Pemberian bobot dilakukan dengan mengalihkan nilai TF dengan IDF. Sebagai contoh implementasi manual TF-IDF, maka *tweet* tersebut akan digunakan sebagai data *training* dan juga data *testing*. Setelah mendapatkan TF-IDF dari sebuah dokumen, hasil tersebut kemudian

dikumpulkan dan dihitung menggunakan rumus (1) dan (2) sehingga didapatkan hasil akhir seperti pada Tabel2.

TABEL2. HASIL AKHIR TF-IDF

TF-IDF	Dokumen/ Tweet					
	T1	T2	T3	T4	T5	T6
Positif	5.0	0.0	7.0	21.6	1.5	1.5
Negatif	1.5	12.4	0.0	1.5	19.3	33.5
Label	1	-1	1	1	-1	-1

Selanjutnya adalah implementasi SVM dengan menggunakan rumus (3). $W = \{W_1, W_2, \dots, W_n\}$ merupakan bobot vektor, yang dalam perhitungan manual ini menggunakan dua W . Jika berdasarkan pada $W_1 = \text{positif}$ dan $W_2 = \text{negatif}$, dengan pemisalan $X(x_1, x_2)$ sebagai nilai atribut W dan jika b dianggap sebagai bobot tambahan W_0 maka persamaan suatu bobot tambahan w_0 , maka persamaan suatu *hyperlane* dapat ditulis ulang sebagai berikut:

$$W_0 + W_1X_1 + W_2X_2 = 0 \quad (4)$$

Berdasarkan hasil perhitungan bobot TF-IDF pencarian *hyperlane* dapat dilakukan dengan pendekatan:

$$y_i(W_0 + W_1X_1 + W_2X_2 + b) \geq 0 \quad (5)$$

Sebagai pendekatan *one-against-all*, sehingga terdapat dua fungsi klasifikasi, yaitu fungsi untuk menentukan kelas perhitungan manual untuk memilih kelas tersebut, penulis menggunakan positif dan fungsi untuk menentukan kelas negatif. Berikut merupakan penerapan fungsi kelas positif dengan memasukan *support vector* T1, T3, dan T6.

$$T_1 = (5.0W_1 + 1.5W_2 + b) \geq 1$$

$$T_3 = (10.6W_1 + 0W_2 + b) \geq 1$$

$$T_6 = (1.5W_1 + 33.5W_2 + b) \geq -1$$

Untuk mendapatkan nilai W_1 , W_2 dan b akan dilakukan penyelesaian persamaan linear untuk fungsi positif. Proses ini dapat dilakukan dengan menggunakan matrix, perkalian antara *inversmatrix* dan *matrix* konstanta, seperti dibawah ini:

$$\begin{bmatrix} 5.0 & 1.5 & 1 \\ 10.6 & 0 & 1 \\ 1.5 & 33.5 & 1 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} -0.192 & 0.184 & 0.008 \\ -0.053 & 0.020 & 0.032 \\ 2.045 & -0.955 & -0.090 \end{bmatrix} \begin{bmatrix} w1 \\ w2 \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} w1 \\ w2 \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} -0.192 & 0.184 & 0.008 \\ -0.053 & 0.020 & 0.032 \\ 2.045 & -0.955 & -0.090 \end{bmatrix}$$

$$\begin{bmatrix} w1 \\ w2 \\ b \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.06 \\ 1.18 \end{bmatrix}$$

Berdasarkan hitungan manual diatas didapatkan nilai $W_1 = -0.02$ $W_2 = -0.06$ dan $b = 1.18$. Apabila dimasukkan kepersamaan sebagai *hyperlane* pada label positif:

$$-0.02X_1 - 0.06X_2 + 1 = 0$$

Hal yang sama dikenakan pada training data negative, sehingga hasil akhir pada perhitungan adalah seperti pada Tabel3.

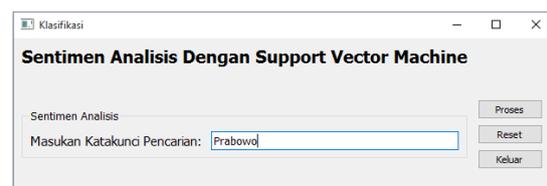
TABEL3. HASIL PERSAMAAN W_1 , W_2 , DAN B PADA SETIAP KELAS

Fitur	Positif	Negatif
W_1	-0.02	0.2
W_2	-0.06	0
B	1	-1

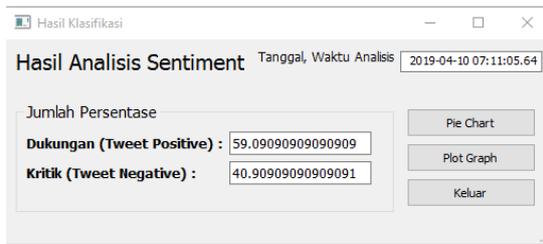
Implementasi SVM pada antarmuka untuk mengetahui dukungan politik dapat memberikan prosentase yang jelas pada pengguna. Gambar 3 dan Gambar 4 merupakan halaman implementasi algoritma SVM untuk klasifikasi sentimen analisis pada calon presiden yang ada. Gambar 5 dan Gambar 6 adalah hasil dari perhitungan SVM yang diimplementasikan ke dalam sebuah antarmuka.



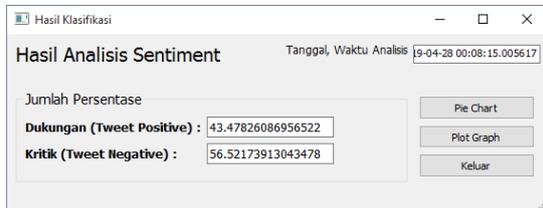
Gambar 3 . Menu Klasifikasi dengan Kata Kunci "Jokowi"



Gambar 4. Menu Klasifikasi dengan Kata Kunci "Prabowo"

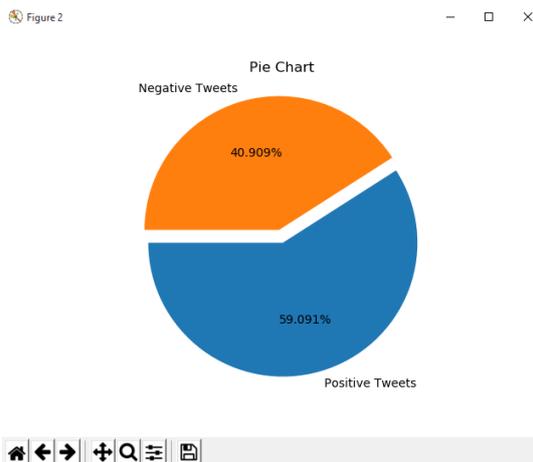


Gambar 5. HasilKlasifikasi “Jokowi”



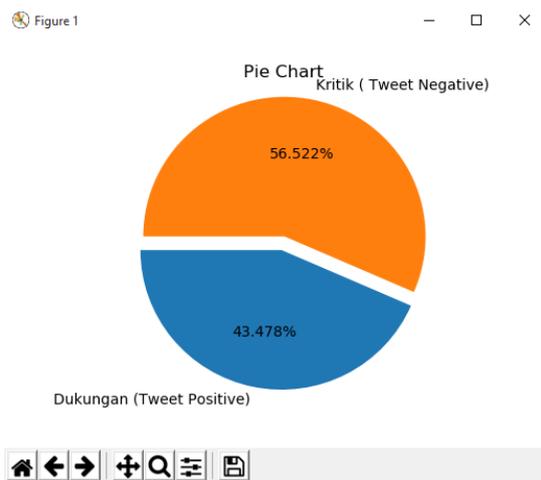
Gambar 6. HasilKlasifikasi “Prabowo”

Hasil ujicoba pencarian kata “Jokowi” dengan waktu *realtime* menghasilkan prosentase dukungan atau *tweet* positif sebesar 59.09% dan kritik dari publik sebesar 40.909%. Sementara pada kata kunci “Prabowo”, dukungan atau *tweet* positif mencapai 43.478% dan kritik sebesar 56.522%. Hal ini menunjukkan bahwa pada detik tersebut saat aplikasi dijalankan, opini publik mendukung Jokowi merupakan representasi dari partai PDIP. Dan sebaliknya, opini publik terhadap Prabowo merupakan representasi dari partai Gerindra. Grafik prosentase hasil klasifikasi seperti yang ada pada Gambar 7 dan Gambar 8.



Gambar 7. DiagramHasilAnalisisSentimenpada Kata Kunci “Jokowi”

Model yang telah dibuat dan di testing secara *realtime* akan melalui proses evaluasi. Hal ini berguna untuk mendapatkan tingkat *accuracy*, *precision*, *recall*, dan *F-score*. Proses evaluasi menggunakan metode *Confussion Matrix* dan *K-Fold Cross Validation* yang telah tersedia pada *library python scikit learn*. Proses evaluasi dibagi menjadi dua skenario.



Gambar 8. DiagramHasilAnalisisSentimenpada Kata Kunci “Prabowo”

a. *Confussion Matrix*

Pengujian model klasifikasi membutuhkan data testing yang telah memiliki label. Data yang digunakan sebagai data testing evaluasi yaitu data yang juga digunakan sebagai data latih, dengan jumlah data 900 yang terbagi atas 450 data positif dan 450 data negatif.

Tabel 4. *Confussion Matrix*Skenario I

	NilaiSebenarnya		Jumlah	
	Positif	Negatif		
NilaiPrediksi	Positif	390	60	450
	Negatif	67	383	450
Jumlah		457	443	900

Pada Tabel 4, terdapat 390 *tweet* yang diklasifikasi positif dan 388 *tweet* yang diklasifikasi negatif. Selain itu terdapat kesalahan pada klasifikasi yaitu *tweet* yang seharusnya bernilai positif sebanyak 67 *tweet* dan yang seharusnya negatif adalah 60 *tweet*. Hasil perhitungan untuk mendapat nilai:

1) *Precision*

$$Precision = \frac{390}{390 + 60} = 0.87 * 100\% = 87\%$$

2) *Recall*

$$Recall = \frac{390}{390 + 67} = 0.85 * 100\% = 85\%$$

3) *F-Score*

$$F - Score = 2 * \frac{0.87 * 0.85}{0.87 + 0.85} = 0.859 * 100\% = 86\%$$

4) Accuracy

$$Accuracy = \frac{390 + 383}{390 + 383 + 60 + 67}$$

$$= 0.86 * 100\% = 86\%$$

b. K-Fold Cross Validation

Proses evaluasi skenario kedua menggunakan *Confussion Matrix* dan *K-Fold Cross Validation* dengan jumlah *K-Fold* =10. Skenario pengujian ini adalah membagi 900 data menjadi 10 bagian/*fold*, sehingga tiap bagian berisi 90 data berlabel yang terdiri dari 45 dari data positif dan 45 data negatif. Hasil akhir dari skenario adalah rata-rata *accuracy*, *preccision*, *recall*, dan *F-score* dari 10 *fold*. Hasil total pengolahan hasil pengujian pada tiap perulangan dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pengujian Model dengan 10-Fold Cross Validation dan Confussion Matrix

10-fold cross validation	Confussion Matrix				
	Accura cy	Error rate	preccissi on	recal l	F- score
Fold 1	73%	27%	82%	70%	76%
Fold 2	71%	29%	80%	68%	73%
Fold 3	68%	32%	69%	67%	68%
Fold 4	71%	29%	73%	70%	72%
Fold 5	74%	26%	69%	78%	73%
Fold 6	69%	31%	71%	68%	70%
Fold 7	72%	28%	80%	69%	74%
Fold 8	70%	30%	64%	73%	68%
Fold 9	61%	39%	62%	61%	62%
Fold 10	78%	22%	80%	77%	78%
Rata-rata	71%	29%	73%	70%	71%

Pengujian kombinasi *K-Fold Cross Validation* dan *Confussion Matrix* memiliki nilai *accuracy* tertinggi pada *fold* ke-10, yaitu 78%. Sementara *error rate* terendah berada pada *fold* ke-10. Nilai *precision* tertinggi pada *fold* ke-1 yaitu 82%, nilai *recall* tertinggi adalah 78% pada *fold* ke-5. Dan nilai *f-score* tertinggi pada *fold* ke-10 yaitu sebesar 78%.

Secara keseluruhan *accuracy* model adalah 71%, *error rate* sebesar 29%, *precision* sebesar 73%, *recall* sebesar 70% dan *f-score* sebesar 71%. Hal ini menunjukkan algoritma Support Vector Machine bagus dan layak digunakan untuk klasifikasi data tekstual.

4. Kesimpulan

Berdasarkan hasil pembahasan yang telah diuraikan di atas maka dapat dibuat kesimpulan sebagai berikut:

- 1) Penelitian ini telah menguji kemampuan algoritma *Support Vector Machine* untuk klasifikasi data *Tweet* berbahasa Indonesia dengan total data 900 dan diperoleh rata – rata akurasi sebesar 71%, *error rate* sebesar 29% *precision* sebesar 73%, *recall* sebesar 70% dan *f-score* sebesar 71%.
- 2) Hasil penelitian ini menunjukkan algoritma *Support Vector Machine* dapat diimplementasikan untuk mengklasifikasikan data *Twitter* berbahasa Indonesia

5. Referensi

- [1] K. P. Umum, “Portal Publikasi Pilkada dan Pemilu Indonesia,” 2018. [Online]. Available: <https://infopemilu.kpu.go.id/>. [Accessed: 20-Jun-2019].
- [2] SimiliarWeb, “Similiar Web: www.kpu.go.id,” 2019. [Online]. Available: <https://www.similarweb.com/website/kpu.go.id#pro>. [Accessed: 20-Jun-2019].
- [3] D. A. Putri, “Penerapan Algoritma Support Vector Machine Berbasis Algoritma Genetika untuk Analisis Sentimen,” *J. Tek. Inform. STMIK Antar Bangsa*, vol. I, no. 01, pp. 1–7, 2015.
- [4] V. Chandani, “Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film,” *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.
- [5] E. E. Pratama and B. R. Trilaksono, “Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM),” *J. Edukasi dan Penelit. Inform.*, vol. 1, no. 2, 2015.
- [6] R. Intan and A. Defeng, “Hard : Subject-Based Search Engine Menggunakan TF-IDF dan Jaccard’s Coefficient,” *J. Tek. Ind.*, vol. 8, no. 1, pp. 61–72, 2006.
- [7] R. S. Perdana and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features,” no. October, 2017.

