ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE

by Journal Manager Methodist

Submission date: 02-May-2023 12:40AM (UTC-0600)
Submission ID: 2081823751
File name: Manuscript_SVM_Methomika_Darwis_marioX.docx (191.98K)
Word count: 4131
Character count: 23524

METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi
Vol. 6 No. 2 (Oktober 2022)
ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE

Darwis Robison Manalu[1], Mario Christofell L. Tobing[2], Margaretha Yohanna[3]
[1,2,3] Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Methodist Indonesia
e-mail: manaludarwis@gmail.com

ABSTRACT

Machine learning plays an important role in managing important issues to classify and predict information that develops ahead of the General Election in Indonesia. Especially in knowing public sentiment on the discourse on the postponement of the 2024 election through Twitter social media. So it is necessary to analyze the discourse by categorizing it as positive or negative. The Support Vector Machine (SVM) model is used for analysis and classification. The sample data used were 100 tweets data which were then scraped in the period January 2022 – May 2022. The processing was done using Python programming and Jupyter Lab tools. Before doing the analysis, do preprocess to eliminate unnecessary words and information so that the level of accuracy of the results of this Twitter sentiment classification can provide a closer picture of reality. As for the results of the grouping carried out positive sentiment in as many as 40 data tweets and negative sentiment in as many as 60 data tweets. The classification test results on tweets data with a good level of accuracy of 92%. These results are expected to be a reference for future researchers who want to improve accuracy or analysis results.

Keyword: Support Vector Machine, Metode SVM, Sentimen Twitter, Penundaan Pemilu

ABSTRAK

Machine learning sangat berperan penting dalam pengelolaan isu-isu penting utnuk melakukan klasifikasi maupun prediksi terhadap informasi yang berkembang menjelang Pemilihan Umum di Indonesia. Terutama dalam mengetahui sentimen masyarakat terhadap wacana Penundaan Pemilu 2024 melalui media sosial twitter. Sehingga perlu dilakukan analisis terhadap wacana tersebut dengan mengkategorikan bersifat positif atau negatif. Model Support Vector Machine (SVM) digunakan untuk menganalisis dan klasifikasi. Data sampel yang digunakan 100 data tweets yang kemudian discraping pada periode Januari 2022 – Mei 2022. Pengolahannya dengan pemrograman Python dan tools Jupyter Lab. Sebelum melakukan analisis terlebih dahulu melakukan preprosesing untuk menghilangkan kata maupun informasi yang tidak diperlukan. Sehingga tingkat keakurasian dari hasil klasifikasi sentimen Twitter ini dapat memberikan gambaran yang lebih mendekati pada kenyataannya. Adapun hasil dari pengelompokan yang dilakukan sentimen positif sebanyak 40 data tweets dan sentimen negatif sebanyak 60 data tweets. Hasil pengujian klasifikasi pada data tweets degnan tingkat akurasi yang baik sebesar 92%. Dari hasil ini diharapkan dapat menjadi referensi bagi peneliti berikutnya yang ingin meningkatkatkan akurasi ataupun hasil analisis.

Keyword: Support Vector Machine, Metode SVM, Sentimen Twitter, Penundaan Pemilu

1. PENDAHULUAN

Media sosial Twitter merupakan media yagn digunakan untuk berkomunikasi dalam bentuk teks atau yang biasa dikenal dengan istilah *tweet*, dan dapat juga mengunggah media berupa foto dan video. Twitter memiliki 206 juta pengguna aktif secara global dan indonesia sendiri menyumbang sebesar 18,4 juta pengguna dari total keseluruhan menjadikan Indonesia berada di peringkat 5 negara pengguna Twitter terbesar di dunia[1]. Salah satu isu yang muncul pada trending topic Twitter di awal tahun 2022 adalah tentang "Wacana Penundaan Pemilu 2024". Ada beberapa perdebatan pro dan kontra terjadi di kalangan masyarakat, ada yang beranggapan bahwa wacana tersebut melanggar konstitusi serta dapat membuat situasi perpolitikan dan demokrasi yang tidak sehat. Upaya agar megetahui pendapat masyarakat tentang isu ini perlu dilakukan analisis analisis sentimen melalui media sosial Twitter. Analisis Sentimen merupakan bidang yang berfokus pada pemanfaatan teknik *Natural Language Proccesing* dan *Text Mining*[2]. Adapun tujuannya adalah untuk menentukan dan menganalisis opini dari suatu dokumen, teks atau serangkain kalimat dengan topik tertentu. Analisis sentimen ini dapat menentukan pendapat atau opini dari suatu teks atau dokumen kedalam kelas positif, netral, maupun negatif. Berbagai penelitian bidang analisis sentimen telah dilakukan oleh beberapa peneliti terdahulu dengan tujuan untuk memperoleh informasi dari suatu kumpulan dataset terhadap penilaian subjek yang diteliti[3]. Juga sudah digunakan untuk menganalisa opini masyarakat di Twitter terhadap wacana pemindahan ibu kota Indonesia[4].

Dari uraian sebelumnya tentang analisis sentimen, maka perlu diteliti lebih mendalam agar menghasilkan analisis dan klasifikasi sentimen/opini pengguna Twitter

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

terhadap wacana penundaan pemilu 2024 di Indonesia degnan metode *SVM*. Sehingga peranan dari kecerdasan buatan melalui machine learning dapat membantu para pegnambil keputusan untuk membuat kebijakan terhadap peraturan dan pendapat masyarakat.

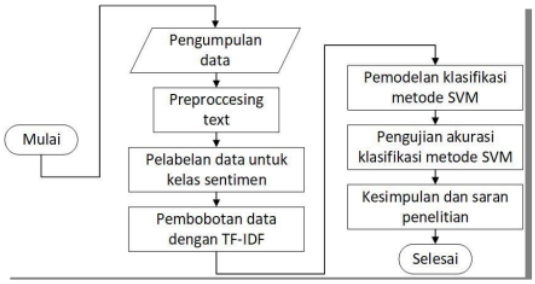
KAJIAN LITERATUR

Beberapa penelitian terkait dengan klasifikasi pada dokumen bahasa indonesia telah dipublikasi seperti penelitian klasifikasi kedalam kelas positif dan negatif menggunakan alogritma *Support Vector Machine*, yang dimana hasil pengujian klasifikasi dengan metode *SVM* didapatkan nilai akurasinya sebesar 96,68%[4]. Pada tahun 2021 dilakukan penelitian untuk menganalisa pendapat masyarakat di Twitter mengenai layanan Telkomsel[5]. Penelitian ini membandingkan penggunaan metode *Unsupervised Learning* yaitu *TextBlob* dengan algoritma *Support Vector Machine* untuk mengklasifikasi sentimen, hasilnya terbukti metode SVM melakukan klasifikasi lebih baik dengan akurasi sebesar 75%. Penelitian lainnya tahun 2022 yaitu dengan menganalisis sentimen opini pengguna Twitter terhadap kinerja dari Gojek/Gofood. Penelitian tersebut menggunakan dua metode klasifikasi yaitu Naïve Bayes dan SVM untuk dibandingkan nilai akurasinya mana yang paling baik, hasil akhirnya didapatkan akurasi metode SVM yang paling baik yakni sebesar 83% sedangkan Naïve Bayes sebesar 74,6%[6]. Penelitan sebeumnya sangat membantu untuk penelitian saat ini karena akan menambah akurasi dan hasil yang lebih maksimal jika karena adanya penambahan parameter pada metode yang digunakan. Teknik yang digunakan adalah teknik dari Natural Languange Proccesing dan Text Mining[7], yang tujuannya adalah menentukan dan menganalisis opini dari suatu dokumen, teks atau serangkain kalimat dengan topik tertentu[8]. Support Vector Machine adalah salah satu metode klasifikasi supervised learning pada machine learning[9][10]. Metode ini awalnya memiliki prinsip linear yang kemudian sudah berkembang menjadi dapat bekerja pada masalah non-linear dengan memasukkan konsep kernel pada ruang yang berdimensi tinggi, tugas atau tujuan utama dari metode ini adalah mencari hyperplane (pembatas) dan memaksimalkan jaraknya antar kedua kelas yang diberi label[11].

METODE PENELITIAN

2.1. Alur Proses Penelitian

Dalam melaksanakan penelitian ini dirancang suatu alur proses tahapan penelitian agar sesuai dan terstruktur, *flowchart* alur proses penelitian dapat dilihat seperti pada Gambar 1.



Gambar 1. Alur Proses Penelitian

Penarikan Data

Penarikan data *tweet* adalah menggunakan teknik *scraping* dengan *tools snscrape* pada *library* bahasa pemrograman *python*, data *tweet* yang diambil adalah yang berbahasa Indonesia dengan kata kunci "Tunda Pemilu". Dataset disimpan dalam format *xlsx* yang kemudian nantinya akan diproses pada tahapan berikutnya.

Preproccesing Text

Dataset yang telah terkumpul sebelumnya merupakan data yang belum terstruktur dengan baik, sehingga diperlukan tahapan penting pada bidang *text mining* yang disebut dengan istilah *preproccesing text* dimana bermaksud mengolah data awal yang masih belum terstruktur dengan baik untuk dijadikan sebuah data yang terstruktur agar dapat dikenali dan diterapkan kedalam beberapa metode *text mining* yang ada, adapun keempat langkah utama dalam *preproccesing* biasa disebut dengan istilah *case folding*, *tokenizing*, *stop-word removal* dan *stemming*[12].

Pelabelan Data

Pelabelan data *tweet* pada penelitian ini hanya membagi kedalam kelas positif dan negatif dimana parameter nilai label positif adalah kalimat yang mengandung kata-kata pujian,dukungan dan lainnya sedangkan label negatif adalah kalimat yang mengandung kata-kata hinaan, cacian dan ejekan. Adapun ketentuan dalam pemberian label pada kalimat data *tweet* adalah dengan mengamati keseluruhan kalimat, jika kata-kata positif lebih mendominasi daripada kata negatif maka kalimat tersebut diberikan label positif (1) dan jika sebaliknya diberikan label negatif (-1).

Pembobotan Data

Penerapan metode *TF-ID* (*Term Frequency-Inverse Document Frequency*) pada penelitian ini adalah bertujuan dalam pembobotan data yang mengubah suatu kata dalam dokumen kedalam bentuk numerik, dimana tugas utama dari algoritma *TF-IDF* adalah sebagai pencarian seberapa sering kemunculan suatu kata dalam dokumen/dataset[13]. Rumus algoritma *TF-IDF* seperti pada persamaan (1) berikut :

$$W_{dt} = tf_{dt} * IDF_{dt} \quad (1)$$

Ket :

W_{dt} = nilai dokumen ke…d pada kata ke…t

tf_{dt} = banyak jumlah kata dalam dokumen

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

$IDF\ td = log(\frac{n}{df})$

df= banyak jumlah dokumen pada kata yang dicari
n = banyak jumlah dokumen

Pemodelan Klasifikasi Support Vector Machine (SVM)

Sebelum proses klasifikasi menggunakan metode *SVM* pembagian dataset dibagi menjadi 2 yaitu dengan skenario 80% data *training* dan 20% data *testing*, data *training* berguna untuk melatih dan membangun model dari metode *SVM* sedangkan data *testing* untuk sebagai data pengujian yang memberikan hasil klasifikasi kedalam kelas positif dan negatif berdasarakan model yang telah dibentuk dari data *training* sebelumnya.

Metode *SVM* merupakan salah satu metode *Supervised Learning* dari bidang *Machine Learning* yang umum digunakan metode ini awalnya memiliki prinsip linear yang kemudian sudah berkembang menjadi dapat bekerja pada masalah non-linear dengan memasukkan konsep kernel pada ruang yang berdimensi tinggi, tugas atau tujuan utama dari metode ini adalah mencari hyperplane (pembatas) dan memaksimalkan jaraknya (margin) antar kedua kelas yang diberi label [10].

Adapun tahapan dalam perhitungan metode *Support Vector Machine* dengan menggunakan fungsi kernel adalah dengan beberapa langkah sebagai berikut :

- Menentukan nilai α = 0.5, C = 1, λ = 0.5, gamma = 0.5 dan epsilon = 0.001
- Menghitung matriks dengan persamaan (2) :

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

Ket :

D_{ij} : matriks data ke…ij
y_i : label data ke…i
y_j : label data ke…j
λ : lamda
$K(x_i \cdot x_j)$: fungsi kernel

- Persamaan (3) untuk perhitungan nilai error :

$$E_i = \sum^n = 1\ \alpha_i\ D_{ij} \quad (3)$$

- Perhitungan untuk nilai delta alpha dengan persamaan (4) :

$$\delta\alpha i = \min\{max[\gamma(1 - Ei), -\alpha i], C - \alpha i\} \quad (4)$$

Keterangan :

Ei : nilai error pada data ke…i
γ : tingkat pembelajaran
max(i) Dij : nilai maksimum matriks hessian

- Menghitung nilai alpha baru dengan persamaan (5):

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (5)$$

- Persamaan (6) untuk mencari nilai bias (*b*) :

$$b = -\frac{1}{2}\ [W * x^+ + W * x^-] \quad (6)$$

- Pengujian terhadap data uji
- Perhitungan keputusan dengan menggunakan persamaan (7), penentuan kelas positif adalah jika hasil perhitungan keputusan memiliki nilai lebih besar sama dengan 0 maka *sign h(x)* adalah 1, sedangkan jika hasil perhitungan keputusan memiliki nilai lebih kecil dari 0 maka *sign h(x)* adalah -1.

$$h(x) = \sum_{i=1}^{m} \alpha_i\ y_i\ K(x, x_i) + b \quad (7)$$

Pengujian dan Evaluasi Model Klasifikasi *SVM*

Pengujian dan evaluasi model klasifikasi *SVM* pada penelitian ini menggunakan metode *confusion matrix* untuk mengukur hasil perbandingan dari klasifikasi yang dilakukan oleh sistem (model) dengan klasifikasi yang sebenarnya [11]. Dalam *confusion matrix* terdapat berbagai performance matrix yang umum digunakan untuk mengukur kinerja sistem (model) seperti *accuracy*, *precision*,*recall* dan *f1-score*. Adapun untuk menghitung akurasi adalah dengan persamaan (8) :

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

PEMBAHASAN

Pengumpulan Data Tweet

Data *tweet* diambil dari media sosial Twitter dengan teknik *scraping* menggunakan *tools snscrape* pada pemrograman *python* menggunakan kata kunci "Tunda Pemilu" dari periode bulan Januari 2022 – Mei 2022. Data *tweet* yang terkumpul pada penelitian ini adalah sebanyak 3681 data, dikarenakan adanya keterbatasan pada penelitian ini sehingga data yang digunakan untuk proses tahapan selanjutnya adalah hanya sebanyak 100 sampel data *tweet*. Sampel data *tweet* dapat dilihat pada Tabel 1.

Tabel 1. Sampel Data Tweet

No.	Tweet
1	PKS Tolak Wacana Tunda Pemilu 2024 https://t.co/vr2PUEpX8r
2	PKS Tegaskan Tolak Usulan Tunda Pemilu dan Perpanjangan Masa Jabatan Presiden https://t.co/64Ae4FOqGX
3	Muncul Ide Tunda Pemilu, KSPSI: Jangan Lempar Wacana yang Cuma Buat Gaduh, Buruh Sudah Pasti Menolak https://t.co/KbVhMxxTgD lewat @wartakotalive
4	BAWA ASPIRASI RAKYAT..! PKB ke Megawati: Izinkan Dengan Segala Hormat Kami Lanjutkan Wacana Tunda Pemilu! https://t.co/JrhyWfCITY
5	Ga punya stok wacana yg lebih cerdas mas? PKB ke Megawati: Izinkan dengan Segala Hormat Kami Lanjutkan Wacana Tunda Pemilu - https://t.co/rJfBvFtMYA

Preproccesing Data

Setelah data *tweet* telah diperoleh tahapan selanjutnya merupakan *preproccesing* yang tujuannya untuk

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

menghasilkan data yang bersih dan terstruktur dengan baik, dimana langkah awalnya adalah :

- *Case Folding* : berfungsi untuk mengubah seluruh jenis huruf *uppercase* (huruf besar) menjadi *lowercase* (huruf kecil), dan juga pada tahap ini dilakukan pembersihan terhadap *emoticon*, *url*, *hashtag*, angka, *RT*, *tag/add* '@' dan tanda baca.
- *Tokenizing* : untuk memecah kalimat pada dataset menjadi per-kata yang disebut dengan *tokens*
- *Stop-word Removal* : untuk menghapus kata-kata umum, imbuhan yang tidak memiliki arti/sentimen dan juga menghapus kata-kata singkat/gaul yang tidak terdapat dalam kamus bahasa Indonesia seperti kata "di","yang","ada" dan lainnya
- *Stemming* : berfungsi untuk mengembalikan suatu kata kedalam bentuk akarnya (*root*)

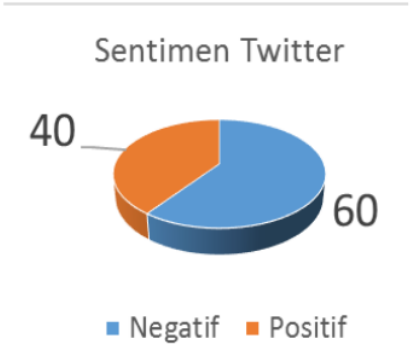
Adapun sampel data hasil *preproccesing* dapat dilihat dalam Tabel 2.

Tabel 2. Sampel Data Tweet Hasil Preproccesing

No.	Tweet
1	tolak wacana tunda
2	tegas tolak usul tunda presiden
3	muncul ide tunda lempar wacana gaduh buruh tolak
4	bawa aspirasi rakyat izin hormat lanjut wacana tunda
5	wacana cerdas izin hormat lanjut wacana tunda

Pelabelan Data Tweet

Pelabelan 100 sampel data *tweet* dilakukan secara manual dalam penelitian ini, dengan mengamati keseluruhan kalimat dengan memperhatikan tiap sentimen kata yang ada. Pelabelan kelas hanya dibagi kedalam dua kelas yaitu sentimen positif (1) dan negatif (-1), hasil keseluruhan pelabelan data *tweet* dalam penelitian ini adalah 40 data dilabeli bersentimen positif dan 60 data dilabeli bersentimen negatif. Gambar 2 merupakan diagram lingkaran/pie hasil pelabelan manual data *tweet* sebanyak 100 sampel.



Gambar 2. Diagram Pelabelan Manual Data *Tweet*

Pembobotan TF-IDF

Penggunaan algoritma *TF-IDF* diperlukan untuk memperoleh bobot nilai terhadap satu dokumen, berupa bentuk kalimat untuk dirubah menjadi dalam bentuk numerik/angka agar dapat diproses ptahapan selanjutnya yaitu pengklasifikasian. Sampel data *tweet* yang digunakan untuk perhitungan *TF-IDF* dengan rumus (1) adalah 5 sampel data *tweet* hasil dari *preproccesing* sebelumnya pada Tabel 2.

Tabel 3. Hasil Pembobotan *TF-IDF*

T	TF-IDF				
	D1	D2	D3	D4	D5
T1	0.204	0.204	0.204	0	0
T2	0.096	0	0.096	0.096	0.192
T3	0	0	0	0	0
T4	0	0.698	0	0	0
T5	0	0.698	0	0	0
T6	0	0.698	0	0	0
T7	0	0	0.698	0	0
T8	0	0	0.698	0	0
T9	0	0	0.698	0	0
T10	0	0	0.698	0	0
T11	0	0	0.698	0	0
T12	0	0	0	0.698	0
T13	0	0	0	0.698	0
T14	0	0	0	0.698	0
T15	0	0	0	0.397	0.397
T16	0	0	0	0.397	0.397
T17	0	0	0	0.397	0.397
T18	0	0	0	0	0.698
	y				
	-1	-1	-1	1	1

Tabel 3 menunjukkan terdapat variabel T yang merupakan sebagai istilah dari *term* atau kumpulan kata dari kelima data *tweet* hasil *preprocessing* diatas yang terdapat sebanyak 18 *term*/kata sedangkan variabel y merupakan label sentimen dari kelima data *tweet* tersebut.

Klasifikasi Metode SVM

Sebelum melakukan klasifikasi dengan metode *SVM* ditentukan terlebih dahulu pembagian dataset yaitu dengan skenario 80% data *training*/latih (80 data) dan 20% data *testing*/uji (20 data). Pengklasifikasian dengan metode *SVM* adalah dengan memanfaatkan hasil dari vektorisasi atau pembobotan kata yang telah dilakukan pada tahapan proses *TF-IDF* terhadap 5 sampel data *tweet* sebelumnya. Dalam penelitian ini proses klasifikasi *SVM* menggunakan fungsi kernel, yang dimana setiap data akan dihitung terhadap data itu sendiri dan antara data lainnya, fungsi kernel dapat dilihat pada Tabel 4.

Tabel 4. Fungsi Kernel

	D1	D2	D3	D4	D5

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

D 1	K(D1, D1)	K(D1, D2)	K(D1, D3)	K(D1, D4)	K(D1, D5)
D 2	K(D2, D1)	K(D2, D2)	K(D2, D3)	K(D2, D4)	K(D2, D5)
D 3	K(D3, D1)	K(D3, D2)	K(D3, D3)	K(D3, D4)	K(D3, D5)
D 4	K(D4, D1)	K(D4, D2)	K(D4, D3)	K(D4, D4)	K(D4, D5)
D 5	K(D5, D1)	K(D5, D2)	K(D5, D3)	K(D5, D4)	K(D5, D5)

Perhitungan data pada fungsi kernel adalah sebagai berikut :

K(D1,D1) = (0.204 × 0.204)+(0.096 × 0.096)+(0 × 0)+(0 × 0)+(0 × 0)+(0 × 0)+(0 × 0)+ (0 × 0)+(0 × 0)+(0 × 0)+(0 × 0)+(0 × 0)+(0 × 0)+(0 × 0) +(0 × 0)+(0 × 0) +(0 × 0)+(0 × 0) +(0 × 0) = 0.050

Seluruh data dihitung terhadap data itu sendiri dan dengan data yang lain, yang dimana pada Tabel 4 diatas terdapat sebanyak 5 data yang kemudian hasil akhir keseluruhan perhitungan fungsi kernelnya menghasilkan matriks ordo 5x5 yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Nilai Fungsi Kernel

	D1	**D2**	**D3**	**D4**	**D5**
D1	0.050	0.041	0.050	0.009	0.018
D2	0.041	1.503	0.041	0	0
D3	0.050	0.041	2.486	0.009	0.018
D4	0.009	0	0.009	1.943	0.491
D5	0.018	0	0.018	0.491	0.996

Langkah berikutnya setelah didapatkan hasil seluruh perhitungan fungsi kernel adalah melakukan perhitungan matriks dengan persamaan (2) sebagai berikut :

(D1,D1) = (-1)(-1)(0.050)+0.25 = 0.3

Perhitungan dilakukan pada setiap data seterusnya yang hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Nilai Matriks

	D1	**D2**	**D3**	**D4**	**D5**
D1	0.3	0.291	0.3	0.241	0.232
D2	0.291	1.753	0.291	0.25	0.25
D3	0.3	0.291	2.736	0.241	0.232
D4	0.241	0.25	0.241	2.193	0.741
D5	0.232	0.25	0.232	0.741	1.246

Selanjutnya adalah mencari nilai *error* menggunakan persamaan (3) dengan perhitungannya sebagai berikut :

D1 = (1)(0.5)(1.364) = 0.682

Hasil keseluruhan perhitungan *error* setiap data dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perhitungan Nilai *Error*

Dokumen	**Error**
D1	0.682
D2	1.4
D3	1.9
D4	1.833
D5	1.350

Berikutnya adalah tahapan mencari nilai dari *delta alpha* menggunakan persamaan (4), berikut adalah perhitungannya :

D1 = $min\{max\ [0.5(1-0.682), -0.5], 1-0.5\}$
= $min\{max\ [0.159, -0.5], 0.5\}$
= $min\ [0.159, 0.5]$
= 0.159

Adapun hasil perhitungan data lainnya dapat dilihat pada Tabel 8.

Tabel 8. Hasil Perhitungan Nilai Delta Alpha

Dokumen	**Delta Alpha**
D1	0.159
D2	− 0.2
D3	− 0.45
D4	− 0.416
D5	− 0.175

Selanjutnya adalah tahapan menghitung nilai *alpha* baru dengan persamaan (5) sebagai berikut :

D1 = 0.5 + 0.159 = 0.659

Tabel 9 merupakan hasil perhitungan data lainnya.

Tabel 9. Hasil Perhitungan *Alpha* Baru

Dokumen	**Alpha Baru**
D1	0.659
D2	0.3
D3	0.05
D4	0.084
D5	0.325

Sebelum masuk pada tahapan pencarian nilai bias terlebih dahulu menentukan nilai W dimana :

W^+ adalah bobot *dot product* dengan nilai *alpha* terbesar pada kelas positif

W^- adalah bobot *dot product* dengan nilai *alpha* terbesar pada kelas negatif

perhitungannya adalah sebagai berikut :

W * x^+ = (0.659 × (-1) × 0.018) + (0.3 × (-1) × 0) + (0.05 × (-1) × 0.018) + (0.084 × 1 × 0.491) + (0.325 × 1 × 0.996) = 0.352

W * x^- = (0.659 × (-1) × 0.050) + (0.3 × (-1) × 0.041) + (0.05 × (-1) × 0.050) + (0.084 × 1 × 0.009) + (0.325 × 1 × 0.018) = − 0.041

Setelah nilai *dot product* kelas positif dan negatif telah didapatkan, maka selanjutnya mencari nilai bias (b) dengan menggunakan persamaan (6) perhitungannya sebagai berikut :

$b = -\frac{1}{2}(0.352 + (-0.041))$

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

$b = -0.155$

Setalah didapatkan nilai α, w, dan b maka berikutnya dapat dilakukan pengujian pada sampel data uji dengan kalimat sebagai berikut :
Dx = "izin hormat lanjut wacana tunda"
Saat dipecah kalimat dari data uji mengandung *term*/kata 'izin', 'hormat', 'lanjut', 'wacana' dan 'tunda' dan dimasukkan kedalam bentuk *vector* untuk proses klasifikasi yang dapat dilihat pada Tabel 10.

Tabel 10. Sampel Data Uji

D	TF-IDF									y
	T1	T2	T3	T4	T5	T6	T7	T8	T9	
	0	0.07	0	0	0	0	0	0	0	
Dx	T10	T11	T12	T13	T14	T15	T16	T17	T18	?
	0	0	0	0	0	0.3	0.3	0.3	0	

Langkah awal dalam menguji adalah dengan menghitung *dot product* data uji dengan semua data latih menggunakan fungsi kernel dengan perhitungannya adalah sebagai berikut :
K(x,y) = x,y
dimana x = data uji
y = data latih
Berikut perhitungannya :

$$D1 = (0 \times 0.204)+(0.079 \times 0.096)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)+(0 \times 0)\ (0 \times 0)+(0.301 \times 0)+(0.301 \times 0)+(0.301 \times 0)+(0 \times 0) = 0.007$$

Hasil keseluruhan perhitungan *dot product* antara data uji dengan data latih dapat dilihat pada Tabel 11.

Tabel 11. Hasil Perhitungan Dot Product Data Uji dengan Data Latih

Dokumen	Dot Product
D1	0.007
D2	0
D3	0.007
D4	0.366
D5	0.373

Tahapan berikutnya adalah melakukan perhitungan fungsi keputusan menggunakan persamaan (7) dengan perhitungannya sebagai berikut :

$$h(x) = sign((0.659 \times (-1) \times 0.007) + (-0.155) + (0.3 \times (-1) \times 0) + (-0.155) + (0.05 \times (-1) \times 0.007) + (-0.155) + (0.084 \times 1 \times 0.366) + (-0.155) + (0.325 \times 1 \times 0.373) + (-0.155))$$
$$= -0.627$$
$$= sign(-0.627) = -1$$

Berdasarkan dari perhitungan fungsi keputusan terhadap data uji maka nilai *h(x) sign* data uji tersebut diklasifikasi kedalam kelas (− 1).

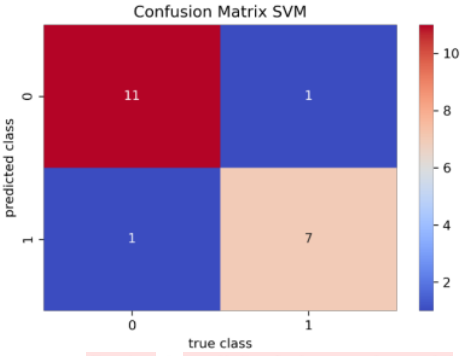
Pengujian dan Evaluasi Model SVM

Pengujian model SVM dilakukan dengan tujuan untuk mengetahui seberapa baik performa model *SVM* yang telah dibentuk sebelumnya pada 80% data training (80 data) dalam mengklasifikasi terhadap 20% data uji (20 data). Pengujian model SVM pada sistem yang dibangun adalah menggunakan confusion matrix yang dimana didapatkan hasil akurasinya sebesar 90% seperti pada Gambar 3.

```
SVM Accuracy: 0.9
SVM Precision: 0.9166666666666666
SVM Recall: 0.9166666666666666
SVM f1_score: 0.9166666666666666
confusion matrix:
 [[11  1]
 [ 1  7]]
```

Gambar 3. Hasil Pengujian Model SVM

Langkah terakhir adalah melakukan evaluasi terhadap model SVM dengan maksud untuk meninjau atau mengukur hasil pengklasifikasian dari model yang telah dibentuk dengan klasifikasi yang sebenarnya. Evaluasi dilakukan dari hasil *confusion matrix* bentuk ordo 2 x 2 yang telah diperoleh sebelumnya, seperti yang dapat dilihat lebih jelas pada Gambar 4.



Gambar 4. Hasil Confusion Matrix Model SVM

Gambar 4 merupakan hasil perbandingan informasi klasifikasi model *SVM* (*Support Vector Machine*) [14] oleh sistem menggunakan metode *confusion matrix*, dengan penjelasannya adalah sebagai berikut :

- Pada bagian yang terdapat angka 11 adalah merupakan hasil dari *True Positive* (TP) yang mengklasifikasi data kedalam kelas positif dengan tepat.
- Pada bagian yang terdapat angka 7 adalah merupakan hasil dari *True Negative* (TN) yang mengklasifikasi data kedalam kelas negatif dengan tepat.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online

- Pada kedua bagian yang terdapat angka 1 merupakan hasil dari *False Positive* (FP) yang mengklasifikasi data tidak tepat kedalam kelas positif yang kelas sebenarnya adalah negatif dan *False Negative* (FN) yang mengklasifikasi data tidak tepat kedalam kelas negatif yang kelas sebenarnya adalah positif.

Pada pengujian akurasi model *SVM* sebelumnya oleh sistem diperoleh nilai akurasi sebesar 90%, untuk dapat mengukur kebenaran dari hasil akurasi model *SVM* tersebut maka dilakukan perhitungan akurasi secara manual menggunakan persamaan (8) sebagai berikut :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$= \frac{11+7}{11+7+1+1}$$
$$= \frac{18}{20} \times 100\%$$
$$= 90\%$$

Berdasarkan perhitungan secara manual yang dilakukan didapatkan hasil yang sama tepat dengan perhitungan melalui sistem, nilai akurasi dari suatu model perlu diketahui agar dapat dinilai seberapa tinggi kemampuan sistem atau model tersebut dalam mengklasifikasi sebuah data dengan benar. Model klasifikasi *SVM* yang dibangun pada penelitian ini terbukti dapat mengklasifikasi data *tweet* dengan baik yakni dengan tingkat keakurasiannya sebesar 90%.

KESIMPULAN

Dari penelitian dan pengujian yang dilakukan dengan metode *Support Vector Machine* pada data *tweet* sebanyak 100 sampel, terbukti metode ini dapat menghasilkan klasifikasi pada tingkat akurasi sebesar 90%. Hasil pelabelan sentimen dengan manual yang dilakukan oleh peneliti pada 100 sampel data *tweet*, diperoleh sentimen positif sebanyak 40 data *tweet* dan sentimen negatif sebanyak 60 data *tweet*. Maka respon dari masyarakat terhadap opini ini cenderung negatif ta kurang sependapat dengan wacana penundaan pelaksanaan pemilu pada tahun 2024. Agar hasil penelitian ini dapat meningkat pada masa berikutnya perlu dilakukan penambahan parameter ataupun penambahan pada proses pembentukan model klasifikasi. Serta meningkatkan kualitas preprosesing supaya mudah mengenali *slang word* (kata gaul) dan *short form* (kata singkat) yang belum masuk di kamus bahasa Indonesia

REFERENSI

[1] Databooks, "Mayoritas Warga Tidak Mau Tunda Pemilu 2024," *LSI*, 2022. databoks.katadata.co.id

[2] P. Pasek, O. Mahawardana, G. Arya, I. P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap ' Figure Pemimpin ' Menggunakan Python," *JITTER-Jurnal Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 810–820, 2022, [Online]. Available: https://ojs.unud.ac.id/index.php/jitter/article/view/82975

[3] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 2, pp. 195–200, 2022, doi: 10.32736/sisfokom.v11i2.1412.

[4] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[5] Kevin Perdana, Titania Pricillia, and Zulfachmi, "Optimasi Textblob Menggunakan Support Vector Machine Untuk Analisis Sentimen (Studi Kasus Layanan Telkomsel)," *J. Bangkit Indones.*, vol. 10, no. 1, pp. 13–15, 2021, doi: 10.52771/bangkitindonesia.v10i1.120.

[6] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022, doi: 10.30865/mib.v6i1.3530.

[7] S. Khomsah and Agus Sasmito Aribowo, "Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *Rekayasa Sist. dan Teknol. Informasi, RESTI*, vol. 4, no. 10, pp. 648–654, 2020.

[8] D. R. Manalu and E. Rajagukguk, "The Development of Document Similarity Detector by Jaccard Formulation," no. 1, pp. 1–3.

[9] D. R. Manalu, M. Zarlis, H. Mawengkang, and O. S. Sitompul, "Forest Fire Prediction in Northern Sumatera using Support Vector Machine Based on the Fire Weather Index," *AIRCC Publ. Corp.*, vol. 10, no. 19, pp. 187–196, 2020, doi: 10.5121/csit.2020.101915.

[10] I. J. Song and W. Heo, "Improving insurers' loss reserve error prediction: Adopting combined unsupervised-supervised machine learning techniques in risk management," *J. Financ. Data Sci.*, vol. 8, pp. 233–254, 2022, doi: 10.1016/j.jfds.2022.09.003.

[11] 荒牧英治, 今井健, 美代賢吾, and 大江和彦, "Support Vector Machine を用いた医学用語の表記ゆれ解消," 言語処理学会, pp. 135–138, 2008.

[12] K. S. Nugroho, "Dasar Text Preprocessing dengan Python," 2019. https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe

[13] M. Nurjannah and I. Fitri Astuti, "PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman," *J. Inform. Mulawarman*, vol. 8, no. 3, pp. 110–113, 2013.

[14] Herdiawan, "Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved K-Nearest Neighbor," *J. Ilm. Komput. dan Inform.* , 2016, [Online]. Available: https://repository.unikom.ac.id/id/eprint/15556

ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE

ORIGINALITY REPORT

90%	90%	13%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejurnal.methodist.ac.id Internet Source	78%
2	Submitted to Universitas Jenderal Soedirman Student Paper	3%
3	ojs.stmik-banjarbaru.ac.id Internet Source	2%
4	ejurnal.stmik-budidarma.ac.id Internet Source	1%
5	ejournal.upnjatim.ac.id Internet Source	1%
6	ijcs.stmikindonesia.ac.id Internet Source	1%
7	www.morris.umn.edu Internet Source	1%
8	bibdigital.epn.edu.ec Internet Source	1%
	ejournal.upnvj.ac.id	

9	Internet Source	<1%
10	nero.trunojoyo.ac.id Internet Source	<1%
11	pdfs.semanticscholar.org Internet Source	<1%
12	Submitted to Universitas Amikom Student Paper	<1%
13	www.methodist.ac.id:8082 Internet Source	<1%
14	jsi.politala.ac.id Internet Source	<1%
15	jsi.stikom-bali.ac.id Internet Source	<1%
16	jurnal.stmik-amik-riau.ac.id Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE