

SENTIMENT ANALYSIS PEMILIHAN CALON PRESIDEN 2024 MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

[SENTIMENT ANALYSIS 2024 PRESIDENTIAL CANDIDATE ELECTION USING SUPPORT VECTOR MACHINE ALGORITHM]

Michael Alfonso, s32190039@student.ubm.ac.id¹⁾, Dionisia Bhisetya Rarasati, S.Kom.,
M.T.I, i1783@lecturer.ubm.ac.id²⁾

¹⁾ Program Studi Informatika, Universitas Bunda Mulia

²⁾ Program Studi Informatika, Universitas Bunda Mulia

Disetujui 15 Agustus 2023

ABSTRACT

*The election of presidential candidates is held every 5 years with various candidates running for office. On social media Twitter there are often arguments about political matters which are the topic of discussion regarding the election of presidential candidates. Thus, this research focuses on the sentiment analysis process of Twitter users responses to the election of presidential candidates followed by validation by looking for correlations between electability survey results and tweets using Pearson Correlation. Method used to search for the best model using 10-Fold Cross Validation, then data weighting using TF-IDF and validating the best model using Confusion Matrix to get the closest *f1-score* to the average *f1-score* across all folds. The algorithm used to create the model is the Support vector machine algorithm with a Gaussian RBF (Radial Basis Function) kernel. The results of analysis were compared with the news portal electability survey containing the 3 candidates using Pearson Correlation. Correlation for candidates Anies Baswedan and Prabowo Subianto, the higher the positive sentiment, the higher the electability survey data, while for candidate Ganjar Pranowo, the lower the positive sentiment, the higher the electability survey data.*

Keywords: NLP, Pearson Correlation, Sentiment analysis, SVM, TF-IDF

ABSTRAK

Pemilihan calon presiden dilaksanakan setiap 5 tahun dengan berbagai kandidat yang mencalonkan diri, pada media sosial Twitter sering terjadi argumen seputar hal-hal politik yang menjadi topik diskusi tentang pemilihan calon presiden ini. Sehingga, penelitian ini berfokus pada proses Analisa sentimen terhadap respon pengguna twitter terhadap pemilihan calon presiden dilanjutkan dengan melakukan validasi dengan mencari korelasi antara hasil survei elektabilitas dan tweet dengan menggunakan Korelasi Pearson. Pencarian model terbaik menggunakan 10-Fold Cross Validation, lalu pembobotan menggunakan TF-IDF serta melakukan validasi dari model terbaik menggunakan Confusion Matrix untuk mendapat *f1-score* terdekat dari rata-rata *f1-score* keseluruhan *fold*. Algoritma yang digunakan untuk membuat model adalah algoritma Support vector machine dengan kernel Gaussian RBF (Radial Basis Function). Hasil analisa dibandingkan dengan hasil survey elektabilitas portal berita yang memuat 3 calon tersebut dengan menggunakan Korelasi Pearson. Pada Korelasi Pearson, pada calon Anies Baswedan dan Prabowo subianto, semakin tinggi sentiment positif, maka semakin tinggi juga data survei elektabilitas, sedangkan untuk calon Ganjar Pranowo, semakin rendah sentiment positif, maka semakin tinggi data survei elektabilitas.

Kata Kunci: NLP, Pearson Correlation, Sentiment analysis, SVM, TF-IDF

*Korespondensi Penulis:

E-mail: s32190039@student.ubm.ac.id

PENDAHULUAN

Pemilu terjadi setiap 5 tahun sekali dengan jumlah pemilih yang semakin bertambah setiap periodenya, penambahan pemilih tetap sejak tahun 2004 sampai 2019 terhitung sebanyak 28,90%. Media sosial seperti Facebook, Twitter, LinkedIn, YouTube sudah menjadi bagian dari kehidupan masyarakat kita, hal ini dibuktikan dengan cara berinteraksi kebanyakan dapat dilakukan dengan sosial media, sampai menjadi identitas diri ke masyarakat umum [1].

Pada sosial media Twitter, terjadi penyebaran informasi, hiburan, bahkan opini politik, salah satu informasi yang tersebar adalah terjadinya pemilu pada tahun 2024 yang disebut sebagai pesta demokrasi. Cara penyampaian opini melalui sosial media Twitter adalah dengan menggunakan tanda pagar tertentu pada tweet yang dipost di Twitter. Twitter digunakan karena kemudahan dalam membalas *tweet* dengan *like*, komentar bahkan *reply*, sehingga topik tersebut dapat dengan mudah tersebar [3].

Kerap kali sosial media didominasi oleh banyak masyarakat yang tidak memiliki pendapat yang kurang didengar lalu dituangkan aspirasi tersebut dalam sosial media, sosial media dapat dikategorikan sebagai kumpulan informasi dan dijadikan pengetahuan politik oleh masyarakat Indonesia [2] khususnya Twitter tentang hak pilih mereka terhadap pemilihan calon presiden 2024. Sebab itu, diperlukannya analisa lebih lanjut untuk mengetahui kesetujuan dan ketidaksetujuan pengguna Twitter terhadap pemilihan calon presiden 2024.

Sentiment analysis atau dapat disebut Opinion Mining merupakan studi yang mempelajari opini, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas [1], [3]. Disebutkan juga oleh sumber lain bahwa analisis sentimen adalah sebuah cara yang dilakukan untuk mendapatkan Opini Publik [4]. Sentiment analysis ini digunakan untuk mengetahui tweet yang diposting mengandung kata positif, negatif atau netral [5].

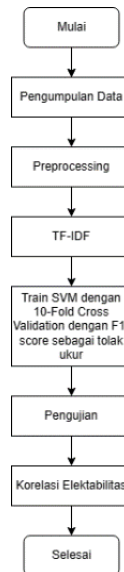
Penelitian ini menggunakan data yang diperoleh dari tweet pengguna Twitter yang berkaitan dengan topik pemilihan umum 2024. Penelitian ini menggunakan metode klasifikasi *supervised* yaitu Support vector machine (SVM). Klasifikasi dilakukan dengan membuat garis pemisah (hyperlane) antara kelas positif dan negatif [1] dengan bantuan garis pemisah yang dibentuk melalui proses pemilihan data yang disebut *support vector* sehingga membentuk garis pemisah terbaik dari sebuah dataset [6].

Sehingga hasil dari penelitian ini adalah sentiment pengguna Twitter terhadap pemilihan calon presiden pada tahun 2024 dengan hasil korelasi antara setiap calon presiden dengan survei elektabilitas portal berita.

Dengan adanya system ini, diharapkan kita sebagai masyarakat yang kurang aktif dalam sosial media, dapat mengetahui ringkasan sentiment pengguna Twitter terhadap pemilihan calon presiden 2024.

METODOLOGI PENELITIAN

Penelitian ini terdiri dari 6 proses secara garis besar yaitu pengumpulan data, preprocessing, pembobotan Term Frequency-Inverse Document Frequency, train Support Vector Machine menggunakan 10-Fold Cross Validation, Memilih fold terbaik untuk dijadikan model, dan pengujian seperti pada Gambar 1.



Gambar 1. Flowchart Penelitian

A. Pengumpulan Data

Tahapan awal dari penelitian ini adalah melakukan scraping data bersumber dari Twitter dengan kata kunci “(Nama Calon) Presiden” untuk mempersempit konteks dari Tweet pengguna Twitter. Proses pengumpulan data dilakukan dengan bantuan library snsrape oleh python. Data dikumpulkan mulai dari tanggal 1 Januari 2023 sampai 31 Maret 2023. Proses scraping ini menghasilkan total 18.267 data dengan 7.810 data yang duplikat dan tidak relevan dengan tema “Pemilihan Calon Presiden 2024” sehingga menghasilkan total 10.457 data bersih dari 3 calon presiden.

B. Preprocessing

Tahapan pre-processing berfungsi untuk membersihkan dan mentransformasi data sehingga dapat diproses oleh sistem dengan mudah dan akurat. Beberapa tahapan pre-processing pada penelitian ini diantaranya:

1. Lowercasing

Lowercasing merupakan tahapan untuk merubah semua huruf menjadi huruf kecil sehingga kata yang diproses menjadi seragam. Tabel 1 menampilkan contoh proses lowercasing dari beberapa tweet Anies Baswedan:

Tabel 1. Tabel Hasil Lowercasing

No	Kalimat Awal	Hasil Proses
1	PKS Mendukung Anies Baswedan Menjadi Calon Presiden 2024	pks mendukung anies baswedan menjadi calon presiden 2024
2	Anies Baswedan Presiden Indonesia	anies baswedan presiden indonesia
3	Walau sekarang didukung Demokrat dan PKS, ada empat	walau sekarang didukung demokrat dan pks, ada empat

	alasan	Anies	alasan	anies
	Baswedan	gak	baswedan	gak
	bakal	bisa	bakal	bisa
	menang	jadi	menang	jadi
	presiden !!		presiden !!	
4	Anies Baswedan		anies baswedan	
	lh yg pantas jadi		lh yg pantas jadi	
	presiden		presiden	
5	Pak	Anies	pak	anies
	Baswedan calon		baswedan calon	
	presiden 2024		presiden 2024	

2. Remove Special Character and Number

Tahapan ini berfungsi menghasilkan kalimat tanpa tanda baca, emoticon, dan angka. Tabel 2 menampilkan contoh proses Remove Special Character and Number:

Tabel 2. Tabel Hasil Remove Special Character and Number

No	Kalimat Awal	Hasil Proses
1	pks mendukung anies baswedan menjadi calon presiden 2024	pks mendukung anies baswedan menjadi calon presiden
2	anies baswedan presiden indonesia	anies baswedan presiden indonesia
3	walau sekarang didukung demokrat dan pks, ada empat alasan anies baswedan gak bakal bisa menang jadi presiden !!	walau sekarang didukung demokrat dan pks ada empat alasan anies baswedan gak bakal bisa menang jadi presiden
4	anies baswedan lh yg pantas jadi presiden	anies baswedan lh yg pantas jadi presiden
5	pak anies baswedan calon presiden 2024	pak anies baswedan calon presiden

3. Stemming

Tahap ini, semua kata berimbuhan akan diproses menjadi kata dasar, seperti ditampilkan pada tabel 3 berikut:

Tabel 3. Tabel Hasil Stemming

No	Kalimat Awal	Hasil Proses
1	pks mendukung anies baswedan	pks dukung anies baswedan

	menjadi	calon	jadi	calon
	presiden		presiden	
2	anies	baswedan	anies	baswedan
	presiden		presiden	
	indonesia		indonesia	
3	walau	sekarang	walau	sekarang
	didukung		dukung	
	demokrat	dan	demokrat	dan
	pks, ada empat		pks ada empat	
	alasan anies		alasan anies	
	baswedan	gak	baswedan	gak
	bakal	bisa	bakal	bisa
	menang	jadi	menang	jadi
	presiden		presiden	
4	anies	baswedan	anies	baswedan
	lh yg pantas	jadi	lh yg pantas	jadi
	presiden		presiden	
5	pak	anies	pak	anies
	baswedan	calon	baswedan	calon
	presiden		presiden	

4. Tokenize

Tahap ini berfungsi untuk memecah kalimat menjadi token per kata. Proses ini dilakukan dengan tujuan agar proses selanjutnya dapat diterapkan dalam level kata. Proses ditunjukkan pada tabel 4 berikut:

Tabel 4. Tabel Hasil Tokenizing

No	Kalimat Awal	Hasil Proses
1	pks dukung anies baswedan jadi calon presiden	[[pks] [dukung] [anies] [baswedan] [jadi] [calon] [presiden]]
2	anies baswedan presiden indonesia	[[anies] [baswedan] [presiden] [Indonesia]]
3	walau sekarang dukung demokrat dan pks, ada empat alasan anies baswedan gak bakal bisa menang jadi presiden	[[walau] [sekarang] [dukung] [demokrat] [dan] [pks] [ada] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [bisa] [menang] [jadi] [presiden]]

4	anies baswedan	[[anies]
	lh yg pantas jadi	[baswedan] [lh]
	presiden	[yg] [pantas]
		[jadi] [presiden]]
5	pak anies	[[pak] [anies]
	baswedan calon	[baswedan]
	presiden	[calon]
		[presiden]]

5. Remove Stopwords

Stopwords merupakan untuk menghapus kata yang tidak memiliki dampak pada sentimen, namun sering muncul pada kalimat. Tabel 5 menunjukkan proses Remove Stopwords sebagai berikut:

Tabel 5. Tabel Hasil *Remove Stopwords*

No	Kalimat Awal	Hasil Proses
1	[[pks] [dukung]	[[pks] [dukung]
	[anies]	[anies]
	[baswedan]	[baswedan]
	[jadi] [calon]	[calon]
	[presiden]]	[presiden]]
2	[[anies]	[[anies]
	[baswedan]	[baswedan]
	[presiden]	[presiden]
	[Indonesia]]	[Indonesia]]
3	[[walau]	[[sekarang]
	[sekarang]	[dukung]
	[dukung]	[demokrat] [pks]
	[demokrat] [dan]	[empat] [alasan]
	[pks] [ada]	[anies]
	[empat] [alasan]	[baswedan]
	[anies]	[gak] [bakal]
	[baswedan]	[menang]
	[gak] [bakal]	[presiden]]
	[bisa] [menang]	
	[jadi] [presiden]]	
4	[[anies]	[[anies]
	[baswedan] [lh]	[baswedan] [lh]
	[yg] [pantas]	[pantas]
	[jadi] [presiden]]	[presiden]]
5	[[pak] [anies]	[[pak] [anies]
	[baswedan]	[baswedan]
	[calon]	[calon]
	[presiden]]	[presiden]]

6. Slangwords Filtering

Tahapan ini melakukan penghilangan kata-kata gaul menjadi kata baku. Kamus gaul diambil dari repository github yang dipost oleh [louisowen6](#) yang ditunjukkan pada Tabel 6 sebagai berikut:

Tabel 6. Tabel Hasil Slangwords Filtering

No	Kalimat Awal	Hasil Proses
1	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan] [presiden] [Indonesia]]	[[anies] [baswedan] [presiden] [Indonesia]]
3	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [gak] [bakal] [menang] [presiden]]	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies] [baswedan] [tidak] [akan] [menang] [presiden]]
4	[[anies] [baswedan] [pantas] [presiden]]	[[anies] [baswedan] [pantas] [presiden]]
5	[[pak] [anies] [baswedan] [calon] [presiden]]	[[pak] [anies] [baswedan] [calon] [presiden]]

7. Remove Synonim

Pada tahapan ini, penulis melakukan penghapusan sinonim dalam kalimat yang sama jika ada. Contoh penerapan proses ini, yaitu:

Tabel 7. Tabel Hasil Remove Synonim

No	Kalimat Awal	Hasil Proses
1	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan] [presiden] [Indonesia]]	[[anies] [baswedan] [presiden] [Indonesia]]
3	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies]	[[sekarang] [dukung] [demokrat] [pks] [empat] [alasan] [anies]

	[baswedan]	[baswedan]
	[tidak] [akan]	[tidak] [akan]
	[menang]	[menang]
	[presiden]]	[presiden]]
4	[[alhamdulillah]	[[alhamdulillah]
	[mudah]	[mudah] [anies]
	[mudah] [anies]	[baswedan]
	[baswedan]	[takdir] [allah]
	[takdir] [allah]	[swt] [presiden]
	[swt] [presiden]	[aamiin] [yra]
	[aamiin] [yra]	
5	[[pak] [anies]	[[pak] [anies]
	[baswedan]	[baswedan]
	[calon]	[calon]
	[presiden]]	[presiden]]

C. TF-IDF

Bobot TF-IDF merupakan penilaian bobot setiap kata yang ada pada sebuah dokumen, TF-IDF merupakan metode pembobotan fitur yang kuat yang mengukur frekuensi dan distribusi kata untuk menentukan bobotnya. TF-IDF terdiri dari TF (Term Frequency) yaitu frekuensi kemunculan kata dalam dokumen dan IDF (Inverse Document Frequency) yaitu kebalikan dari frekuensi dokumen, nilai TF-IDF dapat diperoleh dengan menggunakan persamaan berikut:

Nilai Term Frequency (TF) diperoleh dengan nilai frekuensi kemunculan fitur t pada dokumen d [7] seperti pada persamaan (1).

$$TF_t = (t, d) \dots \dots \dots (1)$$

Nilai Inverse Document Frequency (IDF) diperoleh dari logaritma banyaknya dokumen n dibagi dokumen df yang mengandung fitur t [8] yang ditampilkan pada persamaan (2).

$$IDF_t = \frac{\log n}{df(t)} + 1 \dots \dots \dots (2)$$

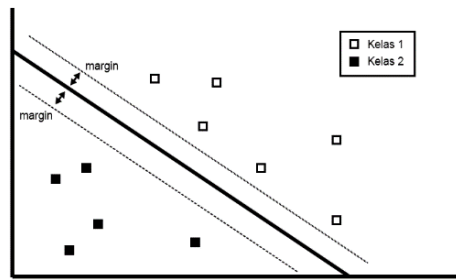
Nilai Term Frequency Inverse Document Frequency (TF-IDF) didapatkan dengan mengalikan nilai TF dan IDF [9] yang ditampilkan pada persamaan (3).

$$W_t = TF_t \cdot IDF_t \dots \dots \dots (3)$$

Hasil TF-IDF sendiri berfungsi untuk mengubah data menjadi bentuk numerikal agar fitur-fitur yang telah diproses dapat dimuat kedalam persamaan. Sehingga metode ini dipilih untuk mengekstrak data tersebut.

D. Support Vector Machine

Support Vector Machine diciptakan oleh Boser, Guyon, dan Vapnik [10] yang merupakan salah satu metode klasifikasi dengan tujuan menemukan hyperplane terbaik untuk memisahkan 2 kelas [11]. Garis pemisah yang akan dibentuk melalui proses pemilihan data yang disebut sebagai *support vector* sehingga nantinya akan menjadi bantuan dalam membentuk garis pemisah terbaik dari sebuah dataset [6]. Algoritma ini akan menghasilkan hyperplane yang memisahkan sebuah plane menjadi 2 bagian pada masing-masing kelas yang dapat dilihat pada gambar 2.



Gambar 2. Contoh Hyperplane Support Vector Machine

Beberapa kernel yang dapat diterapkan pada SVM:

1. Linear Kernel merupakan fungsi kernel yang memisahkan 2 kelas pada persamaan (4).

$$[(w^t \cdot x_i) + b] \geq 1 \text{ for } y_i = 1 \text{ dan } [(w^t \cdot x_i) + b] \leq -1 \text{ for } y_i = -1 \quad (4)$$

Deskripsi :

x_i = dataset training
 i = 1, 2, ..., n
 y_i = label dari x_i [10]

2. Polynomial Kernel merupakan fungsi kernel yang tidak terpisah secara linear yang ditampilkan pada persamaan (5):

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \dots\dots\dots (5)$$

Deskripsi :

x_i, x_j = training data
 $c, d > 0$ = konstanta [10]

3. Radial Basis Function (RBF) Kernel merupakan fungsi untuk menganalisis data yang tidak terpisah secara linear dan memerlukan 2 jenis parameter (Gamma dan Cost).

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \dots\dots\dots (6)$$

Deskripsi :

$\|x_i - x_j\|$ =Euclidean Distance
 2σ =Parameter Independen menentukan tingkat pengurangan $K(x_i, x_j)$ menuju 0 [10]

Untuk menggunakan algoritma Support vector machine, disini perlu dicari terbaik untuk memisahkan tiap kelas. Normalnya dengan data yang tidak terlalu tersebar, hyperplane berupa linear atau polynomial. Namun jika masuk pada kasus penulis, persebaran data yang sangat luas dan dimensi fitur yang tanpa batas, maka diperlukannya penggunaan kernel trick, implementasi yang penulis lakukan untuk mengatasi masalah ini adalah menggunakan Gaussian RBF (Radial Basis Function) yang mampu mengatasi masalah diatas.

E. 10-Fold Cross Validation

K-Fold Cross Validation merupakan salah satu bentuk pengujian yang berfungsi untuk menilai kinerja algoritma dengan cara kerja membagi sampel data secara acak lalu membagi data tersebut sebanyak nilai K dari total fold. Selanjutnya, salah satu kelompok K tersebut akan dijadikan sebagai data testing dan sisa kelompok akan dijadikan sebagai data training [12].

Setelah mesin SVM dilatih, untuk menemukan tingkat untuk menemukan tingkat untuk menemukan tingkat untuk menemukan nilai *f1-score* paling baik dari model SVM, penulis menggunakan metode 10-Fold Cross Validation. Metode ini membagi sebuah dataset menjadi 10 bagian dengan perbandingan data 10% banding 90%. Fold dibagi menjadi 10 bagian karena merupakan angka paling optimal yang memiliki nilai tertinggi [13].

Selanjutnya, akan dilihat dari 10 pembagian data, mana pembagian yang paling mendekati rata-rata *f1-score* dari 10 fold untuk diambil model dari fold tersebut [6].

F. Pengujian

Confusion Matrix merupakan tabel dimana untuk mengevaluasi model klasifikasi, nilai *f1-score* sebagai nilai yang digunakan untuk menentukan model terbaik [14]. Sehingga penulis menerapkan evaluasi pada penelitian dengan metode Confusion Matrix. Nilai yang diambil adalah nilai *f1-score macro* karena lebih merata dalam sebuah dataset yang tidak seimbang [15].

Selanjutnya mencari nilai precision dengan persamaan (7), recall dengan persamaan (8), dan *f1-score* didapat dengan menggunakan persamaan (9).

$$Precision_{(A)} = \frac{TP_A}{TP_A + FP_A} \dots\dots\dots (7)$$

$$Recall_{(A)} = \frac{TP_A}{TP_A + FN_A} \dots\dots\dots (8)$$

$$f1_{(A)} = \frac{2 * Precision_{(A)} * Recall_{(A)}}{Precision_{(A)} + Recall_{(A)}} \dots\dots\dots (9)$$

Dimana:

TP = True Positive

FP = False Positive

FN = False Negative

G. Korelasi Elektabilitas

Pearson Correlation merupakan metode untuk mencari hubungan linear antara 2 variable atau lebih. Hasil dari Pearson Correlation merupakan koefisien korelasi yang berkisar antara angka 0 sampai 1 yang berlaku untuk angka negative [10]. Dalam menguji validitasnya menggunakan Pearson Product Moment Correlation seperti pada persamaan (10).

$$r_{x,y} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \dots\dots\dots (10)$$

Dimana:

n = Jumlah Data

$\sum xy$ = Hasil Jumlah antara X dan Y setiap pasang

$\sum x$ = Hasil Jumlah Variable X

$\sum y$ = Hasil Jumlah Variable Y

$\sum x^2$ = Hasil Jumlah Variable X Kuadrat

$\sum y^2$ = Hasil Jumlah Variable Y Kuadrat

HASIL DAN PEMBAHASAN

Implementasi algoritma SVM (*Support vector machine*) dilakukan persis seperti yang dilampirkan sebelumnya, pembangunan sistem SVM diawali dengan tahap pembersihan data, lalu transformasi data kedalam bentuk numerik, lalu dilakukan pelatihan model dengan pembagian data split 10:90 dengan metode 10-Fold Cross Validation, dan diuji model terbaik dengan mengambil nilai *f1-score* dengan metode Confusion Matrix.

Tahap pertama dalam melakukan train model setelah melakukan pembersihan data adalah melakukan data split Dimana data displit secara acak secara 10 fold. Lalu kita lakukan transformasi data menjadi bentuk numerik dengan menggunakan metode TF-IDF dan train kedalam model SVM dengan kernel RBF. Selanjutnya ditentukan model terbaik berdasarkan hasil *f1-score* dari fold yang mendekati rata-rata *f1-score* keseluruhan diikuti dengan nilai optimal dari parameter C dan Gamma.

Lalu didapat hasil *f1-score* terbaik beserta foldnya dari ketiga calon ditampilkan pada tabel 8.

Tabel 8. Hasil Uji Fold Terbaik Setiap Calon Presiden

Nama Calon	Fold Terbaik	<i>f1-score</i>
Anies Baswedan	8	66,36%
Ganjar Pranowo	5	71,72%
Prabowo	4	77,99%
Subianto		

Dilanjutkan dengan proses mencari korelasi antara data sentiment dengan survey elektabilitas portal berita. Pertama-tama diketahui data sentiment positif, negatif, dan survey elektabilitas portal berita pada calon Anies Baswedan pada tabel 9.

Tabel 9. Data Sentiment dan Elektabilitas Anies Baswedan

Bulan	Positif	Negatif	Elektabilitas
Januari	90,41%	9,59%	15,00%
Februari	91,33%	8,67%	16,95%
Maret	93,23%	6,77%	23,50%

Sehingga didapatkan korelasi antara data sentiment dengan survey elektabilitas untuk calon Anies Baswedan ditampilkan pada tabel 10.

Tabel 10. Korelasi Anies Baswedan

Korelasi Pearson	
Korelasi Sentimen Positif	0,994
Korelasi Sentimen Negatif	-0,994

Lalu diketahui data sentiment dengan survey elektabilitas portal berita untuk calon Ganjar Pranowo ditampilkan pada tabel 11.

Tabel 11. Data Sentiment dan Elektabilitas Ganjar Pranowo

Bulan	Positif	Negatif	Elektabilitas
Januari	96,88%	3,12%	27,58%
Februari	97,58%	2,42%	23,65%
Maret	94,66%	5,34%	28,83%

Sehingga didapatkan korelasi antara data sentiment dengan survey elektabilitas untuk calon Ganjar Pranowo ditampilkan pada tabel 12.

Tabel 12. Korelasi Ganjar Pranowo

Korelasi Pearson	
Korelasi Sentimen Positif	-0,836
Korelasi Sentimen Negatif	0,836

Selanjutnya diketahui data sentiment dengan survey elektabilitas portal berita untuk calon Prabowo Subianto ditampilkan pada tabel 13.

Tabel 13. Data Sentiment dan Elektabilitas Prabowo Subianto

Bulan	Positif	Negatif	Elektabilitas
Januari	99,08%	0,92%	21,68%
Februari	99,69%	0,31%	24,62%
Maret	99,79%	0,21%	23,12%

Sehingga didapatkan korelasi antara data sentiment dengan survey elektabilitas untuk calon Prabowo Subianto ditampilkan pada tabel 14.

Tabel 14. Korelasi Prabowo Subianto

Korelasi Pearson	
Korelasi Sentimen Positif	0,789
Korelasi Sentimen Negatif	-0,789

SIMPULAN

Didapatkan kualitas Mesin Sentiment analysis dengan algoritma Support vector machine menggunakan kernel Gaussian RBF dengan Fold Terbaik untuk masing-masing calon adalah fold ke-6 dengan *f1-score* 0.68, gamma 0.01, dan C 1000 untuk calon Anies Baswedan, fold ke-2 dengan *f1-score* 0.73, gamma 0.1, dan C 100 untuk calon Ganjar Pranowo, dan fold ke-7 dengan *f1-score* 0.70, gamma 1, dan C 10 untuk calon Prabowo Subianto.

Lalu disimpulkan bahwa adanya keterkaitan antara sentimen twitter dengan survei elektabilitas portal berita dengan kesimpulan yang diperoleh untuk calon Anies Baswedan dan Prabowo Subianto, korelasi memiliki hubungan kuat yang searah berarti semakin tinggi sentimen positif, maka semakin tinggi elektabilitas, namun untuk calon Ganjar Pranowo, korelasi dengan hubungan kuat tidak searah sehingga semakin rendah sentimen positif, maka semakin tinggi elektabilitas.

DAFTAR PUSTAKA

- [1] B. W. Sari and F. F. Haranto, "IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, Sep. 2019, doi: 10.33480/pilar.v15i2.699.
- [2] C. N. Zempi, A. Kuswanti, and S. Maryam, "ANALISIS PERAN MEDIA SOSIAL DALAM PEMBENTUKAN PENGETAHUAN POLITIK MASYARAKAT,"

- EKSPRESI DAN PERSEPSI : JURNAL ILMU KOMUNIKASI*, vol. 6, no. 1, pp. 116–123, Jan. 2023, doi: 10.33822/jep.v6i1.5286.
- [3] Fatimah Rahmadayana and Yuliant Sibaroni, “Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.
- [4] I. A. Mastan and Y. Toni, “ANALISIS SENTIMEN TERHADAP TEMPAT KULINER AYAM GEDEBUK DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” *JBASE - Journal of Business and Audit Information Systems*, vol. 3, no. 1, Mar. 2020, doi: 10.30813/jbase.v3i1.2062.
- [5] S. Fendyputra Pratama, R. Andrean, and A. Nugroho, “Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis,” *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 4, no. 2, pp. 2541–3619, 2019, doi: 10.31328/jo.
- [6] F. Feiters Tampinongkol, C. Herdian, H. Basri, and L. Halim, “Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi Identifikasi Penyakit Daun Tomat Menggunakan Gray Level Co-occurrence Matrix (GLCM) dan Support Vector Machine (SVM),” *Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 8, no. 1, 2023, [Online]. Available: <https://www.kaggle.com/>
- [7] N. Arifin, U. Enri, and N. Sulistiyowati, “PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DENGAN TF-IDF N-GRAM UNTUK TEXT CLASSIFICATION,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, pp. 129–136, 2021.
- [8] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, “PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA,” 2020.
- [9] A. S. Arief, “SENTIMENTANALYSIS REVIEW APLIKASI MENGGUNAKAN ALGORITMA SVM PADA APLIKASI MYPERTAMINA,” 2023.
- [10] D. B. Rarasati and J. C. A. Putra, “Correlation Between Twitter Sentiment Analysis with Three Kernels Using Algorithm Support Vector Machine (SVM) Governor Candidate Electability Level,” *COIESE*, pp. 249–256, 2021.
- [11] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, “IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE,” *Jurnal Informatika Polinema*, vol. 4, no. 2, pp. 93–100, 2018.
- [12] A. Hutapea and M. Tanzil Furqon, “Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3957–3961, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>

- [13] H. Florenci Tapikap, B. S. Djahi, and T. Widiastuti, “KLASIFIKASI SPAM E-MAIL MENGGUNAKAN METODE TRANSFORMED COMPLEMENT NAÏVE BAYES (TCNB),” *J-ICON*, vol. 7, no. 1, pp. 21–26, 2019.
- [14] F. Istighfarizkya, N. A. S. ER, I. M. Widiarthaa, L. G. Astutia, I. G. N. A. C. Putra, and I. K. G. Suhartana, “Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur,” *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 11, pp. 167–176, 2022, [Online]. Available: <https://scholar.google.com>
- [15] O. H. Anidjar, A. Barak, B. Ben-Moshe, E. Hagai, and S. Tuvyahu, “A Stethoscope for Drones: Transformers Based Methods for UAVs Acoustic Anomaly Detection,” *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3262702.