

SENTIMENT ANALYSIS PEMILIHAN CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

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

Michael Alfonso¹⁾, Dionisia Bhisetya Rarasati, S.Kom., M.T.I²⁾

^{1, 2)} Program Studi Informatika, Universitas Bunda Mulia

Diterima Tanggal Bulan Tahun / Disetujui Tanggal Bulan Tahun

ABSTRACT

Elections for presidential candidates are held every 5 years with various candidates, especially on Twitter, arguments about political matters often occur that many Twitter users participate in discussions about the election for presidential candidate. Twitter users will tweet to convey arguments and discussions related to the election. Therefore, this study focuses on sentiment analysis to infer user responses to the presidential election and validate it by looking for a correlation between electability survey results and Twitter sentiment data using Pearson Correlation.

In sentiment analysis model, the 10-Fold Cross Validation method is used to find the best model from a dataset with a division of training data and test data with 90:10 split. Then the alphabetic data will be converted into numeric data using the TF-IDF weighting method. To validate the best model, Confusion Matrix is used to get the best f1-score. The model is using Support vector machine algorithm with the Gaussian RBF (Radial Basis Function) kernel. The results of the analysis are compared with the results of the news portal electability survey which contains the 3 candidates using Pearson Correlation.

Based on the search results for the best fold, the best fold was found for each presidential candidate, namely the 8th fold with an f1-score of 0.66 for candidate Anies Baswedan with a total of 2,554 training data and 283 testing data, the 5th fold with an f1-score of 0.72 for the Ganjar Pranowo candidate with a total of 3,330 training data and 370 testing data, and the 4th fold with an f1-score of 0.78 for the Prabowo Subianto candidate with a total of 3,487 training data and 387 testing data. Furthermore, in the Pearson Correlation, a coefficient was found for each presidential candidate, namely Anies Baswedan with a positive sentiment coefficient of 0.876 and a negative sentiment coefficient of -0.876. Furthermore, for the presidential candidate Ganjar Pranowo with a positive sentiment coefficient of 0.894 and a negative sentiment coefficient of -0.894. Then for presidential candidate Prabowo Subianto with a positive sentiment coefficient of 0.97 and a negative sentiment coefficient of -0.97.

This study produces the best fold for each data on each presidential candidate with the f1-score to find the best model for each fold. In the Pearson Correlation result, the higher positive sentiment of each presidential candidate, the higher electability survey data. For further research, research can be discuss about hyper tuning parameters and using other kernels on Support vector machine algorithm.

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

ABSTRAK

Pemilihan calon presiden dilaksanakan setiap 5 tahun dengan berbagai kandidat yang mencalonkan diri, terutama dalam media sosial Twitter lebih sering terjadi argumen seputar hal-hal politik yang tentunya banyak pengguna Twitter turut ikut berdiskusi tentang pemilihan calon presiden ini. Pengguna Twitter akan melakukan tweet untuk menyampaikan argumentasi dan diskusi terkait dengan pemilihan calon presiden ini. Oleh karena itu, penelitian ini berfokus pada sentiment analysis untuk melakukan penyimpulan respon pengguna terhadap pemilihan calon presiden serta melakukan validasi dengan mencari korelasi antara hasil survei elektabilitas dan data sentiment Twitter dengan menggunakan Korelasi Pearson.

Dalam pembangun mesin sentiment, metode 10-Fold Cross Validation digunakan untuk mencari model mesin terbaik dari suatu dataset dengan pembagian data training dan data test sebesar 90:10. Lalu data alfabet akan diubah menjadi bentuk numerik dengan menggunakan metode pembobotan TF-IDF. Selanjutnya, untuk melakukan validasi dari model terbaik menggunakan Confusion Matrix untuk mendapat f1-score. 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.

Berdasarkan hasil pencarian fold terbaik, ditemukan fold terbaik untuk masing-masing calon presiden yaitu fold ke-8 dengan f1-score 0,66 untuk calon Anies Baswedan dengan total 2.554 data training dan 283 data testing, fold ke-5 dengan f1-score 0,72 untuk calon Ganjar Pranowo dengan total 3.330 data training dan 370 data testing, dan fold ke-4 dengan f1-score 0,78 untuk calon Prabowo Subianto dengan total 3487 data training dan 387 data testing. Selanjutnya pada Korelasi Pearson, ditemukan koefisien untuk masing-masing calon presiden yaitu Anies Baswedan dengan koefisien sentiment positif sebesar 0,876 dan koefisien sentiment negatif sebesar -0,876. Selanjutnya untuk calon presiden Ganjar Pranowo dengan koefisien sentiment positif sebesar 0,894 dan koefisien sentiment negatif sebesar -0,894. Lalu untuk calon presiden Prabowo Subianto dengan koefisien sentiment positif sebesar 0,97 dan koefisien sentiment negatif sebesar -0,97.

Penelitian ini menghasilkan fold terbaik untuk tiap data pada masing-masing calon presiden dengan ukuran f1-score untuk mencari model terbaik dari tiap fold. Pada Korelasi Pearson, semakin tinggi sentiment positif setiap calon presiden, maka semakin tinggi juga data survei elektabilitas. Untuk penelitian selanjutnya, dapat dilakukan penelitian yang membahas hyper tuning parameter dan menggunakan kernel lain pada algoritma Support vector machine.

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

PENDAHULUAN

Pemilu terjadi setiap 5 tahun sekali dengan jumlah pemilih yang semakin bertambah setiap periodenya, tercatat penambahan pemilih tetap sejak tahun 2004 sampai 2019 sebanyak 28,90% [1]. Media sosial seperti Facebook, Twitter, LinkedIn, YouTube sudah menjadi bagian dari kehidupan manusia, dibuktikan dengan cara berinteraksi kebanyakan dapat dilakukan dengan sosia media, sampai menjadi identitas diri ke masyarakat umum [2]. Pada sosial media terutama Twitter, banyak penyebaran informasi, hiburan, bahkan opini politik, hal ini tidak luput dari terjadinya pemilu yang merupakan kegiatan politik yang akan dilaksanakan pada tahun 2024, disebut sebagai pesta demokrasi, hal tersebut pastinya ramai diperbincangkan di sosial media. Salah satu cara menyampaikan opini dan lewat media sosial adalah dengan menggunakan tanda pagar tertentu pada tweet yang dipost di Twitter. Twitter sendiri digunakan karena kemudahan dalam membalas tweet dengan like, komentar bahkan reply, sehingga topik yang sedang ramai diperbincangkan dapat dengan mudah tersebar [3]. Oleh karena itu, maka diperlukannya analisa lebih lanjut untuk melihat seberapa besar kesetujuan dan ketidaksetujuan pengguna Twitter terhadap

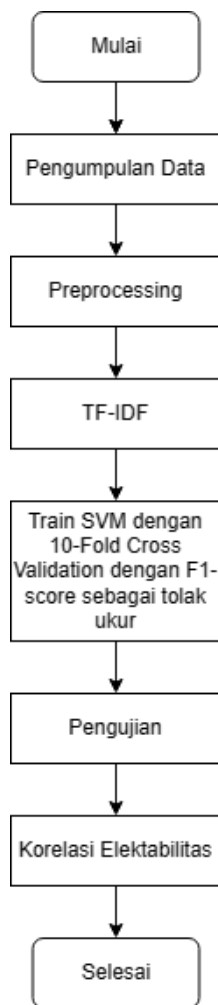
pemilihan calon presiden 2024, lalu data akan diperoleh dari tweet pengguna Twitter yang berkaitan dengan topik ini yaitu pemilihan umum 2024. Sentiment analysis atau dapat disebut Opinion Mining merupakan studi yang mempelajari opini, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas [2], [3]. Sentiment analysis ini digunakan untuk mengetahui tweet yang diposting mengandung kata positif, negatif atau netral [4]. Pada penelitian ini, akan menggunakan metode Support vector machine (SVM) yaitu salah satu metode klasifikasi pada machine learning (supervised learning) yang memprediksi kelas berdasarkan pola dari hasil training. Klasifikasi dilakukan dengan membuat garis pemisah (hyperlane) antara kelas positif dan negatif [2]. Sehingga hasil dari penelitian ini adalah sentiment pengguna Twitter terhadap pemilihan calon presiden pada tahun 2024.

*Korespondensi Penulis:

E-mail: 32190039@student.ubm.ac.id

METODE 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 dibawah ini.



Gambar 1. Flowchart Penelitian

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.

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 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 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 menjadi calon presiden	pks dukung anies baswedan jadi 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 dukung 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	pak anies 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 ditunjukan pada tabel 4 berikut:

Tabel 4. Tabel Hasil *Tokenize*

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] [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	[[pak] [anies] [baswedan] [calon] [presiden]]

5. *Remove Stopwords*

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

Tabel 5. Tabel Hasil *Remove Stopwords*

No	Kalimat Awal	Hasil Proses
1	[[pks] [dukung] [anies] [baswedan] [jadi] [calon] [presiden]]	[[pks] [dukung] [anies] [baswedan] [calon] [presiden]]
2	[[anies] [baswedan]	[[anies] [baswedan]

	[presiden]	[presiden]
	[Indonesia]]	[Indonesia]]
3	[[walau]	[[sekarang]
	[sekarang]	[dukung]
	[dukung]	[demokrat] [pks]
	[demokrat] [dan]	[empat] [alasan]
	[pks] [ada]	[anies] [baswedan]
	[empat] [alasan]	[gak] [bakal]
	[anies] [baswedan]	[menang]
	[gak] [bakal]	[presiden]]
	[bisa] [menang]	
	[jadi] [presiden]]	
4	[[anies]	[[anies]
	[baswedan] [lh]	[baswedan] [lh]
	[yg] [pantas] [jadi]	[pantas]
	[presiden]]	[presiden]]
5	[[pak] [anies]	[[pak] [anies]
	[baswedan]	[baswedan]
	[calon] [presiden]]	[calon] [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]	[[pks] [dukung]
	[anies] [baswedan]	[anies] [baswedan]
	[calon] [presiden]]	[calon] [presiden]]
2	[[anies]	[[anies]
	[baswedan]	[baswedan]
	[presiden]	[presiden]
	[Indonesia]]	[Indonesia]]
3	[[sekarang]	[[sekarang]
	[dukung]	[dukung]
	[demokrat] [pks]	[demokrat] [pks]
	[empat] [alasan]	[empat] [alasan]
	[anies] [baswedan]	[anies] [baswedan]
	[gak] [bakal]	[tidak] [akan]
	[menang]	[menang]
	[presiden]]	[presiden]]
4	[[anies]	[[anies]
	[baswedan]	[baswedan]
	[pantas]	[pantas]
	[presiden]]	[presiden]]
5	[[pak] [anies]	[[pak] [anies]
	[baswedan]	[baswedan]
	[calon] [presiden]]	[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]	[[pks] [dukung]
	[anies] [baswedan]	[anies] [baswedan]
	[calon] [presiden]]	[calon] [presiden]]
2	[[anies]	[[anies]
	[baswedan]	[baswedan]
	[presiden]	[presiden]
	[Indonesia]]	[Indonesia]]
3	[[sekarang]	[[sekarang]
	[dukung]	[dukung]
	[demokrat] [pks]	[demokrat] [pks]
	[empat] [alasan]	[empat] [alasan]
	[anies] [baswedan]	[anies] [baswedan]
	[tidak] [akan]	[tidak] [akan]
	[menang]	[menang]
	[presiden]]	[presiden]]
4	[[alhamdulillah]	[[alhamdulillah]
	[mudah] [mudah]	[mudah] [anies]
	[anies] [baswedan]	[baswedan]
	[takdir] [allah]	[takdir] [allah]
	[swt] [presiden]	[swt] [presiden]
	[aamiin] [yra]	[aamiin] [yra]
5	[[pak] [anies]	[[pak] [anies]
	[baswedan]	[baswedan]
	[calon] [presiden]]	[calon] [presiden]]

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:

1. Nilai *Term Frequency* (TF) diperoleh dengan nilai frekuensi kemunculan fitur t pada dokumen d [5]

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

2. Nilai Inverse Document Frequency (IDF) diperoleh dari logaritma banyaknya dokumen n dibagi dokumen df yang mengandung fitur t [6]

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

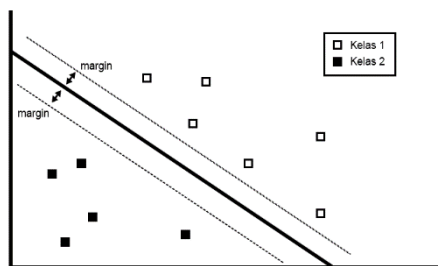
3. Nilai Term Frequency Inverse Document Frequency (TF-IDF) didapatkan dengan mengalikan nilai TF dan IDF [7]

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

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.

Support Vector Machine

Support Vector Machine diciptakan oleh Boser, Guyon, dan Vapnik [8] yang merupakan salah satu metode klasifikasi dengan tujuan menemukan hyperplane terbaik untuk memisahkan 2 kelas [9]. Algoritma ini akan menghasilkan hyperplane yang memisahkan sebuah plane menjadi 2 bagian pada masing-masing kelas yang dapat dilihat pada gambar 2.



Gambar 1. Contoh Hyperplane Support Vector Machine

Beberapa kernel yang dapat diterapkan pada SVM:

1. Linear Kernel merupakan fungsi kernel yang memisahkan 2 kelas dengan persamaan sebagai berikut:

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

Deskripsi :

x_i = dataset training

I = 1, 2, ..., n

y_i = label dari x_i [8]

2. Polynomial Kernel merupakan fungsi kernel yang tidak terpisah secara linear, persamaan sebagai berikut:

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

Deskripsi :

x_i, x_j = training data

$c, d > 0$ = konstanta [8]

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

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

Deksripsi :

$\|x_i - x_j\|$ = Euclidean Distance

2σ = Parameter Independen menentukan tingkat pengurangan $K(x_i, x_j)$ menuju 0 [8]

Untuk menggunakan algoritma Support vector machine, disini perlu dicari \bar{w} 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.

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 [10].

Setelah mesin SVM dilatih, 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%.

Selanjutnya, akan dilihat dari 10 pembagian data, mana pembagian yang paling baik $f1$ -scorenya untuk diambil model dari fold tersebut

Pengujian

Confusion Matrix merupakan tabel dimana untuk mengevaluasi model klasifikasi, nilai $f1$ -score sebagai nilai yang digunakan untuk menentukan model terbaik [11]. Sehingga penulis menerapkan evaluasi pada penelitian dengan metode *Confusion Matrix*. Penerapan *Confusion Matrix* dapat dilihat pada tabel berikut:

Tabel 8. Contoh Tabel *Confusion Matrix*

	Prediksi Positif	Prediksi Netral	Prediksi Negatif
Label Positif	True	False	False
Label Netral	Positif	Netral	Negatif
Label Negatif	False	True	False
Label Positif	Positif	Netral	Negatif
Label Netral	False	False	True
Label Negatif	Positif	Netral	Negatif

Selanjutnya mencari nilai *precision*, *recall*, dan $f1$ -score didapat dengan menggunakan persamaan:

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

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

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

Dimana:

TP = True Positive

FP = False Positive

FN = False Negative

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 [8]. Dalam menguji validitasnya menggunakan *Pearson Product Moment Correlation* dengan persamaan sebagai berikut:

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

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

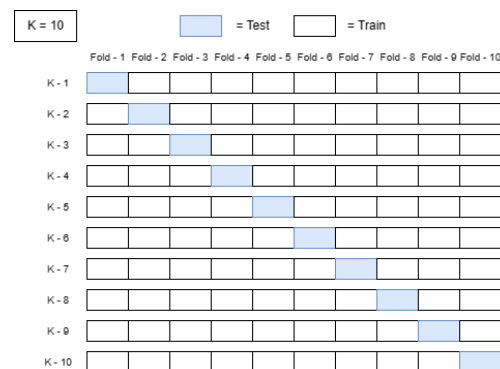
$\sum y^2$ = Hasil Jumlah Variable Y Kuadrat

HASIL DAN PEMBAHASAN

Hasil

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.



Gambar 2. Ilustrasi Cara Kerja 10-Fold Cross Validation

Lalu kita lakukan transformasi data menjadi bentuk numerik dengan menggunakan metode TF-IDF dan train kedalam model SVM dengan kernel RBF.

Selanjutnya dibawah ini adalah rangkuman hasil Korelasi Pearson untuk calon presiden Anies Baswedan.

Tabel 11. Korelasi Anies Baswedan

Korelasi Pearson			
Korelasi	Sentimen	0,876	
Positif			
Korelasi	Sentimen	-0,876	
Negatif			

Diketahui data sentiment positif, negatif, dan survei elektabilitas terhadap data sentiment calon Ganjar Pranowo ditampilkan pada tabel 12

Tabel 12. Data *Sentiment* dan Elektabilitas Ganjar Pranowo

Bulan	Positif	Negatif	Elektabilitas
Januari	96,22%	3,78%	27,58%
Februari	97,30%	2,70%	23,65%
Maret	94,32%	5,68%	28,83%

Selanjutnya dibawah ini adalah rangkuman hasil Korelasi Pearson untuk calon presiden Ganjar Pranowo.

Tabel 13. Korelasi Ganjar Pranowo

Korelasi Pearson			
Korelasi	Sentimen	0,894	
Positif			
Korelasi	Sentimen	-0,894	
Negatif			

Diketahui data sentiment positif, negatif, dan survei elektabilitas terhadap data sentiment calon Prabowo Subianto ditampilkan pada tabel 14

Tabel 14. Data *Sentiment* dan Elektabilitas Prabowo Subianto

Bulan	Positif	Negatif	Elektabilitas
Januari	99,05%	0,95%	21,68%
Februari	99,23%	0,77%	24,62%
Maret	99,79%	0,21%	23,12%

Selanjutnya dibawah ini adalah rangkuman hasil Korelasi Pearson untuk calon presiden Prabowo Subianto.

Tabel 15. Korelasi Prabowo Subianto

Korelasi Pearson

Korelasi	Sentimen	0,970
Positif		
Korelasi	Sentimen	-0,970
Negatif		

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa Implementasi Mesin Sentiment Analysis dengan algoritma Support Vector Machine menggunakan kernel Gaussian RBF ditemukan fold terbaik untuk masing-masing calon presiden yaitu fold ke-8 dengan f1-score 0,66 untuk calon Anies Baswedan dengan total 2.554 data training dan 283 data testing, fold ke-5 dengan f1-score 0,72 untuk calon Ganjar Pranowo dengan total 3.330 data training dan 370 data testing, dan fold ke-4 dengan f1-score 0,78 untuk calon Prabowo Subianto dengan total 3487 data training dan 387 data testing.

Juga korelasi antara survei elektabilitas dengan hasil analisis sentiment positif calon presiden Anies Baswedan adalah 0,876 dan sentiment negatif adalah -0,876. Sehingga kesimpulan yang diperoleh korelasi memiliki hubungan kuat yang searah. Lalu untuk korelasi antara survei elektabilitas dengan hasil analisis sentiment positif calon presiden Ganjar Pranowo adalah 0,894 dan sentiment negatif adalah -0,894. Sehingga kesimpulan yang diperoleh korelasi memiliki hubungan kuat yang searah. Selanjutnya korelasi antara survei elektabilitas dengan hasil analisis sentiment positif untuk calon presiden Prabowo Subianto adalah 0,97 dan sentiment negatif adalah -0,97. Sehingga kesimpulan yang diperoleh korelasi memiliki hubungan kuat yang searah. Sehingga dapat disimpulkan adanya keterkaitan antara sentiment dengan survei elektabilitas portal berita. Disimpulkan dari 3 calon presiden bahwa semakin tinggi sentiment positif, maka semakin tinggi elektabilitas calon tersebut.

DAFTAR PUSTAKA

- [1] BPS, "Hasil Penghitungan Suara Sah Partai Politik Peserta Pemilu Legislatif Tahun 1955-2019.", 13 Jul 2020 [4 Mei 2023].

- [2] 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.
- [3] Fatihah 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] 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.
- [5] D. W. Seno and A. Wibowo, "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Dengan Metode Lexicon Based Dan Support Vector Machine," *Jurnal Ilmiah FIFO*, vol. 11, no. 2, p. 144, Nov. 2019, doi: 10.22441/fifo.2019.v11i2.004.
- [6] 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.
- [7] A. S. Arief, "SENTIMENTANALYSIS REVIEW APLIKASI MENGGUNAKAN ALGORITMA SVM PADA APLIKASI MYPERTAMINA," 2023.
- [8] 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.
- [9] 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.
- [10] 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>
- [11] 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>