

# Penggunaan Natural Language Processing dan Support Vector Machine dalam Analisis Sentimen Bakal Calon Presiden RI Tahun 2024

Mesti<sup>1\*</sup>, Muhammad Arhami<sup>2</sup>, Mahdi<sup>3</sup>

<sup>1,2,3</sup> *Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe  
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA*

<sup>1\*</sup>mesty017@gmail.com

<sup>2</sup>muhammad.arhami@pnl.ac.id

<sup>3</sup>mahdi@pnl.ac.id

*Abstrak*— Perkembangan informasi yang sangat cepat membuat masyarakat Indonesia harus tetap *update* terhadap informasi terkini. Pada tahun 2024 mendatang Indonesia akan mengadakan pesta demokrasi. Beberapa tokoh politik akan maju menjadi bakal calon presiden. Beragam isu dan opini bermunculan, baik terkait prestasi, keunggulan, maupun kekurangan dari masing-masing bakal calon presiden di media sosial. Media sosial yang sering membahas tentang politik yaitu Twitter. Beragam pandangan atau pendapat dicurahkan melalui *tweet*. Kumpulan pandangan orang-orang di Twitter dapat diolah menjadi sebuah informasi baru yang bermanfaat, yaitu dengan cara membuat sistem cerdas yang mampu menganalisis apakah sebuah *tweet* memiliki sentimen positif, netral, atau negatif menggunakan Natural Language Processing (NLP) dan metode support vector machine (SVM). Pengujian ini menggunakan sebanyak 1050 *tweet* yang berkaitan tentang 3 bakal calon kandidat presiden yaitu Prabowo Subianto, Ganjar Pranowo, dan Anies baswedan. Menggunakan penerapan metode *support vector machine* untuk analisis sentimen *tweet*, metode *cross validation* untuk menilai kinerja *support vector machine*, dan metode *confusion matrix* untuk evaluasi model. Berdasarkan hasil evaluasi metode *support vector machine* menggunakan *confusion matrix* diperoleh akurasi sebesar 91% dengan nilai *cross validation* sebesar 89%. Jadi dapat disimpulkan bahwa metode *support vector machine* cocok diterapkan untuk analisis sentimen bakal calon presiden Indonesia tahun 2024.

*Kata kunci*— *Machine Learning, NLP, SVM, Sentimen, Tweet.*

*Abstract*— *The rapid development of information necessitates that the Indonesian public stays updated with the latest news. In 2024, Indonesia will hold a democratic election. Several political figures will run as presidential candidates. Various issues and opinions are emerging, concerning the achievements, strengths, and weaknesses of each presidential candidate on social media. Twitter is a social media platform frequently discussing politics. Diverse views or opinions are expressed through tweets. The collection of people's views on Twitter can be processed into valuable new information by creating an intelligent system capable of analyzing whether a tweet has a positive, neutral, or negative sentiment using Natural Language Processing (NLP) and the support vector machine (SVM) method. This test uses 1050 tweets related to three presidential candidates, Prabowo Subianto, Ganjar Pranowo, and Anies Baswedan. The implementation of the support vector machine method for tweet sentiment analysis, the cross-validation method for evaluating the performance of the support vector machine, and the confusion matrix method for model evaluation are used. Based on the evaluation results of the support vector machine method using the confusion matrix, an accuracy of 91% was obtained with a cross-validation score of 89%. Therefore, it can be concluded that the support vector machine method is suitable for sentiment analysis of Indonesia's 2024 presidential candidates.*

*Keywords*— *Machine Learning, NLP, SVM, Sentiment, Tweet.*

## I. PENDAHULUAN

Indonesia akan mengadakan pesta demokrasi pemilihan presiden pada tahun 2024. Beberapa tokoh politik akan maju menjadi bakal calon presiden pada pemilu tersebut. Beberapa tokoh tersebut diantaranya adalah Ganjar Pranowo, Prabowo Subianto, Anies Baswedan, Agus Harimurti Yudhoyono, Sandiaga Salahudin Uno, dan lain-lain, seperti yang dilansir Kompas tanggal 03/09/2022. Kemunculan beberapa tokoh tersebut ke publik telah menjadi topik yang hangat diperbincangkan dikalangan masyarakat terkait siapa yang cocok untuk dipilih menjadi presiden Indonesia pada pemilu 2024 mendatang. Beragam isu dan opini bermunculan, baik terkait prestasi, keunggulan, maupun kekurangan dari masing-masing bakal calon kandidat di media sosial. Media sosial yang paling umum digunakan masyarakat untuk bebas beropini adalah Twitter. Twitter merupakan salah satu media sosial yang sering digunakan oleh politisi untuk kampanye politik di Indonesia [2-3]. Media kampanye yang digunakan cukup banyak, twitter merupakan salah satu media paling terkenal yang digunakan oleh politisi [3]. Twitter telah digunakan oleh banyak politisi Indonesia untuk menarik simpati publik dan meningkatkan popularitas menjelang pemilihan [4]. Melalui Twitter, politisi memposting pendapat, pemikiran, dan aktivitas mereka untuk menarik dan memengaruhi orang agar memilih mereka dalam pemilihan umum. Selain itu, Twitter telah digunakan sebagai sumber data untuk menganalisis sentimen publik dalam pemilu di beberapa negara seperti Indonesia, India, Jerman, Amerika Serikat, Inggris, dan Bulgaria [2][5-9].

Perkembangan penggunaan Twitter yang sangat cepat mampu menarik perhatian masyarakat Indonesia. Hasil *survey* Monavia Ayu Rizaty yang dilansir dari situs dataindonesia.id pada 10 Agustus 2022, menyebutkan bahwa Indonesia menjadi salah satu negara dengan pengguna Twitter terbanyak di dunia. Laporan *We Are Social*, mencatat bahwa jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada 2022. Jumlah tersebut setara dengan 4,23% dari total pengguna Twitter di dunia yang mencapai 436 juta. Jumlah pengguna Twitter di Indonesia juga naik 31,3% dibandingkan tahun sebelumnya yaitu 14,05 juta pengguna. Kemudahan dan kebebasan dalam penggunaan merupakan alasan Twitter digemari masyarakat Indonesia dalam beropini.

Ada beberapa permasalahan yang berkaitan dengan opini atau sentimen masyarakat terhadap calon presiden

RI pada tahun 2024 diantaranya:

- 1) Ada 3 bakal calon kandidat yang akan bertarung di pilpres 2024 nanti, sehingga masyarakat memerlukan informasi atau tanggapan dari masyarakat lainnya tentang bagaimana prestasi ataupun kekurangan dari ketiga bakal calon selama menjabat di ranah pemerintahan Indonesia.
- 2) Belum adanya sistem yang secara khusus membahas tentang sentimen masyarakat Indonesia terhadap bakal calon presiden pada tahun 2024.

Analisis sentimen bakal calon presiden Indonesia tahun 2024 diharapkan dapat memberikan sudut pandang tentang bakal calon presiden kepada masyarakat agar lebih kritis dalam menentukan pilihan pada pemilu 2024 mendatang, sehingga melahirkan sosok pemimpin yang sesuai harapan bangsa dan membawa kemajuan untuk negara Indonesia.

Penelitian ini menggunakan penerapan *Natural Language Processing* (NLP) dan metode *support vector machine*, serta beberapa algoritma pendukung yaitu *cross validation*, dan *confusion matrix* untuk menunjang proses pembelajaran *machine*. *Natural Language Processing* (NLP) adalah Bahasa yang digunakan manusia untuk berinteraksi [9]. NLP berkaitan dengan bagaimana komputer dapat memahami dan mengerti Bahasa manusia [10]. Beberapa proses NLP yang diterapkan pada penelitian ini adalah [11]:

- 1) *Case Folding*: berperan dalam pembentukan data yang terstruktur dan konsisten. Digunakan untuk mengubah seluruh kalimat menjadi huruf kecil.
- 2) *Cleaning*: merupakan proses pembersihan kalimat dari tanda, emoticon, hastaq, link, dll.
- 3) *Stopword Removal*: merupakan proses pembersihan kata hubung atau kata yang tidak memiliki makna. *Stopwords* di anggap sebagai *noise* dalam teks dan mengacu pada kata-kata yang paling umum seperti kata “pada”, “dengan”, “yang”, “ke” dan lainnya.
- 4) Tokenisasi: merupakan proses yang digunakan untuk memecah sebuah kalimat menjadi perkata. Dengan adanya tokenisasi dapat dilihat perbedaan kata yang memiliki makna dan tidak memiliki makna seperti kata hubung.
- 5) *Stemming*: adalah proses penghapusan kata awalan dan akhiran, bertujuan untuk mengembalikan kata ke bentuk dasarnya sehingga walaupun sudah dalam bentuk kata yang berbeda tetap memiliki makna yang sama. Contohnya “membuka” dan “bukalah” menjadi “buka”.

Proses Analisis sentimen menggunakan metode

*support vector machine* pernah dilakukan oleh A. Tane dan Kemas Muslim Lhaksmana pada penelitiannya yang berjudul “Analisis Sentimen pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode SVM (Support Vector Machine)”, penelitiannya menghasilkan akurasi sebesar 91.5%[12]. Dan penelitian Digna Tata Lukmana dkk yang berjudul “Analisis Sentimen Terhadap Calon Presiden 2019 dengan *Support Vector Machine* di Twitter” menggunakan dua dataset yaitu dataset Jokowi dan Prabowo. Untuk dataset Jokowi menghasilkan akurasi sebesar 86,82% dan dataset Prabowo dengan akurasi sebesar 86,37% [12].

Berdasarkan referensi di atas, maka pada penelitian ini dilakukan penelitian menggunakan *Natural Language Processing* (NLP) dan metode *support vector machine*. Tujuannya untuk melihat seberapa akurat metode ini menganalisis sentimen bakal calon presiden Indonesia tahun 2024.

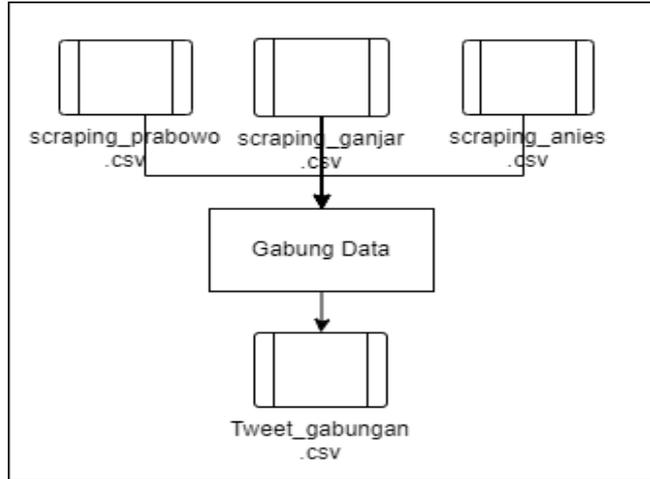
II. METODOLOGI PENELITIAN

Bagian ini menjelaskan secara rinci tentang penelitian yang dilakukan.

A. Teknik Pengumpulan Data

Teknik pengumpulan data adalah tahapan yang dilakukan untuk memperoleh dataset yang akan digunakan pada proses *training* dan *testing* model.

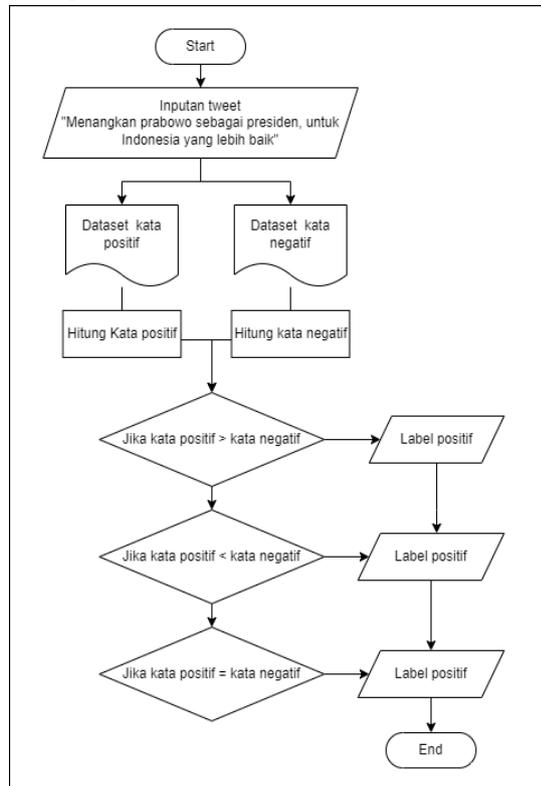
- 1) Dataset Kata Positif dan Negatif: diambil dari link <https://github.com/prasastoadi/IDOpinionWords/blob/master/readme.md>. Terdapat sebanyak 1182 kata yang mengandung makna positif dan sebanyak 2402 kata mengandung makna negatif. Dataset ini digunakan untuk proses labeling data secara otomatis dengan sistem.
- 2) *Scraping* Data: menggunakan data *tweet* yang diperoleh dari Twitter. *Scraping* data menggunakan bahasa pemrograman *python* dan *web driver* (*selenium*). Proses *scraping* dilakukan untuk setiap nama kandidat dengan *keyword* #prabowosubianto, #ganjarpranowo, dan #aniesbaswedan. Setiap *keyword* hasil *scraping* disimpan ke dalam file csv yang berbeda. Sehingga menghasilkan 3 file yaitu data hasil *scraping* Prabowo Subianto, Ganjar Pranowo, dan Anies Baswedan.
- 3) Gabung Data: gambar 1 menunjukkan rancangan proses penggabungan data *scraping* ketiga bakal calon presiden Indonesia tahun 2024.



Gambar 1. Gabung Data

Penggabungan data bertujuan untuk menggabungkan ketiga file hasil *scraping tweet* tentang Prabowo Subianto, Ganjar Pranowo, dan Anies Baswedan menjadi satu file yang sama, sehingga proses labeling tidak perlu 3 kali penginputan data.

- 1) Labeling data: dilakukan secara tersistem menggunakan dataset kata positif dan negatif dengan jumlah 1182 kata positif dan 2402 kata negatif. Gambar 2 mendeskripsikan alur pelabelan otomatis.

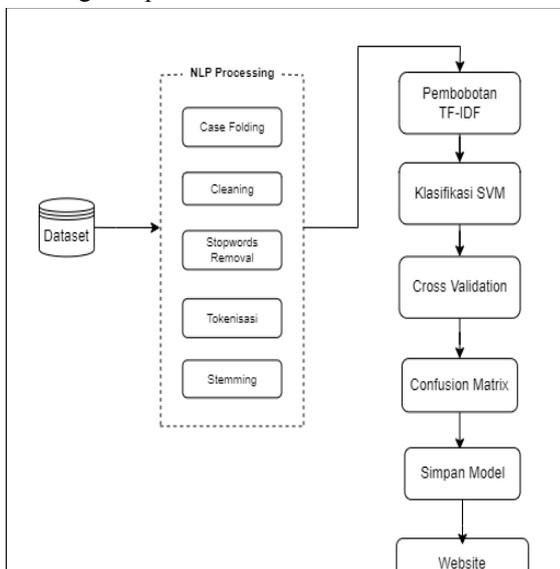


Gambar 2. Alur Label Otomatis

Sistem akan mengecek setiap kata yang ada pada *tweet* dan menyesuaikannya dengan dataset kata positif dan negatif, masing-masing akan dihitung jumlah kata positif dan negatif. Jika kata positif lebih banyak dibanding kata negatif maka *tweet* akan dilabeli sebagai label positif. Jika sebaliknya kata positif lebih sedikit dibanding kata negatif maka *tweet* dilabel sebagai negatif, dan jika jumlah kata positif dan negatif sama banyak maka *tweet* dilabel sebagai netral.

### B. Arsitektur Umum

Perancangan sistem analisis sentimen bakal calon presiden Indonesia tahun 2024 terdiri atas beberapa tahapan seperti yang ditunjukkan pada Gambar 3. Berdasarkan analisa hasil teknik pengujian data, pada penelitian ini diambil sebanyak 1050 *tweet*. Proses *preprocessing* atau pembersihan data dilakukan dengan beberapa tahapan yaitu *case folding*, *cleaning*, *stopword removal*, tokenisasi, dan *stemming*. Proses pembobotan kata menggunakan metode TF-IDF. Pembobotan ini dilakukan untuk melihat seberapa tinggi frekuensi kemunculan sebuah kata pada proses analisis. Tahap analisis sentimen dengan metode *support vector machine*. Validasi data menggunakan metode *cross validation* dan evaluasi model menggunakan metode *confusion matrix*. Kemudian dihubungkan dengan sistem *website* agar dapat divisualisasikan.



Gambar 3. Arsitektur Umum

### C. Dataset

Dataset yang digunakan dalam penelitian ini berjumlah 1050 *tweet*. Dataset terbagi atas 3 kandidat yaitu Prabowo Subianto dengan jumlah 350 *tweet*,

Ganjar Pranowo 350 *tweet*, dan Anies Baswedan 350 *tweet*. Setiap data kandidat memiliki 3 kategori label yaitu 120 *tweet* positif, 115 *tweet* netral, dan 115 *tweet* negatif.

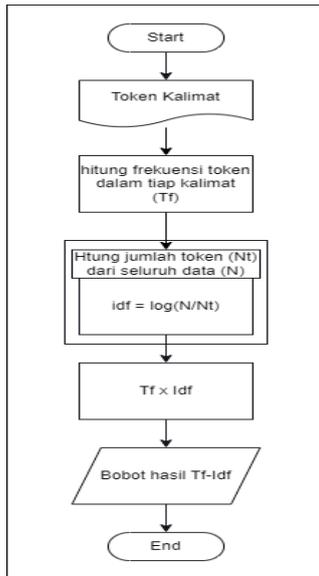
### D. Preprocessing

*Preprocessing* dilakukan untuk membersihkan data yang belum terstruktur sehingga data tersebut dapat digunakan untuk merancang model *machine learning*. Tahapan yang dilakukan sebagai berikut :

- 1) *Case Folding*: mengubah seluruh huruf besar menjadi huruf kecil. Langkah ini dilakukan agar membantu proses normalisasi dengan membuat semua karakter pada dataset seragam
- 2) *Cleaning*: adalah proses pembersihan segala karakter yang tidak berguna dan dapat mempengaruhi performa seperti angka, emoji, tanda baca, *link*, simbol.
- 3) *Stopword Removal*: adalah proses penghapusan kata-kata umum yang dianggap tidak penting, seperti kata hubung “dan”, “atau”, “di”, “dari”, “pada”, “sehingga”, “ke”, “yang”, “lalu”, “sehingga”, “maka”, dan sebagainya.
- 4) Tokenisasi: adalah proses memecah kalimat menjadi kepingan kata yang disebut token. Tokenisasi berguna saat ingin mengekstrak makna dari sebuah kata. Pada proses ini, penulis menggunakan *function word tokenize* yang disediakan oleh *library* NLTK. Tokenisasi sangat berperan penting dalam proses kerja *Neural Language Preprocessing* atau pengolahan kata. Dimana proses ini menjadi bagian penting ketika proses pembobotan kata, setiap kalimat yang telah ditokenisasi akan menghasilkan setiap kata yang memiliki nilai masing-masing.
- 5) *Stemming*: bertujuan untuk mengembalikan kata ke bentuk dasarnya dengan cara menghilangkan awalan dan akhiran

### E. Pembobotan TF-IDF

Pembobotan TF-IDF dilakukan untuk mencari seberapa tinggi frekuensi kemunculan sebuah kata. Alur kerja metode ini diilustrasikan pada Gambar 4 berikut.

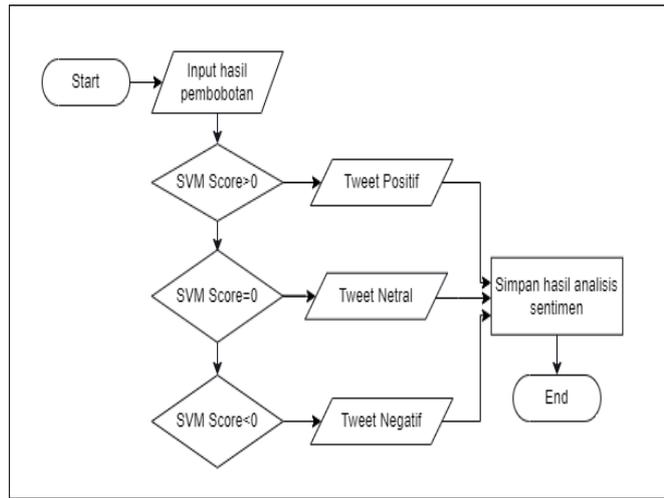


Gambar 4. Pembobotan TF-IDF

Setelah melalui tahap *preprocessing* dihasilkan dataset yang telah memiliki token pada masing-masing *tweet* / kalimat. Kemudian akan dihitung *frekuensi* kata terhadap sebuah *tweet* dari seluruh *tweet*. Berdasarkan proses tersebut menghasilkan nilai TF dan IDF. Kedua nilai tersebut kemudian dikalikan untuk memperoleh nilai bobot TF-IDF.

F. Analisis Sentimen Menggunakan Metode Support Vector Machine

Proses analisis sentimen bakal calon presiden Indonesia tahun 2024 menggunakan metode *support vector machine* ditunjukkan pada Gambar 5. SVM adalah algoritma pembelajaran mesin yang dapat digunakan untuk membangun model klasifikasi dengan memaksimalkan margin antara kelas yang berbeda dalam ruang fitur. Pada penelitian ini, melatih satu model SVM dengan beberapa kelas sekaligus. SVM akan mencari *hyperplane* yang dapat memisahkan ketiga kelas sentimen di ruang fitur. Dalam berbagai kasus analisis sentimen, *hyperplane* pada SVM berfungsi sebagai batas pemisah antara sentimen positif, netral, dan negatif. jika nilai > 0 maka *tweet* akan di analisis sebagai *tweet* positif, jika nilai= 0 maka *tweet* akan di analisis sebagai *tweet* netral, dan jika nilai<0 maka *tweet* negatif.



Gambar 5. Analisis Sentimen Menggunakan Metode Support Vector Machine

G. Cross Validation

Mengukur kinerja metode *Support Vector Machine* (SVM) menggunakan *cross-validation* adalah langkah penting untuk mengevaluasi seberapa baik model SVM bekerja di berbagai subdataset yang berbeda.

Proses *cross-validation* sebagai berikut:

- 1) Bagi data pelatihan menjadi k bagian yang sama besar (k-fold).
- 2) Untuk setiap fold, pilih satu fold sebagai data pengujian dan sisanya sebagai data pelatihan.
- 3) Latih model SVM pada data pelatihan yang tersedia
- 4) Ulangi langkah b-d untuk semua fold.
- 5) Hitung rata-rata yang dihasilkan dari semua fold untuk mendapatkan estimasi kinerja model secara keseluruhan.

H. Confusion Matrix

*Confusion matrix* adalah alat evaluasi yang digunakan dalam analisis klasifikasi untuk mengukur kinerja model atau sistem klasifikasi dengan membandingkan hasil prediksi dengan nilai sebenarnya dari data yang diuji. Metode *confusion matrix* memetakan berbagai hasil klasifikasi menjadi empat komponen utama :

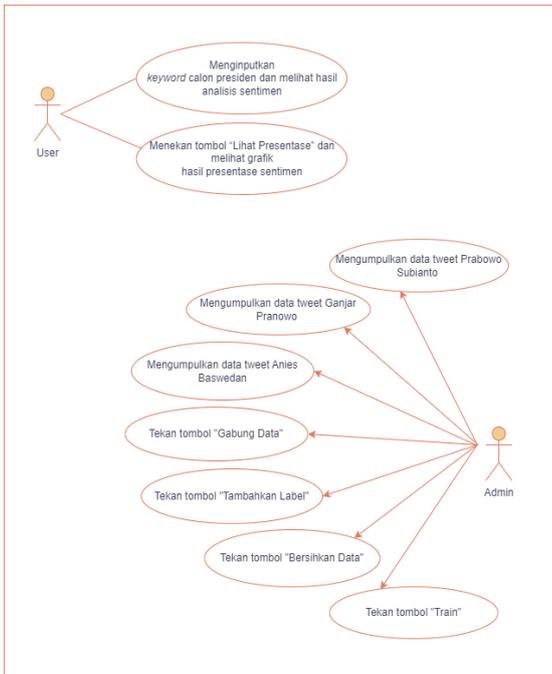
- 1) *True Positive* (TP) : Ini adalah jumlah kasus di mana model dengan benar memprediksi hasil positif ketika hasil sebenarnya juga positif.
- 2) *True Negative* (TN) : Ini adalah jumlah kasus di mana model dengan benar memprediksi hasil negatif ketika hasil sebenarnya juga negatif.
- 3) *False Positive* (FP) : Ini adalah jumlah kasus di mana model salah memprediksi hasil positif, yaitu

model memprediksi positif padahal hasil sebenarnya negatif.

- 4) *False Negative (FN)* : Ini adalah jumlah kasus di mana model salah memprediksi hasil negatif, yaitu model memprediksi negatif padahal hasil sebenarnya positif.

I. Use Case Diagram

Use case menggambarkan interaksi antara aktor dengan sistem.



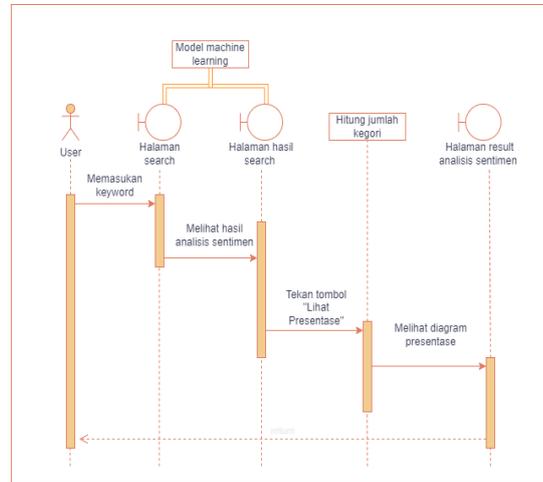
Gambar 6. Use Case Diagram

Pada Gambar 6 dideskripsikan bahwa user dapat melakukan beberapa *activity* pada sistem yaitu memasukkan *keyword* calon presiden, melihat hasil analisis sentimen, melihat grafik persentase sentimen dengan menekan tombol hasil analisis. Sedangkan admin berperan dalam *activity* mengumpulkan dataset ketiga calon, dan menekan beberapa tombol instruksi ke sistem.

J. Sequence Diagram User

*Sequence diagram* user merupakan semua interaksi yang dapat dilakukan user di dalam sistem. Pada tahapan *sequence diagram*, pertama user memasukkan *keyword* yang berhubungan dengan nama Prabowo Subianto, Anies Baswedan, dan Ganjar Pranowo pada halaman *searching*. Sistem akan mengambil 100 data *tweet* secara random dari database yang berkaitan dengan *keyword* inputan. Selanjutnya 100 data tersebut akan di analisis dengan model, hasil analisis kemudian ditampilkan ke

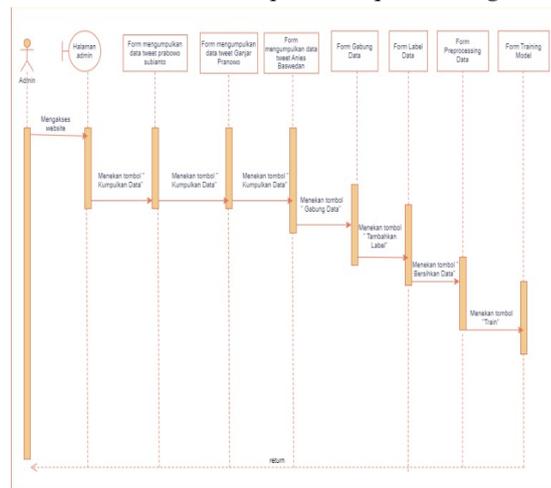
dalam sebuah tabel di halaman hasil *search*, pada halaman ini juga user dapat menekan tombol “Lihat Presentase” untuk melihat presentase dari ketiga kategori sentimen yaitu positif, netral, dan negatif.



Gambar 7. Sequence Diagram User

K. Sequence Diagram Admin

*Sequence diagram* admin merupakan semua interaksi yang dapat dilakukan admin di dalam sistem. Gambar 8 berikut mendefinisikan proses *Sequence diagram* admin.



Gambar 8. Sequence Diagram Admin

Admin mengakses *website* halaman admin, Menekan tombol “Kumpulkan Data” pada form Prabowo subianto, Ganjar Pranowo, atau Anies Baswedan. Setelah data di *scraping*, admin dapat menekan tombol “Gabung Data” pada form gabungan data, dilanjutkan dengan menekan tombol “Tambahkan Label” untuk melabel data. Kemudian menekan tombol “Bersihkan Data” agar data yang akan di *training* menjadi setara. Terakhir admin menekan tombol “Train” pada form *training* data untuk mendapatkan model baru.

### III. HASIL DAN PEMBAHASAN

#### A. Implementasi Proses

Implementasi proses menjabarkan prosedur dan tahapan yang terlibat berdasarkan perancangan model *machine learning* yang telah dijabarkan.

##### 1. Dataset

Contoh dataset yang digunakan pada penelitian ini adalah sebagai berikut.

	username	pesan
0	1632687766416035840	Sepekan Pak Bowo, jejaring dan nama Pak Bowo m...
1	1632988592875380737	Didikan orang tua Pak Bowo yang menjadi bekal ...
2	1633662046867963905	Bravo Pak ❤️???????\n\nMenteri Pertahanan RI ...
3	1634101865108631553	Romo Syafeei Ceritakan Ketulusan Pak Bowo untu...
4	1634117147604717568	Totalitas & Loyalitas Pak Bowo di Era Peme...

Gambar 9. Dataset

##### 2. Case Folding

Proses *case folding* menjadikan semua karakter menjadi huruf kecil. Hasil proses *case folding* ditunjukkan pada Gambar 10.

	case folding result :
0	sepekan pak bowo, jejaring dan nama pak bowo m...
1	didikan orang tua pak bowo yang menjadi bekal ...
2	bravo pak ❤️???????\n\nmenteri pertahanan ri ...
3	romo syafeei ceritakan ketulusan pak bowo untu...
4	totalitas & loyalitas pak bowo di era peme...

Gambar 10. Hasil Case Folding

##### 3. Cleaning

Data dibersihkan dari segala karakter yang tidak berguna dan dapat mempengaruhi performa seperti angka, emoji, tanda baca, *link*, simbol. Hasil proses *data cleaning* ditunjukkan pada Gambar 11.

	cleaning result :
0	sepekan pak bowo jejaring dan nama pak bowo ma...
1	didikan orang tua pak bowo yang menjadi bekal ...
2	bravo pak menteri pertahanan ri pak prabowo su...
3	romo syafeei ceritakan ketulusan pak bowo untu...
4	totalitas amp loyalitas pak bowo di era pemer...

Gambar 11. Hasil Cleaning

##### 4. Stopword Removal

*Stopword removal* adalah proses penghapusan kata hubung yang tidak memiliki makna dalam sebuah sentimen.

	stopword removal result :
0	sepekan bowo jejaring nama bowo menguat
1	didikan orang tua bowo bekal bowo bersikap dem...
2	bravo menteri pertahanan ri prabowo subianto m...
3	romo syafeei ceritakan ketulusan bowo rakyat i...
4	totalitas loyalitas bowo era pemerintahan joko...

Gambar 12. Hasil Stopword Removal

##### 5. Tokenisasi

Melalui proses tokenisasi kalimat dipecah menjadi

per kata. Keluaran proses tokenisasi ditunjukkan pada Gambar 13

	tokenisasi result :
0	[sepekan, bowo, jejaring, nama, bowo, menguat]
1	[didikan, orang, tua, bowo, bekal, bowo, bersi...
2	[bravo, menteri, pertahanan, ri, prabowo, subi...
3	[romo, syafeei, ceritakan, ketulusan, bowo, ra...
4	[totalitas, loyalitas, bowo, era, pemerintahan...

Gambar 13. Hasil Tokenisasi

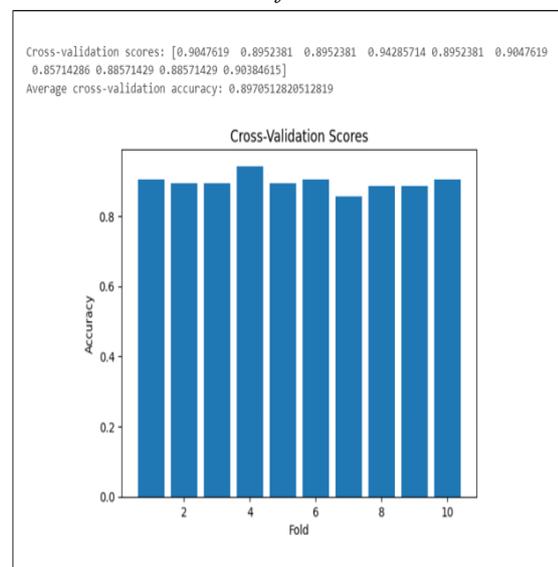
##### 6. Stemming

Kata yang sudah mengalami transformasi, seperti penambahan awalan dan/atau akhiran, dikembalikan ke bentuk dasarnya. Hasil dari proses *stemming* ditunjukkan pada Gambar 14.

	stemming result :
0	[pek, bowo, jejaring, nama, bowo, kuat]
1	[dik, orang, tua, bowo, bekal, bowo, sikap, de...
2	[bravo, menteri, tahan, ri, prabowo, subianto,...
3	[romo, syafeei, cerita, tulus, bowo, rakyat, i...
4	[totalitas, loyalitas, bowo, era, perintah, jo...

Gambar 14. Hasil Stemming

##### 7. Hasil Analisis sentimen menggunakan metode *support vector machine* dan dua metode pendukung yaitu *cross validation* dan *confusion matrix*



Gambar 15. Hasil Cross Validation

Gambar 15 menunjukkan hasil pengujian kinerja metode *support vector machine* menggunakan *cross validation*. Terdapat 10 iterasi pengujian yang masing-masing iterasi menghasilkan akurasi masing-masing berdasarkan baris data yang diuji, secara keseluruhan

diperoleh akurasi rata-rata sebesar **89%**.

Sedangkan pada Gambar 16 merupakan hasil akurasi pembelajaran model menggunakan metode *support vector machine* dan hasil evaluasi model menggunakan metode *confusion matrix*.

```

Confusion matrix:
[[91 1 4]
 [ 5 23 3]
 [ 4 1 78]]
Classification report:
      precision    recall  f1-score   support

-1         0.91     0.95     0.93         96
 0         0.92     0.74     0.82         31
 1         0.92     0.94     0.93         83

 accuracy                   0.91         210
 macro avg         0.92     0.88     0.89         210
 weighted avg     0.91     0.91     0.91         210

Test set accuracy: 0.9142857142857143
    
```

Gambar 16. Evaluasi Model Menggunakan *Confusion Matrix*

Gambar di atas menunjukkan hasil evaluasi metode *support vector machine* menggunakan *confusion matrix* memperoleh akurasi sebesar 91% dengan nilai *precision*: negatif 91%, netral 92%, dan positif 92%. Nilai *recall*: negatif 95%, netral 74%, dan positif 94%. Nilai *f1-score*: negatif 93%, netral 82%, dan positif 93%. Jadi dapat disimpulkan bahwa metode *support vector machine* akurat diterapkan untuk data analisis sentimen bakal calon presiden Indonesia tahun 2024.

8. Perhitungan Manual Confusion Matrix

Untuk membuktikan tingkat kebenaran akurasi dari hasil pembelajaran model dilakukan perhitungan manual *confusion matrix*. Dalam proses ini kita perlu menjumlahkan jumlah prediksi yang benar (*True Positive*) dan membaginya dengan total jumlah data.

Tabel 1. *Confusion Matrix*

	Ng	Nt	P
Ng	91 Tng	1	4
Nt	5	23 TNt	3
P	4	1	78 TP

1. Recall

Rumus :  $Tx/(Tx+FH)$

Recall Ng(Negatif) =  $TNg/(TNg+FHNg) = 91/(91+5) = 0,95$

Recall Nt(Neutral) =  $TNt/(TNt+FHNT) = 23/(23+8) = 0,74$

Recall P(Positif) =  $TP/(TP+FHP) = 78/(78+5) = 0,94$

2. Presisi

Rumus :  $Tx/(Tx+FV)$

Presisi Ng(Negatif) =  $TNg/(TNg+FVNg) = 91/(91+9) = 0,91$

Presisi Nt(Neutral) =  $TNt/(TNt+FVNt) = 23/(23+2) = 0,92$

Presisi P(Positif) =  $TP/(TP+FVP) = 78/(78+5) = 0,92$

3. F1 – score

Rumus :  $2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$

F1 – score Negatif =  $2 \times (0,91 \times 0,95) / (0,91 + 0,95) = 0,93$

F1 – score Netral =  $2 \times (0,92 \times 0,74) / (0,92 + 0,74) = 0,82$

F1 – score Positif =  $2 \times (0,92 \times 0,94) / (0,92 + 0,94) = 0,93$

4. Akurasi

Rumus :  $(TNg(\text{negatif}) + TNt(\text{netral}) + TP(\text{positif})) / (\text{Total Data})$

Jumlah TNg(negatif) = 91

Jumlah TNt(netral) = 23

Jumlah TP(positif) = 78

Total Data = Jumlah semua elemen dalam *confusion matrix* :

$(91 + 1 + 4 + 5 + 23 + 3 + 4 + 1 + 78) = 210$

Akurasi =  $(91 + 23 + 78) / 210 = 0.91$

Berdasarkan perhitungan manual di atas didapatkan hasil yang sama dengan perhitungan sistem yaitu akurasi sebesar **91%**.

B. Implementasi Sistem

Implementasi sistem menjabarkan implementasi dari perancangan sistem.

1. Halaman *Searching*

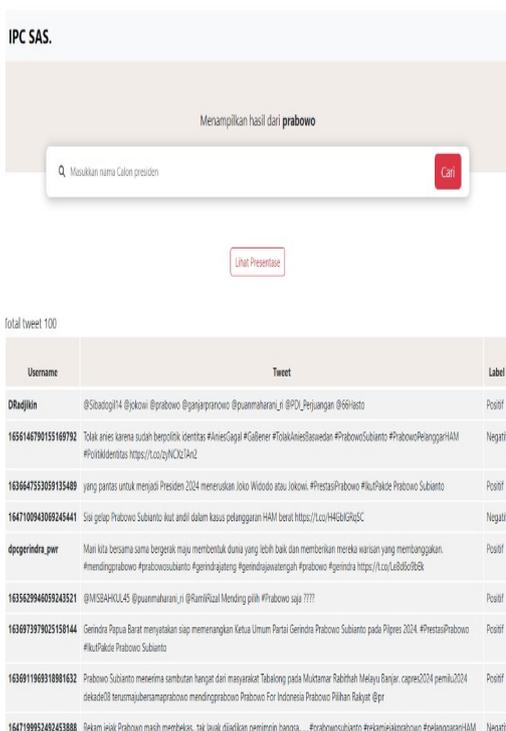
Halaman *searching* adalah halaman utama yang dapat diakses oleh pengguna. Pada halaman ini pengguna menginputkan *keyword* tentang Prabowo, Ganjar, atau Anies.



Gambar 17. Halaman Searching

## 2. Halaman Hasil Searching

Halaman hasil *searching* adalah halaman yang menampung semua *tweet* beserta dengan analisis sentimennya.

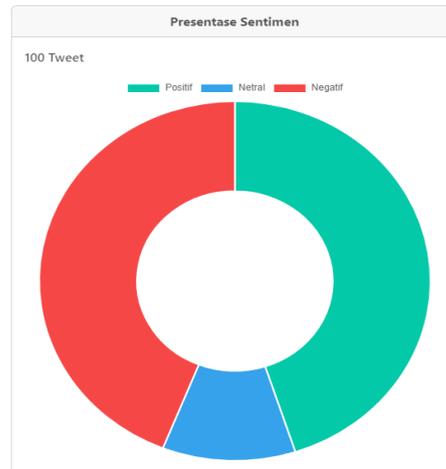


Gambar 18. Halaman Hasil Searching

## 3. Lihat Presentase

Lihat presentase adalah fitur yang digunakan untuk memvisualisasikan hasil analisis sentimen dari ketiga kategori sentimen yaitu, positif, netral, dan negatif. Dengan adanya grafik presentase memudahkan pengguna dalam mendapatkan kesimpulan terkait

bagaimana pandangan masyarakat di Twitter tentang ketiga calon presiden Indonesia tahun 2024.



Gambar 19. Lihat Presentase

## 4. Halaman Admin

Halaman admin berfungsi untuk *scraping* data, label data, *preprocessing* data, dan *training* model. Setiap langkah-langkah disediakan tombol dan urutan masing-masing. Implementasi halaman admin ditunjukkan pada Gambar 20 berikut.



Gambar 20. Halaman Admin

## IV. KESIMPULAN

Setelah melakukan perancangan dan pengujian pada sistem analisis sentimen bakal calon presiden Indonesia tahun 2024. Hasil evaluasi metode *support vector machine* menggunakan *confusion matrix* diperoleh akurasi sebesar **91%** dengan nilai *precision*: negatif 91%, netral 92%, dan positif 92%. Nilai *recall*: negatif

95%, netral 74%, dan positif 94%. Nilai *f1-score*: negatif 93%, netral 82%, dan positif 93%. Dan Analisis kinerja metode dengan *cross validation* menghasilkan akurasi rata-rata sebesar 89%. Jadi dapat disimpulkan bahwa metode *support vector machine* dapat memberikan akurasi yang tinggi ketika diterapkan pada analisis sentimen bakal calon presiden Indonesia tahun 2024.

#### REFERENSI

- [1] Alashri S, Parriott E, Kandala S S, Awazu Y, Bajaj V dan Desouza K C 2018 Pemilihan Presiden AS 2016 di Facebook: Analisis Eksplorasi Sentimen Proc. dari Hawaii Int. ke-51 tentang Ilmu Sistem hlm 1771–80
- [2] Bansal B dan Srivastava S 2018 Tentang memprediksi pemilu dengan analisis sentimen berbasis topik hybrid dari tweet Procedia Computer Sci. 135 hlm 346–53
- [3] Hidayatullah A F dan Sn A 2014 Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada Twitter Seminar Nasional Informatika 2014 (semnasIF 2014) hlm 115–22
- [4] Almatrafi O, Parack S dan Chavan B 2015 Penerapan analisis sentimen berbasis lokasi menggunakan Twitter untuk mengidentifikasi tren menuju pemilihan umum India 2014 Proc. of the 9th Int. Conf. on Ubiquitous Information Management and Communication - IMCOM '15 (Bali, Indonesia: ACM Press) hlm 1–5
- [5] Budiharto W dan Meiliana M 2018 Prediksi dan analisis pilpres Indonesia dari Twitter menggunakan analisis sentimen J. dari Big Data 5
- [6] Heredia B, Prusa J D dan Khoshgoftaar T M 2018 Analisis sentimen Twitter Berbasis Lokasi untuk Memprediksi Pemilihan Presiden AS 2016 The Thirty-First Int. Florida Artificial Intelligence Research Society Conf. (FLAIRS-31) hlm 265–70
- [7] Burnap P, Gibson R, Sloan L, Southern R dan Williams M 2016 140 karakter menuju kemenangan?: Menggunakan Twitter untuk memprediksi Studi Pemilihan Umum Inggris 2015 41 hlm 230–3
- [8] Smailović J, Kranjc J, Grčar M, Žnidaršič M dan Mozetič I 2015 Memantau sentimen Twitter selama pemilu Bulgaria 2015 IEEE Int.
- [9] R. M. Yanti, I. Santoso, and L. H. Suadaa, “Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study: Power Failure in the Special Region of Yogyakarta),” *Indonesian Journal of Information Systems (IJIS)*, vol. 4, no. 1, 2021.
- [10] I. Roldós, “NLP, Machine Learning & AI, Explained.” <https://Monkeylearn.com/blog/nlp-ai/> (accessed Jan. 19, 2023).
- [11] Agnes, “Cara Implementasi Teknik Analisis Data untuk Text Preprocessing,” *Dqlab.id*, May 09, 2022 <https://www.dqlab.id/begini-cara-implementasi-teknik-analisis-data-untuk-text-preprocessing> (accessed Jan. 19, 2023).
- [12] A. Tane and Kemas Muslim Lhaksmana, “Analisis Sentimen pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode SVM (Support Vector Machine),” *eProceedings of Engineering*, vol. 6, no. 2, 2019, Accessed: Dec. 20, 2022. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/1070>
- [13] Lukmana, Digna Tata, Sri Subanti, and Yuliana Susanti. "Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter." *Seminar & Conference Proceedings of UMT*. 2019.