

# Penerapan Algoritma Naive Bayes dan PSO pada Analisis Sentimen Kandidat Calon Presiden 2024

<sup>1</sup>\*Endrik, <sup>2</sup>Agung Nugroho, <sup>3</sup>Ahmad Turmudi Zy  
Universitas Pelita Bangsa  
Bekasi, Indonesia

<sup>1</sup>endrikbinsuwari@gmail.com, <sup>2</sup>agung@pelitabangsa.ac.id  
<sup>3</sup>turmudi@pelitabangsa.ac.id

## \*Penulis Korespondensi

Diajukan : 04/07/2023

Diterima : 26/07/2023

Dipublikasi : 01/08/2023

## ABSTRAK

Pemilihan umum merupakan salah satu sarana demokrasi dan bentuk perwujudan kedaulatan rakyat untuk menghasilkan wakil rakyat dan pemimpin yang aspiratif, berkualitas, serta bertanggung jawab untuk mensejahterakan rakyat. Tahun 2024 Indonesia akan mengadakan pesta demokrasi pemilihan kepala negara Indonesia. Setiap tokoh politik yang dicalonkan menjadi kepala negara akan mempertimbangkan popularitas mereka berdasarkan opini masyarakat. Banyaknya opini pada *twitter* dapat digunakan untuk mengetahui sentimen masyarakat mengenai calon Presiden Indonesia 2024. Penelitian ini bertujuan menganalisa opini masyarakat terkait mengenai 3 kandidat teratas Calon Presiden 2024 untuk mengetahui Hasil klasifikasi senimen analisis masyarakat dengan menggunakan *naive bayes method* dan *feature selection particle swarm optimization* serta hasil evaluasi akurasi terbaik pada analisis sentimen. Parameter yang digunakan untuk pengukuran adalah *accuracy*, *precision*, *recall*. Pada pengujian pertama dengan dataset Anies Baswedan menghasilkan Nilai *accuracy* yang didapatkan Anies Baswedan 63,02%, nilai *recall* yang didapatkan 65,13%, dan nilai *precision* yang didapatkan 64,61%. Nilai *accuracy* yang didapatkan Ganjar Pranowo 87,14%, nilai *recall* yang didapatkan 87,46%, dan nilai *precision* yang didapatkan 85,43%. Nilai *accuracy* yang didapatkan Prabowo Subianto 83,17%, nilai *recall* yang didapatkan 83,17%, dan nilai *precision* yang didapatkan 84,17%.

**Kata Kunci:** *Accuracy, Naive Bayes, Precision, Recall*

## I. PENDAHULUAN

Banyaknya respon masyarakat Indonesia dalam menilai padangan politik menuai opini pro dan kontra, terdapat sejumlah besar pengguna jejaring sosial memungkinkan peneliti untuk menggunakan pendapat ini dalam data sentimen untuk diproses dan dianalisis. *Analisis sentimen* dari pengolahan status atau tweet pengguna twitter menunjukkan opini publik terhadap calon presiden.

Pada penelitian yang dilakukan oleh Junseok Song dkk mengenai “A novel

*classification approach based on Naïve Bayes for Twitter sentiment analysis*” pemilihan fitur untuk sentimen Twitter analisis telah disajikan berdasarkan *Naïve Bayes*. Sebagian besar pembobotan dan fitur atribut pendekatan seleksi berdasarkan algoritma *Naïve Bayes* menggunakan jumlah atribut yang sama untuk memperkirakan bobot setiap kelas. Untuk mengevaluasi akurasi, himpunan bagian dari 1.600.000 kumpulan data yang disediakan oleh Sentimen 140 adalah digunakan dalam simulasi komputer. Total kumpulan data berisi 800.000 tweet kelas positif dan 800.000 tweet kelas negatif. Dalam percobaan ini perbandingan positif dan negatif kelas adalah 5:5 dalam set pelatihan dan tes. Dengan hasil SVM 75,39% dan MNB 85,33% (Srivastava et al., 2019).

Penelitian ini dilakukan oleh Meylan Wongkar dkk mengenai “*Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler:Twitter*” Pada penelitian ini, analisis sentimen pada analisis Twitter menggunakan bahasa pemrograman python. untuk melakukan analisis sentimen ini dilakukan beberapa langkah yaitu pengumpulan data menggunakan library Python, pengolahan teks, pengujian data latih dan klasifikasi teks menggunakan metode *Naive Bayes*. Metode *Naive Bayes* digunakan untuk mengklasifikasikan kelas atau tingkat sentimen masyarakat. Hasil penelitian ini menunjukkan pasangan Jokowi-Ma'ruf Amin memiliki skor polaritas sentiment positif 45,45 dengan skor negatif 54,55%, sedangkan pasangan Prabowo-Sandiaga memiliki skor sentiment positif 44,32 negatif. 55,68%. Data gabungan tersebut kemudian dibandingkan dengan data pendidikan masing-masing capres dengan akurasi 80,90%  $\approx$  80,1%. Pada penelitian ini dilakukan perbandingan dengan metode *naive bayes*, *svm* dan *K-nearest neighbor* (K-NN) yang diuji dengan *RapidMiner* dengan nilai akurasi naive bayes sebesar 75,58%, *svm* dengan nilai akurasi sebesar 63,99. Nilai K-NN sebesar 73,34% (Wongkar & Angdresey, 2021).

Berdasarkan penelitian-penelitian terdahulu yang telah dipaparkan di atas, peneliti tertarik untuk menggunakan metode yang memiliki nilai akurasi yang baik yaitu *naive bayes method* dan seleksi *feature particle swarm optimization* dalam analisis opini publik twitter di media sosial menunjukkan bahwa berdasarkan hasil literatur, metode *Naive Bayes* merupakan metode sederhana dengan akurasi dan efisiensi tinggi untuk klasifikasi teks. Untuk mendapatkan hasil yang lebih akurat menggunakan metode *Naive Bayes* dan *Feature Selection Particle Swarm Optimization*. Pada penelitian ini menggunakan dataset yang bersumber dari media sosial *twitter* yang didapatkan dengan cara crawling data dengan bantuan API twitter menggunakan aplikasi *RapidMiner Studio* dengan kata kunci 3 kandidat teratas menurut survei lembaga yaitu “Anies Baswedan”, “Prabowo Subianto”, “Ganjar Pranomo”.

## II.STUDI LITERATUR

### Penelitian Terdahulu

Penelitian ini dilakukan oleh Ratih Yulia Hayuningtyas dkk mengenai “Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan *Naïve Bayes dan PSO*” dalam penelitian ini dengan tujuan meningkatkan nilai akurasi dengan *Naïve Bayes* dan *Particle Swarm Optimization* dengan ulasan tempat wisata Taman Mini Indonesia Indah. Pada penelitian ini menggunakan data dari 50 ulasan positif dan 50 ulasan negatif (Hayuningtyas & Sari, 2019). Pengujian metode menggunakan *Naive Bayes* dalam eksperimen yang menggunakan validasi silang 1 hingga 10 kali lipat dan populasi 1 hingga 10 kali lipat. Hasil dari eksperimen dan pengujian metode *Naïve Bayes* dengan *fold cross validation* dan population dimana hasil akurasi yang didapat

menggunakan metode *Naïve Bayes* sebesar 70% sedangkan akurasi menggunakan *Naïve Bayes* dan *Particle Swarm Optimization* sebesar 94,02% dengan menggunakan 9 *Fold Cross Validation* (Hayuningtyas & Sari, 2019).

### Pemilihan Umum

Dapat dikatakan bahwa pemilihan umum merupakan salah satu sarana demokrasi dan bentuk ekspresi kedaulatan rakyat untuk mendidik para wakil rakyat dan pemimpin yang memiliki ambisi, kualitas dan tanggung jawab yang cukup untuk kepentingan rakyat. Salah satu kelompok pemilih yang patut dicermati dan dipelajari lebih lanjut adalah pemilih pemula. Pemilih pemula adalah pemilih yang akan memberikan suaranya untuk pertama kali dalam suatu pemilihan. Sebagaimana dijelaskan dalam Modul I Pemilu Pemilih Pemula Komisi Pemilihan Umum (2013), pemilih pemula adalah warga negara yang akan menggunakan hak pilihnya untuk pertama kali dalam kegiatan pemilihan umum (pemilu). Mereka dapat menjadi Warga Negara Indonesia (WNI) meskipun genap berusia 17 tahun atau belum berusia 17 tahun tetapi sudah pernah menikah (Sukma & Wardhani, 2018).

### Calon Presiden 2024

Pada survei yang dilakukan oleh Indikator Politik Indonesia skema 19 nama bakal calon presiden 2024 berdasarkan hasil survei terhadap 1.230 responden selama 26-30 Mei 2023 dengan metode *random digit dialing (RDD)*. Survei memiliki margin of error +/- 2,9 persen dengan tingkat kepercayaan 95 persen (Melani, 2023). Prabowo Subianto 25,3%, Ganjar Pranowo 25,2%, Anies Baswedan 12,5%.

Survei selanjutnya dilakukan oleh Saiful Mujani *Research and Consulting (SMRC)*. Pada 5 Juni 2023, MRC merilis hasil survei bertajuk "Kualitas Popularitas dan Elektabilitas Bacapres di Pemilih Kritis" pada hari ini. Survei yang dilakukan pada 30-31 Mei 2023 melalui telepon ini dipaparkan oleh Direktur *Riset SMRC*, Deni Irvani. Ganjar Pranomo 37,9%, Prabowo Subianto 33,5%, Anies Baswedan 19,2%.

Pada penelitian ini akan memilih 3 calon kandidat teratas yaitu Ganjar Pranomo, Prabowo Subianto dan Anies Baswedan.

### *Naïve Bayes Classifier*

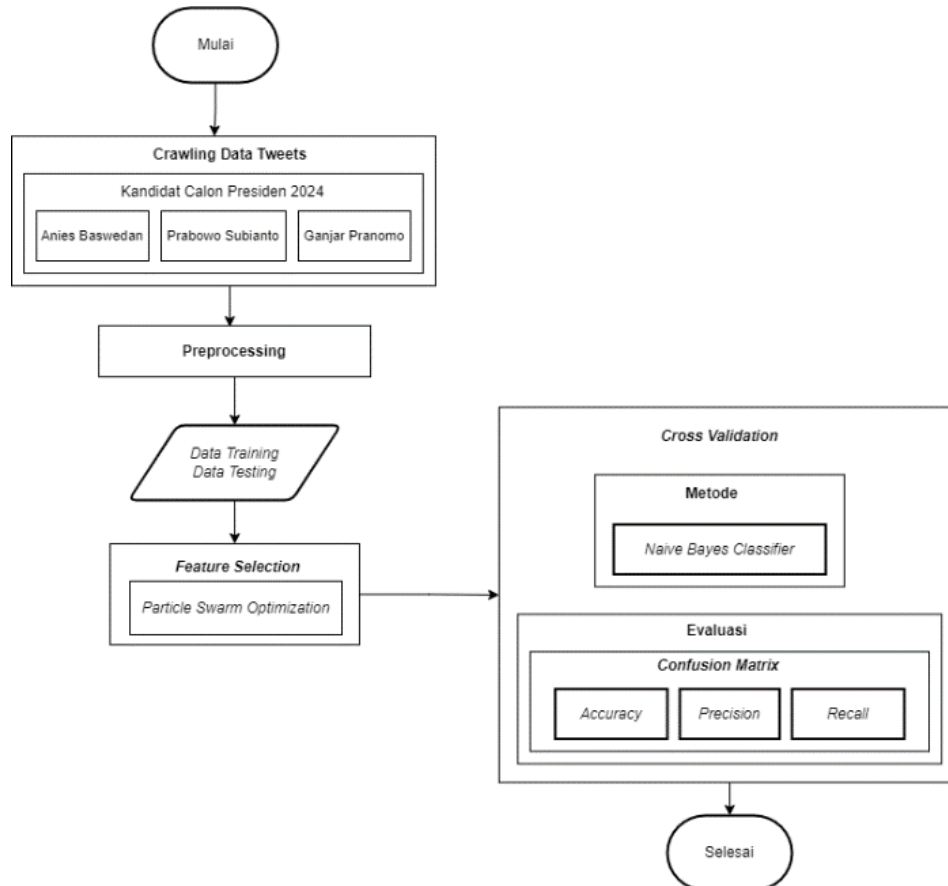
Metode *Naïve Bayes Classifier* adalah teknik klasifikasi sederhana, yang pada prosesnya dilakukan perhitungan terhadap jumlah data dan kombinasi data pada sebuah dataset yang bertujuan untuk mengetahui nilai probabilitas dari dataset tersebut (Prayoga et al., 2019). *Naïve Bayes* adalah metode yang sangat sederhana dalam *machine learning* yang biasanya digunakan untuk menghitung hasil performa suatu klasifikasi (Prayoga et al., 2019).

### *Particle Swarm Optimization*

*Particle Swarm Optimization (PSO)* adalah metode optimisasi proses yang lebih sederhana daripada *algoritma genetika*, terinspirasi oleh perilaku kawanan ikan dan unggas dalam mencari sumber makanannya (Effendi et al., 2022). Partikel adalah setiap individu yang merupakan bagian dari suatu kelompok. Setiap partikel dalam *Particle Swarm Optimization* dapat menyesuaikan posisi dan kecepatannya sehingga cenderung berpindah ke posisi yang dianggap terbaik selama pencarian (Prayoga et al., 2019).

### III. METODE

Pada penelitian ini menggunakan algoritma *Naive Bayes* dengan menambahkan metode seleksi fitur yaitu *Particle Swarm Optimization*. Berikut adalah gambar metode penelitian:



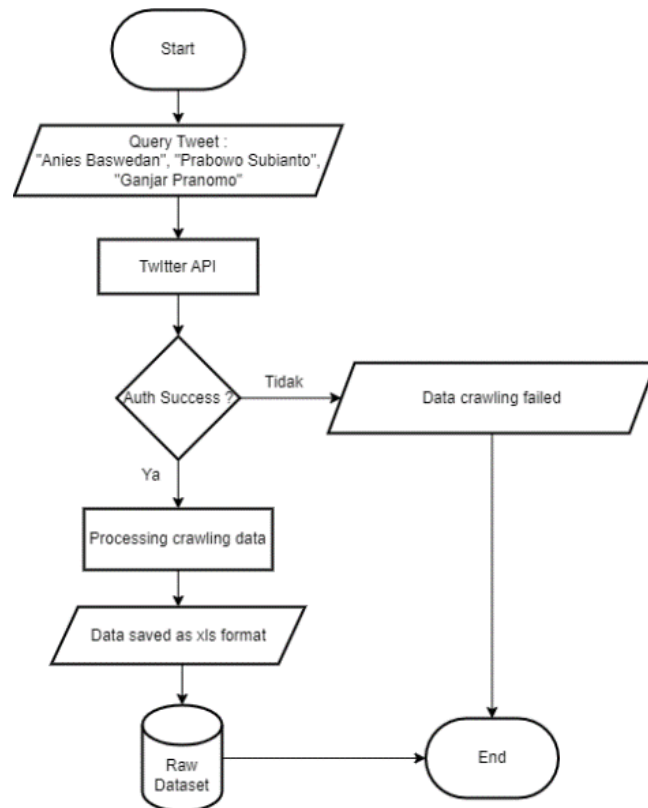
Gambar 1 Metode Penelitian

Uraian dari gambar diatas dapat dilihat sebagai mana penjelasan berikut ini :

#### Pengumpulan Data

Pada akumulasi data atau pengumpulan data twitter dilakukan dengan menggunakan operator pencarian *Twitter* pada perangkat lunak RapidMiner Studio. Data *tweet* yang diambil berisi opini publik atas konflik Capres 2024 di Indonesia yang diambil pada 30 Juni 2024 malam dengan *query* “Anies Baswedan, Ganjar Pranomo, Prabowo”, jenis hasil “*recent or popular*”, batas “10000”, dan Bahasa “*id*”. proses akan selesai ketika *crawling* data yang diambil dari database Twitter mencapai limit yang sudah ditentukan.

*Crawling data* adalah metode pengumpulan atau pengambilan data dari blog, media sosial, atau situs web lainnya (Ramadhan & L, 2018). *Crawling data* terkadang juga disebut *web scraping* atau *spidering*, adalah metode yang memungkinkan secara otomatis mengekstrak data dari internet. Menggunakan sistem otomatis untuk mengekstrak data banyak aplikasi praktis . Jadi berdasarkan pengertian diatas *crawling data* adalah suatu metode untuk mendapatkan dengan cara mengambil atau menarik data dari internet (Claussen & Peukert, 2019).



Gambar 2 Alur proses *crawling* data

**Cleansing**

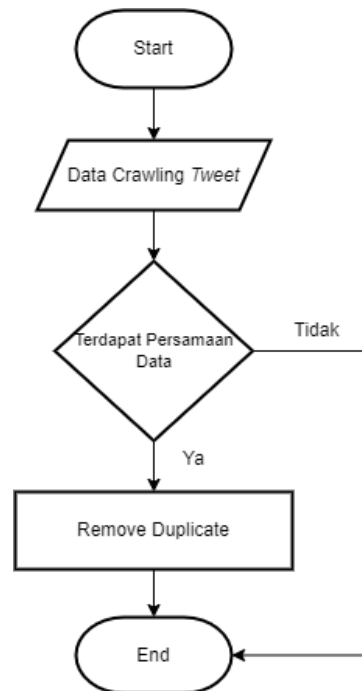
Tujuan dari fase pembersihan (*Cleansing*) data adalah untuk menghilangkan karakter khusus (simbol), *retweet* (RT), *username* dan *link* yang ada pada dataset. Proses *cleansing* dilakukan dengan menggunakan operator *replace* pada *RapidMiner* dengan memasukkan kata yang ingin di *replace* (*replace what*) dan mengkosongkan kata pengganti (*replace by*).

Tabel 1 Kalimat setelah *Cleansing*

Sebelum Proses Cleansing	Sesudah Proses Cleansing
Oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia, yaitu kompromi. #PrabowoNeruskePakdhe Prabowo Subianto #PrabowoMenjagaKerukunan	Oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia, yaitu kompromi.
Anies bukan salah data!! Tapi emang niat berbohong!! #AniesBaswedan <a href="https://t.co/ptWxbOGjjk">https://t.co/ptWxbOGjjk</a>	Anies bukan salah data Tapi emang niat berbohong

**Remove duplicate**

Proses penghapusan atribut dengan nilai yang sama pada data yang di *crawling*. Dibawah ini merupakan gambar alur proses penghapusan data yang bernilai sama (*duplicate*).



Gambar 3 *Remove Duplicate*

**Seleksi data**

Proses seleksi data adalah proses penghapusan beberapa data *tweet* yang tidak diperlukan atau data *tweet* yang tidak berhubungan dengan Capres 2024 contohnya seperti *tweet* promosi penjualan suatu produk. Tahap ini dilakukan dikarenakan masih banyak data yang tidak diperlukan yang didapat dari hasil *crawling*. Data yang tidak berhubungan dengan konflik Capres 2024 akan di hapus secara manual dengan menggunakan software *Microsoft Excel*.

**Normalisasi**

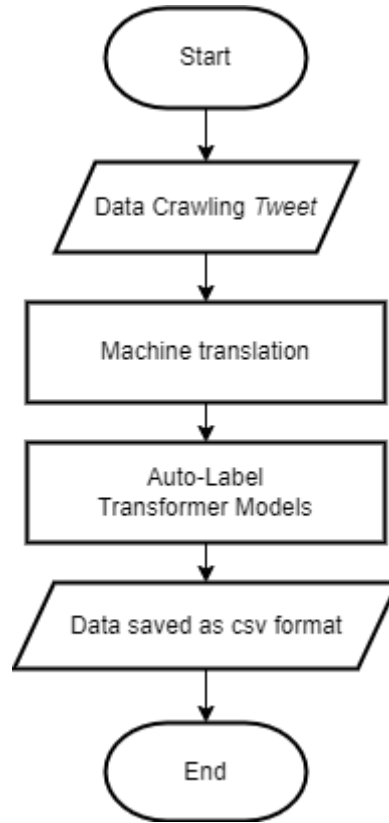
Normalisasi adalah tahapan untuk merubah kata-kata yang terkandung pada *dataset* yang sebelumnya tidak baku menjadi kata yang baku atau sesuai dengan ejaan yang benar.

Tabel 2 Kalimat setelah Normalisasi

Sebelum di Normalisasi	Sesudah di Normalisasi
Seorng	Seorang
Taun	Tahun

**Labeling**

Pemberian label (*Labeling*) merupakan proses untuk menandakan respon *tweet* yang ada di dalam dataset yang dilakukan menggunakan *python* dengan menggunakan *library transformers, pipeline, dan pandas*, komentar yang ada didalam dataset dibagi menjadi dua yaitu koemntar positif, dan negatif. Pendapat yang berisikan kalimat pujian dan dukungan akan dikelompokan kedalam sentimen positif. Pendapat yang berisikan kata ketidak setujuan, ataupun kata-kata kasar akan dimasukan kedalam sentimen negatif.



Gambar 4 Alur proses Labeling data

Tabel 3 Data Tweet Labeling

Nama Capres 2024	Negatif	Positif
Anies Baswedan	1481	953
Ganjar Pranowo	2250	1662
Prabowo Subianto	2286	2413

Tabel 4 Kalimat setelah Labeling

Sebelum Proses Labeling	Sesudah Labeling
Langkah anies mas aniesbaswedan sangat hormat pada pak jokowi apapun keputusan itu yg terbaiknyaallah langkah2 mas anies 2024	Positive
sesungguhnya tongkat itu akan memukul balik aniesbaswedan karena keseringan memanipulasi kata demi lari dari kenyataan dan kegagalan yang dialaminyahidup diselimuti manipulasi kata, hanya akan membawa sengsara karena masalahnya belum sirna	Negative

**Data Uji (Testing) dan Latih (Training)**

Pada proses ini merupakan tahap pemisahan data dengan menggunakan operator *split-data*, bertujuan untuk memisahkan total data menjadi dua bagian, yaitu data latih dan data uji. Tahap pemisahan data dilakukan karena proses klasifikasi membutuhkan data latih dan data uji. Setelah melewati tahap seleksi data, data *tweet* yang berisi opini publik dibagi menjadi dua bagian, 70% sebagai data latih dan 30% sebagai data uji.

Tabel 5 Data Testing dan Training

NAMA CAPRES2024	Training	Testing	TRAINING		TESTING	
			NEGATIF	POSITIF	NEGATIF	POSITIF
ANIES BASWEDAN	1704	730	1037	667	444	286
GANJAR PRANOWO	2738	1174	1575	667	675	995
PRABOWO SUBIANTO	3289	1410	1600	667	686	1746

### Text Preprocessing

*Preprocessing* adalah tahapan yang didalamnya terdapat beberapa proses penghapusan kata yang tidak diperlukan atau tidak penting dalam proses klasifikasi teks (Sidiq et al., 2020). *Preprocessing* dilakukan bertujuan untuk mengubah data yang tidak terstruktur menjadi data terstruktur sehingga dapat mempermudah dalam pemrosesan data (Pramukti et al., 2022). Terdapat beberapa tahap pada *preprocessing* yaitu: *tokenized*, *transform cases*, *filter tokens*, *filter stopword*, dan *stemming* (Yanis, 2018).

#### 1. Transform Case

proses ini merupakan tahap untuk merubah bentuk semua kata yang ada pada teks menjadi *lowercase* atau huruf kecil (Saleh & Nasari, 2018).

Tabel 6 Kalimat setelah Transform Case

Sebelum Proses <i>Transform Case</i>	Sesudah <i>Transform Case</i>
Oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia, yaitu kompromi.	oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia, yaitu kompromi.
Anies bukan salah data Tapi emang niat berbohong	anies bukan salah data tapi emang niat berbohong

#### 2. Tokenize

*Tokenized* merupakan tahap pemecahan sekumpulan karakter dalam suatu teks ke dalam satuan kata, tokenisasi menghilangkan *delimiter* contohnya seperti tanda titik, koma, spasi, dan karakter angka yang ada pada teks tersebut (Purwanto et al., 2019).

Tabel 7 Kalimat setelah *Tokenize*

Sebelum Proses <i>Tokenize</i>	Sesudah <i>Tokenize</i>
oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia yaitu kompromi	“oleh”, “sebab”, “itulah”, “menurut”, “prabowo”, “dunia”, “harus”, “belajar”, “pada”, “tradisi”, “asia”, “yaitu” “kompromi”
anies bukan salah data tapi emang niat bohong	“anies”, “bukan”, “salah”, “data”, “tapi”, “emang”, “niat”, “berbohong”

#### 3. Filtering

*Filter tokens*, merupakan tahap menghilangkan kata-kata yang banyak muncul tetapi tidak punya pengaruh ke dalam proses ekstraksi klasifikasi teks, pada proses ini kata-kata yang akan dihilangkan contohnya seperti kata penunjuk waktu, tanya dan sambung (Muzaki & Witanti, 2021).

#### 4. *Stopwords*

*Stopwords* merupakan proses menghilangkan kata sambung yang terdapat pada dataset (Pramukti et al., 2022). Data *stopwords* yang didapatkan bersumber dari website <https://github.com/masdevid/ID-Stopwords> dengan total data berjumlah 758 kata. Dibawah ini merupakan tahapan yang ada pada proses *stopwords*. Gambar dibawah ini merupakan alur dari proses *stopwords* (Kurniawan & Susanto, 2019).

Tabel 8 Kalimat setelah *Stopwords*

Sebelum Proses <i>Stopwords</i>	Sesudah <i>Stopwords</i>
oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia yaitu kompromi	oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia kompromi
anies bukan salah data agak tapi emang niat berbohong	anies bukan salah data tapi niat berbohong

#### 5. *Stemming*

*Stemming*, merupakan proses perubahan bentuk suatu kata menjadi bentuk kata dasar, pada proses *stemming* bertujuan untuk menghilangkan imbuhan-imbuhan seperti prefix (ber-), sufiks (-an, -kan, -i), maupun konfiks yang ada pada masing-masing kata (Wirayasa et al., 2019).

Tabel 9 Kalimat setelah *Stemming*

Sebelum Proses <i>Stemming</i>	Sesudah <i>Stemming</i>
oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia kompromi	oleh sebab itulah menurut prabowo, dunia harus belajar pada tradisi asia kompromi
anies bukan salah data tapi niat berbohong	anies bukan salah data tapi niat bohong

#### *Evaluasi*

Untuk mengetahui hasil evaluasi, pada penelitian ini menggunakan metode *Cross Validation* dengan nilai  $K=10$  yang nantinya akan diukur dengan *confusion matrix* untuk mengetahui hasil *akurasi*, *precision*, dan *recall*.

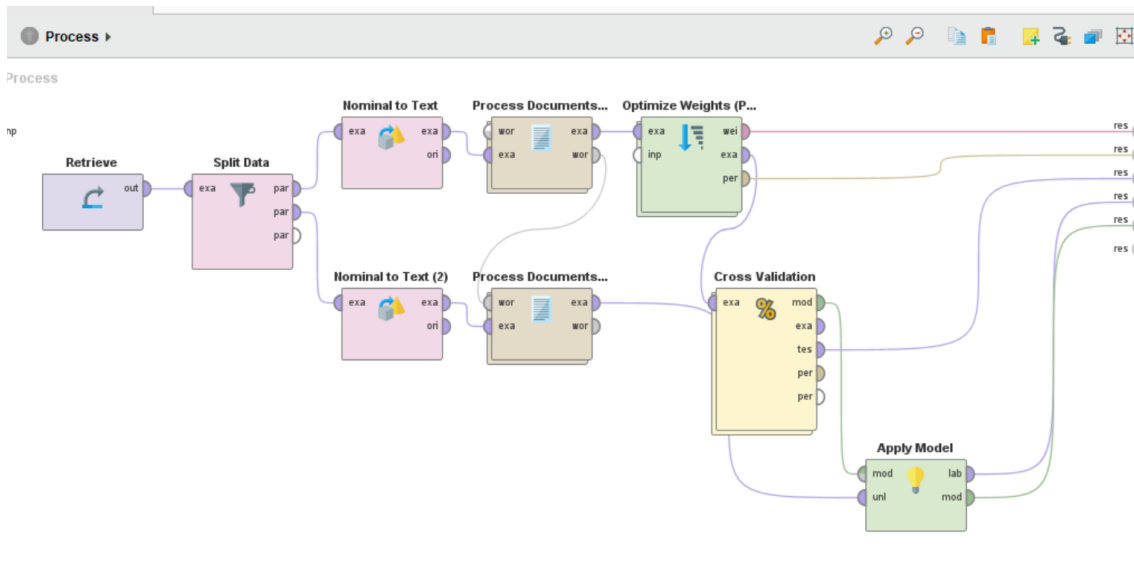
## IV. HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan menggunakan perangkat lunak *RadpidMiner Studio* versi 9.9.0. Di bawah ini adalah hasil pengujian yang dilakukan.

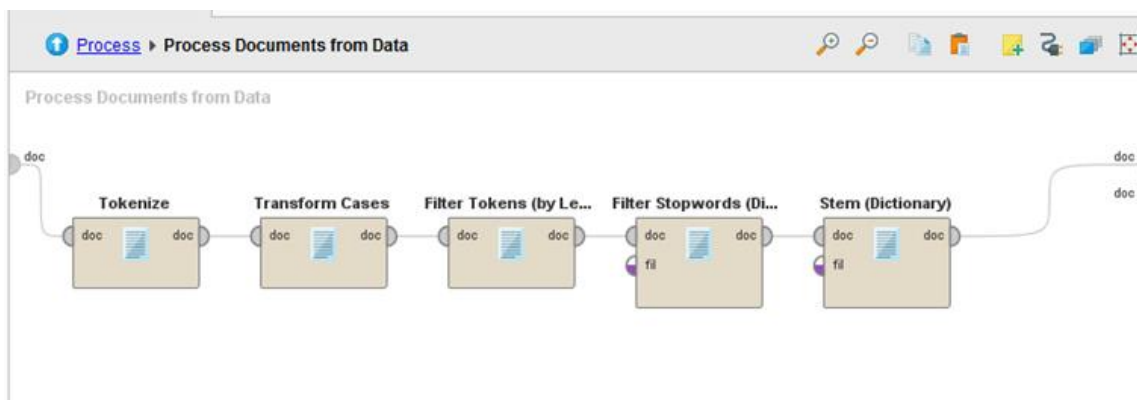
### Pengujian

Pada tahap ini menggunakan metode *Naive Bayes* dan pemilihan fungsi “*Particle Swarm Optimization*” untuk tahap training atau training. Pada metode tersebut diuji sebanyak tiga kali pada dataset calon presiden untuk mengetahui hasil perbandingan akurasi, presisi dan recall masing-masing calon presiden. Pada proses pengujian pertama dilakukan pengujian terhadap Anies Baswedan, pada proses pengujian kedua dilakukan pengujian terhadap Prabowo Subianto, dan pada proses pengujian ketiga dilakukan pengujian terhadap Ganjar Pranowo.

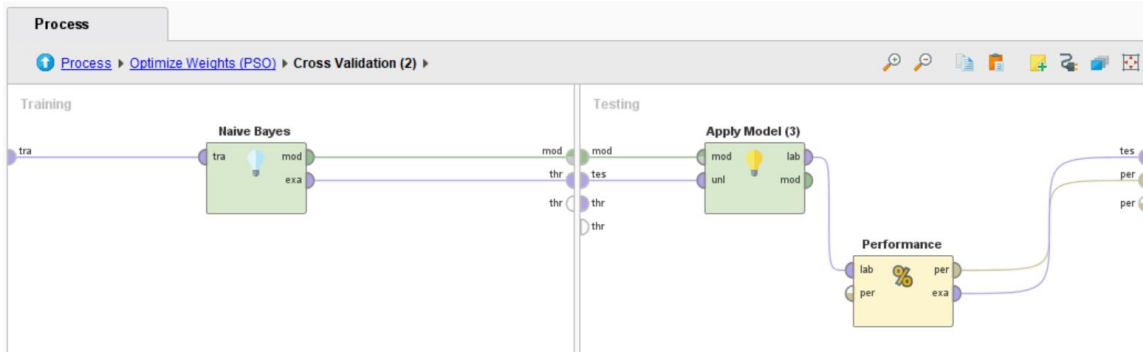
Proses pengujian pertama, kedua dan ketiga dilakukan dengan menggunakan *Naive Bayes* dan *Particle Swarm Optimization*. Gambar 5 adalah tahapan pengujian metode, Gambar 6 adalah dokumen proses, dan Gambar 7 adalah model *cross validation* yang terdiri dari dua bagian yang terpisah yaitu bagian pengujian (*testing*) dan bagian pelatihan (*training*). Bagian pelatihan menggunakan metode *naive bayes*, dan bagian pengujian digunakan untuk operator *apply model* dan operator *performance*, yang menunjukkan *confusion matrix*, yang kemudian digunakan untuk menghitung hasil *akurasi*, *presisi*, dan *recall*.



Gambar 5 Proses pengujian *Naive Bayes* dan *PSO*



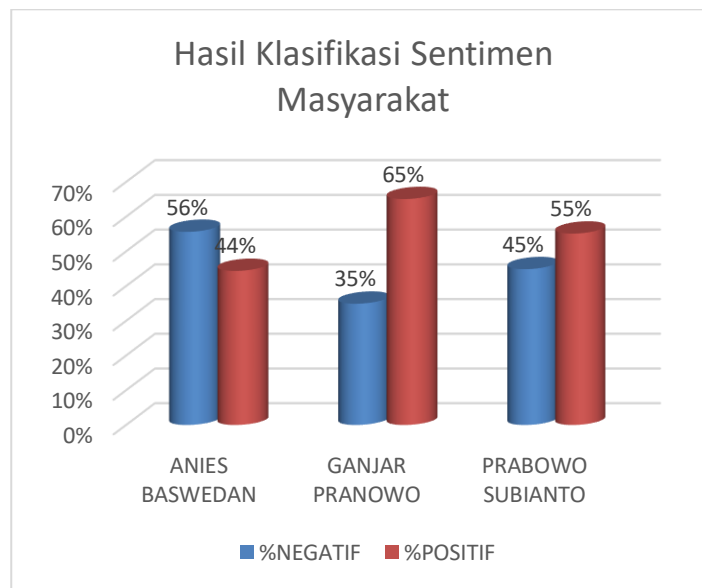
Gambar 6 Proses *Document from data*



Gambar 7 Pemodelan *Cross Validation* pada metode *Naive Bayes* dan *PSO*

**Hasil Klasifikasi Sentimen**

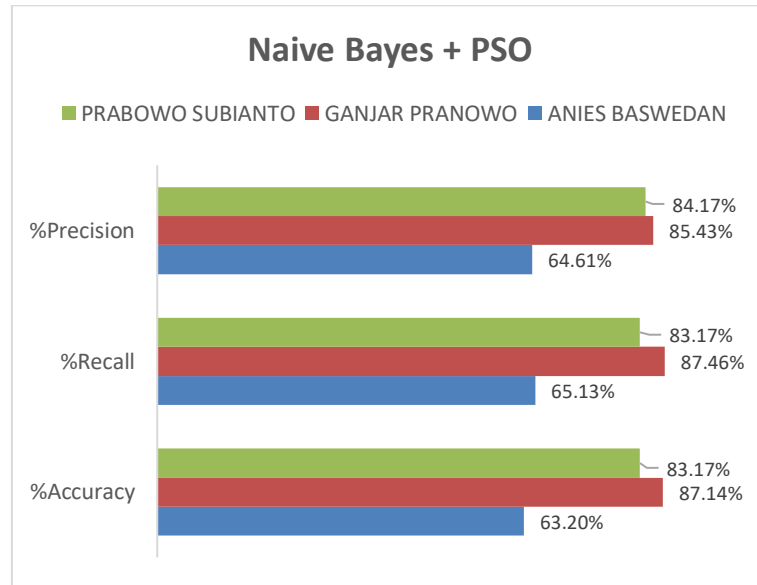
Hasil klasifikasi sentimen masyarakat terhadap Capres 2024, data yang diperoleh berasal dari data latih yang sebelumnya diuji menggunakan metode *naive bayes* dan *PSO*. Di bawah ini adalah grafik batang (kolom) hasil klasifikasi *sentiment* masyarakat terkait capres 2024.



Gambar 8 Hasil Klasifikasi Sentimen Masyarakat

**Hasil Analisis Sentimen**

Pengujian dilakukan dengan menggunakan teknik *cross validation* dengan nilai  $k=10$ . Di bawah ini adalah hasil dari *confusion matrix* saat menggunakan metode *Naive Bayes* di perangkat lunak *RapidMiner Studio*. Pada penelitian ini dilakukan pengujian sebanyak 3 kali, pengujian pertama, kedua dan ketiga dilakukan dengan menggunakan metode yang sama yaitu *Naive Bayes* dengan pemilihan fitur *Particle Swarm Optimization*. Di bawah ini adalah gambar grafik yang membandingkan hasil *akurasi*, *presisi*, dan *recall* dari 3 dataset yang berbeda.



Gambar 9 Hasil Klasifikasi Sentimen Masyarakat

Tabel 10 Hasil Klasifikasi Sentimen

Hasil	Anies Baswedan	Ganjar Pranowo	Prabowo Subianto
%Accuracy	63.20%	87.14%	83.17%
%Recall	65.13%	87.46%	83.17%
%Precision	64.61%	85.43%	84.17%

Menggunakan *particle swarm optimization* di *Naive Bayes* untuk menganalisis opini publik tentang tiga calon presiden 2024 pada social media twitter mendapatkan nilai *accuracy*, *precision*, dan *recall*. Nilai *accuracy* yang didapatkan Anies Baswedan 63,02%, nilai *recall* yang didapatkan 65,13%, dan nilai *precision* yang didapatkan 64,61%. Nilai *accuracy* yang didapatkan Ganjar Pranowo 87,14%, nilai *recall* yang didapatkan 87,46%, dan nilai *precision* yang didapatkan 85,43%. Nilai *accuracy* yang didapatkan Prabowo Subianto 83,17%, nilai *recall* yang didapatkan 83,17%, dan nilai *precision* yang didapatkan 84,17%.

**V. KESIMPULAN**

Hasil penelitian analisis sentimen menggunakan algoritma *naive bayes* dan *feature selection particle swarm optimization* dengan *cross validation*  $k = 10$  dan pelabelan data menggunakan *method transformers* memudahkan pelabelan otomatis untuk pelabelan ribuan data secara cepat. Dapat Diketahui bahwa dataset anies baswedan 60,85% pengguna *twitter* berkomentar negatif dan 39,15% pengguna *twitter* berkomentar positif, dataset Ganjar Pranowo 57,52% pengguna *twitter* berkomentar negatif dan 42,48% pengguna *twitter* berkomentar positif, dataset Ganjar Prabowo Subianto 48,65% pengguna *twitter* berkomentar negatif dan 51,35% pengguna *twitter* berkomentar positif. Perbandingan hasil dari ketiga kandidat calon presiden, nilai akurasi anies baswedan lebih kecil dari pada kedua kandidat lainnya karena komentar negatif dari anies baswedan lebih banyak daripada kedua kandidat lainnya. Anies baswedan mendapatkan nilai *akurasi* 63,02%, ganjar pranowo sebesar 87,14%, dan prabowo subianto 83,17%.

Banyaknya data mempengaruhi hasil dari analisa sentimen, semakin banyak data akan semakin baik nilai akurasi yang didapat.

## VI. REFERENSI

- Cahyono, Y. (2017). Analisis Sentiment pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency. *Jurnal Informatika Universitas Pamulang*, 2(1), 14. <https://doi.org/10.32493/informatika.v2i1.1500>
- Claussen, J., & Peukert, C. (2019). Obtaining Data from the Internet: A Guide to Data Crawling in Management Research. *SSRN Electronic Journal*, 1–38. <https://doi.org/10.2139/ssrn.3403799>
- Effendi, M. M., Mustofa, Z., & Turmudi, A. (2022). Analisis Sentimen Masyarakat Indonesia Dalam Konflik Rusia-Ukraina Di Twitter. *Bulletin of Information ...*, 3(4), 355–366. <https://journal.fkpt.org/index.php/BIT/article/view/418%0Ahttps://journal.fkpt.org/index.php/BIT/article/download/418/261>
- Hayuningtyas, R. Y., & Sari, R. (2019). Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso. *Jurnal Techno Nusa Mandiri*, 16(1), 37–42. <https://doi.org/10.33480/techno.v16i1.115>
- Kurniawan, I., & Susanto, A. (2019). Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019. *Eksplora Informatika*, 9(1), 1–10. <https://doi.org/10.30864/eksplora.v9i1.237>
- Melani, A. (2023, June). Hasil Survei Elektabilitas Capres Pilpres 2024 Ganjar Pranowo, Anies Baswedan, Prabowo Subianto, Erick Thohir Terupdate Hari Ini. *Liputan6*. <https://www.liputan6.com/pemilu/read/5196559/hasil-survei-elektabilitas-capres-pilpres-2024-ganjar-pranowo-anies-baswedan-prabowo-subianto-erick-thohir-terupdate-hari-ini>
- Muzaki, A., & Witanti, A. (2021). Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier. *Jurnal Teknik Informatika (Jutif)*, 2(2), 101–107. <https://doi.org/10.20884/1.jutif.2021.2.2.51>
- Pramukti, S. D., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization. *Techno.Com*, 21(1), 61–74. <https://doi.org/10.33633/tc.v21i1.5332>
- Prayoga, K. J., Nugroho, A., Tri, D., & Wiyatno, N. (2019). Komparasi Feature Selection Particle Swarm Optimization (Pso) Dengan Genetic Algorithm (Ga) Terhadap Algoritma Naïve Bayes Pada Analisis Sentiment Twitter. *E-Prosiding SNasTekS*, 1(1), 299–314. <https://journal.unusida.ac.id/index.php/snts/article/view/99>
- Purwanto, C. N., Hermawan, A. T., Santoso, J., & Gunawan. (2019). Distributed Training for Multilingual Combined Tokenizer using Deep Learning Model and Simple Communication Protocol. *2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019*. <https://doi.org/10.1109/ICORIS.2019.8874898>
- Ramadhan, K., & L, K. M. (2018). Analisis Sentimen Terhadap Toko Online Menggunakan Naïve Bayes pada Media Sosial Twitter. *E-Proceeding of Engineering*, 5(3), 8141–8151.
- Saleh, A., & Nasari, F. (2018). Penggunaan Teknik Unsupervised Discretization pada Metode Naive Bayes dalam Menentukan Jurusan Siswa Madrasah Aliyah. *Jurnal*

- 
- Teknologi Informasi Dan Ilmu Komputer*, 5(3), 353.  
<https://doi.org/10.25126/jtiik.201853705>
- Sidiq, R. P., Dermawan, B. A., & Umaidah, Y. (2020). *Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes*. 5(3), 356–363.
- Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, 14(2), 1–16. <https://doi.org/10.4018/IJHISI.2019040101>
- Sukma, P., & Wardhani, N. (2018). *Jurnal Pendidikan Ilmu-Ilmu Sosial Partisipasi Politik Pemilih Pemula dalam Pemilihan*. 10(1), 57–62.
- Wirayasa, I. P. M., Wirawan, I. M. A., & Pradnyana, I. M. A. (2019). Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani Untuk Stemming Teks Bahasa Bali. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 8(1), 60. <https://doi.org/10.23887/janapati.v8i1.13500>
- Wongkar, M., & Angdresey, A. (2021). *Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter. October 2019*, 3–8. <https://doi.org/10.1109/ICIC47613.2019.8985884>
- Yanis, R. yana. (2018). Sentiment Analysis of Bpjs Kesehatan Services To Smk Eklesia and Bina Insani Jailolo Teachers. *Jurnal Terapan Teknologi Informasi*, 2(2), 113–122. <https://doi.org/10.21460/jutei.2018.22.105>