

Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes



Risha Nur Mauliza¹, Yoannes Romando Sipayung²
Fakultas Komputer Dan Pendidikan, Universitas Ngudi Waluyo, Indonesia
E-mail: rshmauliza@gmail.com , yoannesromando@unw.ac.id

Notifikasi Penulis
28 Desember 2023
Akhir Revisi
05 Februari 2024
Terbit
01 Juni 2024

Mauliza, R. N., & Sipayung, Y. R. Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes: Application of Text Mining in Analyzing Public Opinions on the 2024 Election on Social Media X Using the Naive Bayes Method. *Technomedia Journal*, 9(1).

<https://doi.org/10.33050/tmj.v9i1.2212>

ABSTRACT

Text mining is a process to utilize the ocean of data in the Industrial Age 4.0. The rapid growth in the use of social media has generated a lot of data in the form of text analysis, one of which is sentiment analysis. This research uses social media X in analyzing the sentiment of opinions about the 2024 election. This analysis was taken from X social media user comments as much as 300 review data divided into 2 categories, namely 100 training data and 200 test data, then tested using the naïve bayes method. The text mining method with the naïve bayes algorithm can be applied to analyze public opinion and sentiment towards the 2024 election on the X social media platform. The results of data testing with the naïve bayes method obtained results with the acquisition of 103 positive sentiments, 47 negative sentiments and 50 neutral sentiments.

Keywords: Text mining, X social media, Election 2024, Naïve bayes.

ABSTRAK

Text mining merupakan sebuah proses untuk memanfaatkan lautan data di Era Industri 4.0. Pertumbuhan yang sangat pesat dalam penggunaan media sosial telah menghasilkan banyak data berupa text analisis salah satunya adalah analisis sentimen. Pada penelitian ini menggunakan media sosial X dalam menganalisis sentimen pendapat tentang pemilu 2024. Analisis ini diambil dari komentar pengguna media sosial X sebanyak 300 data ulasan yang dibagi menjadi 2 kategori yaitu 100 data latih dan 200 data ujikemudian dilakukan pengujian menggunakan metode naïve bayes. Metode text mining dengan algoritma naïve bayes dapat diterapkan untuk melakukan analisis pendapat dan sentimen masyarakat terhadap pemilu 2024 di platform media sosial X. Hasil pengujian data dengan metode naïve bayes didapati hasil dengan perolehan 103 sentimen positif, 47 sentimen negatif dan 50 sentimen netral.

Kata Kunci : Text mining, Media sosial X, Pemilu 2024, Naïve bayes.



PENDAHULUAN

Pemilu adalah singkatan dari Pemilihan Umum, yang merupakan suatu proses di mana warga negara dalam suatu negara memilih pemimpin atau wakil mereka dalam pemerintahan dan lembaga legislative [1]. Pemilu juga merupakan salah satu aspek fundamental dari sistem demokrasi, di mana warga negara memiliki hak untuk memilih sebagai bentuk partisipasi dalam proses pengambilan keputusan politik [2]. Pemilu tidak hanya menjadi arena bagi masyarakat untuk menyalurkan hak politiknya, tetapi juga menjadi refleksi dari partisipasi aktif dan pemahaman masyarakat terhadap proses demokratis [3]. Dalam konteks di Negara Indonesia, Pemilu Presiden 2024 dianggap sebagai peristiwa politik yang sangat signifikan, dimana masyarakat memiliki peran penting dalam menentukan kepala negara atau presiden mereka di tingkat nasional .

Dalam era digital seperti saat ini, internet dan media sosial menjadi sarana utama bagi masyarakat untuk berbagi informasi, pendapat, dan pandangan terkait pemilu [4], [5]. Jumlah besar data teks yang dihasilkan oleh masyarakat melalui platform-media sosial seperti X, Facebook, dan Instagram menciptakan potensi besar untuk memahami dinamika opini publik terkait Pemilu 2024 [6]. Namun, mengelola dan menganalisis volume besar data teks ini secara manual menjadi tugas yang sangat rumit [7]. Oleh karena itu, penerapan teknik text mining menjadi solusi yang relevan dan inovatif dalam menghadapi tantangan tersebut [8]. Text mining, juga dikenal sebagai text data mining atau text analytics, adalah suatu proses ekstraksi informasi berharga dan pengetahuan yang terkandung dalam dokumen atau teks [9]. Tujuan utama dari text mining adalah untuk mengubah teks yang tidak terstruktur menjadi bentuk yang dapat diolah oleh komputer sehingga dapat dilakukan analisis lebih lanjut [10]. Teknik ini memungkinkan analisis otomatis terhadap data teks untuk mengekstrak sentimen dan pendapat masyarakat terkait peristiwa Pemilu 2024 [11].

Adapun salah satu algoritma yang digunakan untuk text mining adalah Naïve Bayes Classifier [12]. Naïve Bayes Classifier merupakan algoritma untuk pengklasifikasian dengan metode probabilitas dan statistic untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [13]. Algoritma Naïve Bayes menggunakan metode probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset. Menurut Nursyi'ah 2021 keuntungan penggunaan Naïve Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (data training) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Alasan menggunakan algoritma Naïve Bayes tersebut karena Naive Bayes merupakan algoritma klasifikasi yang efektif dalam mengategorikan teks. Dalam konteks analisis pendapat masyarakat, di mana data berupa teks (misalnya, komentar atau tweet), Naive Bayes dapat membantu mengidentifikasi dan mengklasifikasikan sentimen atau pendapat yang terkandung dalam teks tersebut [14]. Selain itu Algoritma Naive Bayes dapat bekerja dengan baik meskipun fitur atau kata-kata dalam dataset teks sangat banyak. Ini cocok untuk analisis media sosial di mana data dapat mencakup berbagai topik, isu, dan variasi dalam penggunaan bahasa [12].

Penelitian sebelumnya yang membahas tentang keunggulan algoritma naïve bayes dalam analisis sentiment dilakukan oleh Artanti Inez Tanggraeni yang berjudul “Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes”. Penelitian tersebut mencoba untuk melakukan analisis sentiment terhadap aplikasi Sentuh Tanahku berdasarkan hasil ulasan di Google Play Store. Dari proses analisis sentiment tersebut didapati hasil bahwa aplikasi Sentuh Tanahku cenderung mendapatkan respon positif dari pengguna dengan perbandingan 407 ulasan positif dan 235 ulasan negatif [15]. Dan dari hasil pengujian analisis sentimen menggunakan algoritma Naïve Bayes dengan pembobotan TF-IDF menghasilkan akurasi sebesar 89%, precision sebesar 83%, dan recall sebesar 87%.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan teknik text mining dengan algoritma Naive Bayes Classifier dalam menganalisis pendapat masyarakat terhadap Pemilu 2024 yang terungkap melalui media social X. Alasan menggunakan platform X tersebut karena menyediakan data secara real-time, memungkinkan analisis yang lebih dinamis dan responsif terhadap perubahan sentimen atau opini masyarakat seiring waktu, terutama dalam konteks peristiwa politik seperti Pemilu [16]. Penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan metodologi analisis opini publik, tetapi juga memberikan wawasan mendalam tentang dinamika politik dan pemahaman masyarakat terkait proses demokrasi yang sedang berlangsung. Penelitian ini bertujuan menerapkan metode text mining dengan algoritma Naïve Bayes Classifier untuk mengklasifikasi sentiment dan pendapat masyarakat terhadap peristiwa Pemilu 2024 [17].

METODOLOGI PENELITIAN

Prosedur Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah kumpulan opini berupa tweet terhadap topik mengenai Pemilu 2024. Dataset diambil dalam rentang waktu pada tanggal 01 November – 10 Desember 2023 sejumlah 300 tweet dengan hashtag #pemilu2024, #PemiluIndonesia, #PemiluAmanDamai, #Pilpres2024 [18]. Proses pengumpulan data menggunakan bahasa pemrograman python dengan library tweet-harvest dimuat dalam bentuk format CSV file.

Rapid Miner

Rapid miner adalah perangkat lunak yang bersifat terbuka (open source). Rapid miner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining untuk analisis prediksi. Rapid miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapid miner memiliki kurang lebih 500 operator data mining, termasuk operator input, output, data preprocessing dan visualisasi [19]. Rapid miner sebagai software pengolah data mining menyediakan tool untuk membuat decision tree. Hal ini tentu akan memudahkan dalam membuat decision tree dengan menggunakan rapidminer dibandingkan dengan decision tree secara manual. Berikut sekilas dataset ulasan X pada tabel 1 berikut:

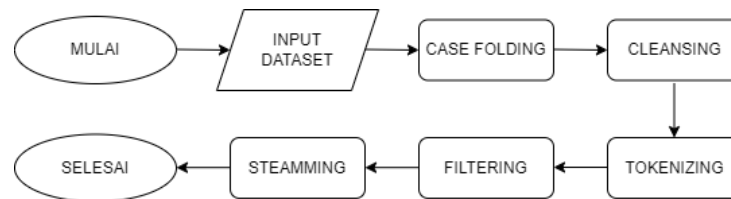
Tabel 1. Ulasan X

created_at	id_str	full_text
Thurs Nov 30 22:03:15 +0000 2023	1,65E+18	Data bisa bocor gitu, berarti privasi pemilu masih sangat minim.

Teknik Analisis Data

Data Preprocessing

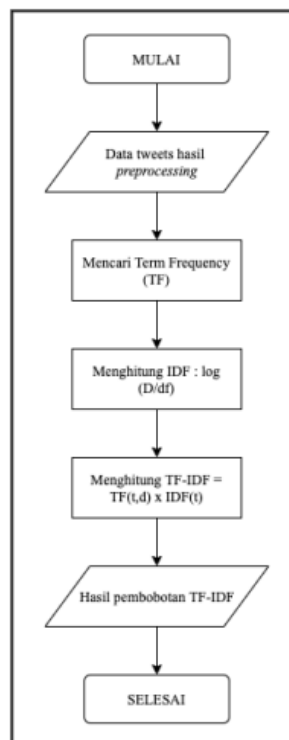
Sebelum masuk dalam proses klasifikasi, kumpulan data tweet harus melalui tahapan preprocessing untuk melakukan transformasi data yang awalnya berbentuk tidak terstruktur atau terdapat banyak noise, menjadi data yang terstruktur sehingga proses analisis menjadi lebih mudah. Beberapa metode yang akan diterapkan dalam proses data preprocessing dapat dilihat pada Gambar 1.



Gambar 1. Data Preprocessing

Pembobotan Term/Kata

Dalam proses ini, kumpulan term/kata akan di vektorisasi atau diubah menjadi bentuk numerik. Hal tersebut perlu untuk dilakukan karena dalam klasifikasi hanya dapat memproses data dalam bentuk numerik saja. Proses pembobotan dalam penelitian ini menggunakan metode Term Frequency-Inverse Document Frequency (TF- IDF). Alur proses pembobotan dengan TF-IDF dapat dilihat pada Gambar 2.



Gambar 2. Pembobotan Kata TF-IDF

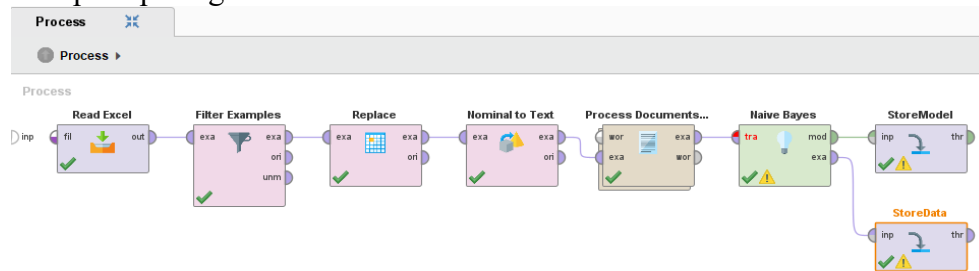
Metode yang Diusulkan

Dalam proses klasifikasi, digunakan algoritma Naïve Bayes Classifier untuk mengenali pola pada data training yang telah diberi label dan kemudian menghasilkan machine learning model [20]. Model yang dihasilkan selanjutnya digunakan untuk mengenali data tweet baru yang belum diberi label sehingga dihasilkan sebuah prediksi kelas sentimennya apakah itu positif, negatif, atau netral .

Cara Pengujian Model

Pengujian model algoritma Naïve Bayes dilakukan menggunakan aplikasi Rapid Miner. Sebelum melakukan pengujian model, terlebih dahulu melakukan proses labeling sentiment terhadap 100 data training untuk mencari model yang sesuai, selanjutnya dilakukan pengujian terhadap 200 data testing untuk mengetahui performa model yang didapatkan menggunakan algoritma Naïve Bayes. Pada Rapid Miner untuk melakukan proses pengujian tersebut

menggunakan operator Naïve Bayes dan menambahkan 2 operator lanjutan yaitu store model dan store data seperti pada gambar 3 berikut.



Gambar 3. Pengujian Naïve Bayes di Rapid Miner

HASIL DAN PEMBAHASAN

Alat Penelitian

Alat penelitian yang digunakan meliputi dua kebutuhan yaitu perangkat keras (Hardware) dan perangkat lunak (Software) diantaranya sebagai berikut :

1. Hardware

Bagian ini akan menjelaskan perangkat keras yang digunakan untuk mendukung penelitian, seperti yang tertera pada tabel 2 dibawah ini

Tabel 2. Kebutuhan Hardware

No	Nama Perangkat	Spesifikasi
1	Laptop	Acer Aspire E5- 471 Processor Intel® Core™ i3-4005U (3MB, 1.7GHz) 64-bit processor
2	Flashdisk USB	16 Gb
3	Handphone	Samsung Galaxy A50 64 GB

2. Software

Bagian ini akan menjelaskan mengenai perangkat lunak (software) yang digunakan untuk mendukung penelitian, seperti yang tertera pada tabel 3 dibawah ini

Tabel 3. Kebutuhan Software

No	Nama Perangkat	Spesifikasi
1	Operating Sistem	Windows 10
2	Bahasa Pemrograman	Python
3	Tools Analisa	Rapid Miner
4	Data Penelitian	Ulasan twitter

Implementasi dan Pengujian

Pengumpulan Data

Pada penelitian ini menggunakan dataset ulasan twitter mengenai topic Pemilu 2024 menggunakan library python bernama tweet-harvest. Dataset diambil dalam rentang waktu pada tanggal 01 November – 10 Desember 2023 sejumlah 300 tweet dengan hashtag

#pemilu2024, #PemiluIndonesia, #PemiluAmanDamai, #Pilpres2024 Berikut beberapa tampilan kode pemrograman beserta hasilnya pada gambar 4 dan 5 berikut

```
[ ] #title Twitter Auth Token
twitter_auth_token = '49fb0124e9661b6deee25327c7fe54e088e31195'

[ ] # Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo tee /etc/apt/sources.list.d
!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v

# Crawl Data
filename = 'topikpemilu2024.csv'
search_keyword = 'Pemilu 2024'
limit = 300

!npx --yes tweet-harvest@2.2.0 -o "{filename}" -s "{search_keyword}" -l {limit} --token {twitter_auth_token}
```

Gambar 4. Tampilan Code Scrape

```
import pandas as pd

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation
0	Thu Dec 14 06:54:50 +0000 2023	1735191598333837161	Wujudkan pemilu damai tanpa hoax #pemiludamai	0	0	0	0	in	1602297929636749318	1735191598333
1	Thu Dec 14 06:54:31 +0000 2023	1735191521153187928	Forum Badan Eksekutif Mahasiswa (BEM) Jawa Tim...	0	0	0	0	in	1294002970334380032	1735191521153
2	Thu Dec 14 06:54:27 +0000 2023	1735191506041098729	Wujudkan pemilu damai tanpa hoax #pemiludamai	0	0	0	0	in	1602297929636749318	1735191506041

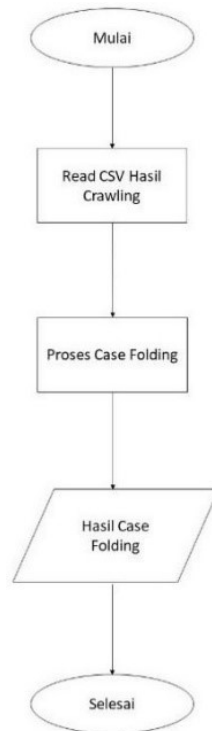
Gambar 5. Hasil Scrape Ulasan Twitter

Data Preprocessing

Data preprocessing adalah proses yang paling penting dalam analisis sentimen, yaitu proses untuk membersihkan data text yang telah dikumpulkan untuk selanjutnya di cari insight-nya. Adapun tahap-tahap preprocessing yang akan diterapkan pada kasus ini adalah sebagai berikut :

1. Case Folding

Tahapan pertama dari preprocessing adalah Case Folding. Pada tahap ini teks ulasan dari hasil crawling data akan diubah menjadi huruf kecil atau diseragamkan dengan mengubah karakter huruf kapital menjadi huruf kecil [21]. Adapun operator yang akan digunakan dalam proses ini adalah operator transform case dari ekstension Text Preprocessing pada Rapid Miner [22]. Setiap ulasan akan masuk satu persatu pada operator dan akan diproses tiap baris data dan hasil case folding kemudian disimpan dengan menggantikan data sebelumnya. Adapun flowchart case folding adalah sebagai berikut.

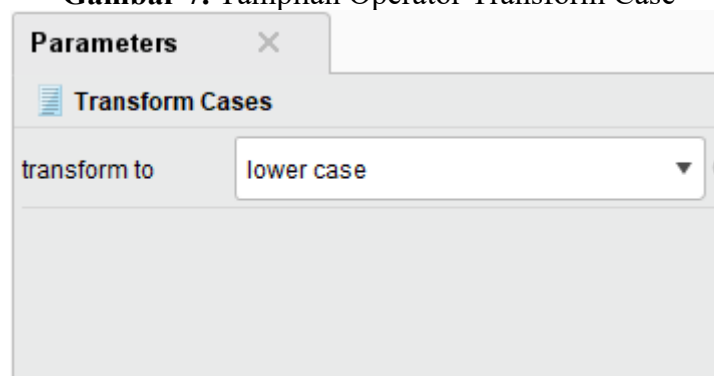


Gambar 6. Alur Case Folding

Berikut adalah tampilan operator transform case pada rapid miner



Gambar 7. Tampilan Operator Transform Case



Gambar 8. Transform Case Lower Case

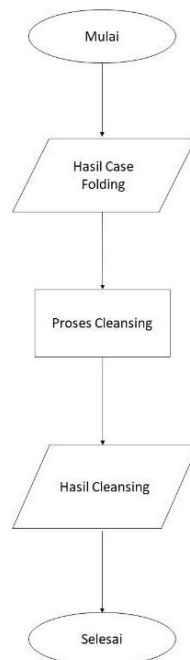
Dari proses case folding diatas maka diperoleh hasil yang disimpan menggantikan data ulasan sebelumnya. Adapun beberapa hasil dari proses case folding ditampilkan dalam table dibawah ini.

Tabel 4. Hasil Proses Case Folding

Sebelum Casefolding	Sesudah Casefolding
Tuyul anggaran Kemenhan! Wowo adalah dalang pengadaan alutsista untuk kepentingan Politik	Tuyul anggaran kemenhan! wowo adalah dalang pengadaan alutsista untuk kepentingan politik

2. Cleaning

Cleansing merupakan proses untuk menghapus URL, @mention, #hashtag, link, double whitespaces, emoticon, serta angka. Karakter-karakter ini dihapus karena dianggap sebagai noise dalam data dan tidak diperlukan dalam alanisis sentimen [23]. Adapun operator yang digunakan adalah operator Replace pada Rapid Miner. Flowchart proses cleansing adalah sebagai berikut.

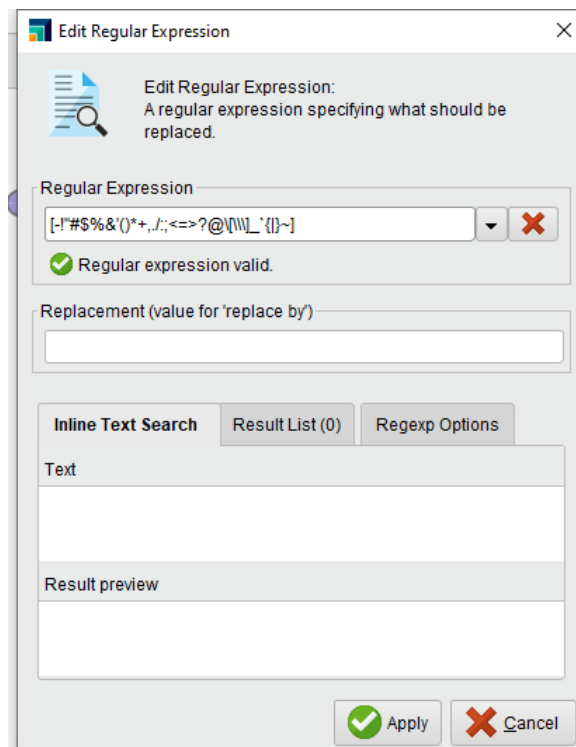


Gambar 9. Alur Cleansing

Berikut gambaran pada proses cleansing menggunakan operator Replace



Gambar 10. Operator Replace Rapid Miner



Gambar 11. Tampilan Edit Regular Expression

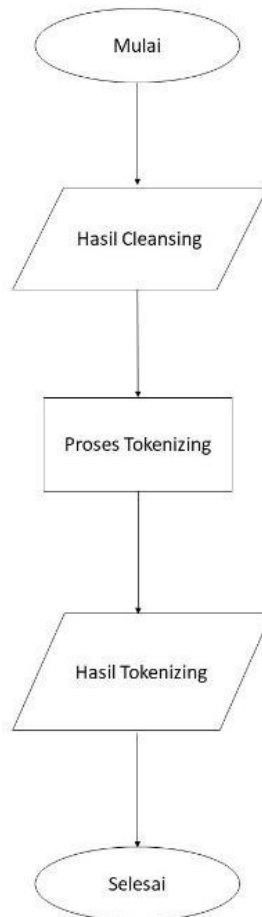
Data tiap baris nya akan masuk pada proses ini dan kemudian diproses satu persatu dan kemudian hasil cleansing akan menggantikan data hasil casefolding sebelumnya. Adapun beberapa hasil cleansing ditampilkan dalam tabel dibawah ini.

Tabel 5. Hasil Proses Cleansing

Sebelum Cleansing	Sesudah Cleansing
Kenaikan anggaran kemenhan jelang pemilu sangat tidak wajar, sarat akan politisasi #pemilu #pilpres #pemilu2024 #pilpres2024 https://t.co/GzcdcM3Ipm	Kenaikan anggaran kemenhan jelang pemilu sangat tidak wajar sarat akan politisasi pemilu pilpres pemilu2024 pilpres2024

3. Tokenizing

Setelah data ulasan dibersihkan selanjutnya masuk pada tahap Tokenizing. Proses Tokenizing merupakan proses yang dilakukan untuk memotong-motong kalimat atau teks menjadi kata perkata atau token [24]. Proses tokenizing ini menggunakan operator Tokenize pada ekstension Teks Preprocessing Rapid Miner. Adapun flowchartnya sebagai berikut



Gambar 12. Flowchart Tokenizing

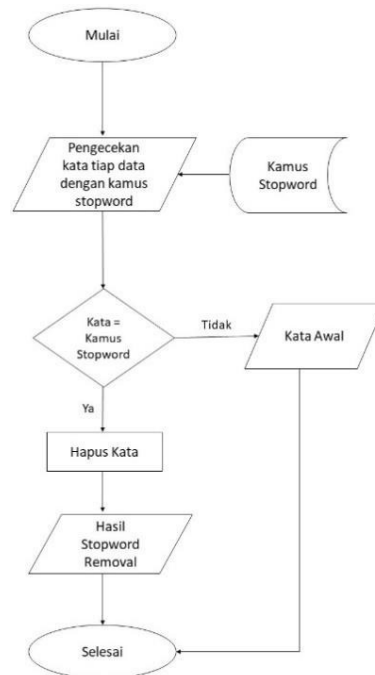
Tiap baris dari data ulasan yang telah diproses pada tahap Tokenizing akan ditampilkan pada tabel berikut

Tabel 6. Hasil Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
Tuyul anggaran kemenhan wowo adalah dalang pengadaan alutsista untuk kepentingan politik	['tuyul', 'anggaran', 'kemenhan', 'wowo', 'adalah', 'dalang', 'pengadaan', 'alutsista', 'untuk', 'kepentingan', 'politik']

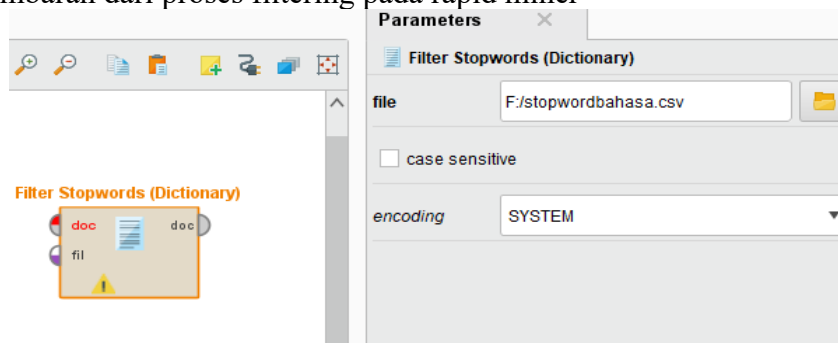
4. Filtering

Setelah kalimat di pecah-pecah menjadi kata perkata maka selanjutnya adalah proses Filtering dengan operator Stopword [25]. Tahap *stopword removal* adalah proses untuk menghapus kata-kata yang tidak penting atau tidak memberikan pengaruh besar terhadap proses klasifikasi nantinya. Contoh-contoh kata yang akan dihapus diantaranya adalah kata hubung 'dan', 'atau', 'ke', kemudian juga nama-nama bulan seperti 'jan', 'feb', 'mar', dan masih banyak lainnya [26]. Tahapan *stopword removal* ini akan menggunakan 2 metode, yaitu yang pertama dengan menggunakan kamus stopwords yang didownload lewat situs <https://www.kaggle.com/oswinrh/indonesian-stoplist>, yang kedua adalah menggunakan operator tambahan yang bernama Filter Token By Length yaitu dengan cara menghapus suatu kata yang memiliki chars huruf kurang dari 4. Berikut alur flowchart prosesnya.



Gambar 13. Flowchart Filtering

Berikut gambaran dari proses filtering pada rapid miner



Gambar 14. Operator Filtering Rapid Miner

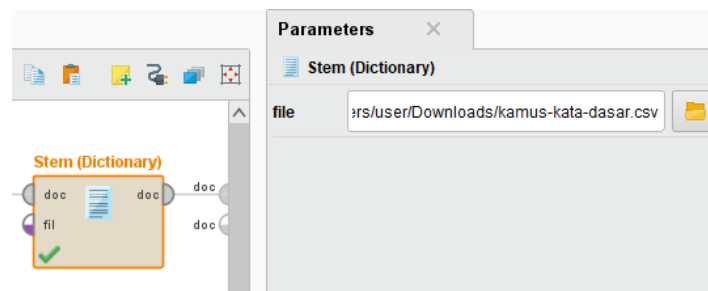
Adapun hasil dari proses ini ditampilkan pada table dibawah ini

Tabel 7. Hasil Filtering

Sebelum Filtering	Setelah Filtering
imin ikutan ngawur gara-gara	imin ikutan ngawur gabung
gabung sama anies	sama anies

5. Steaming

Tahap Stemming adalah tahap untuk mengubah kata-kata kedalam bentuk dasarnya menggunakan operator Stem Dictionary pada Rapid Miner [27]. Sehingga kata-kata dengan imbuhan akan dikembalikan menjadi kata dasar [14]. Kamus kata dasar untuk stemming tersebut didownload melalui <https://github.com/prasastoadi/serangkai>. Berikut gambaran proses stemming menggunakan Rapid Miner.



Gambar 15. Proses Stemming

Berikut adalah beberapa hasil dari proses Stemming yang ditampilkan dalam bentuk table dibawah ini.

Tabel 8. Hasil Stemming

Sebelum Steamming		Sesudah Steamming	
Budiman	sudjatmiko	Budiman	sudjatmiko anggap
menganggap	sekjen pdip hasto	sekjen pdip hasto	kristiyanto
kristiyanto	tidak sadar diri ketika	tidak sadar diri ketika	tuduh
menuduh	prabowo subianto	prabowo subianto	usaha tiru
berusaha	meniru gaya dari	gaya dari	presiden jokowi
presiden jokowi			

Proses Labelling Sentimen dan Pembagian Data

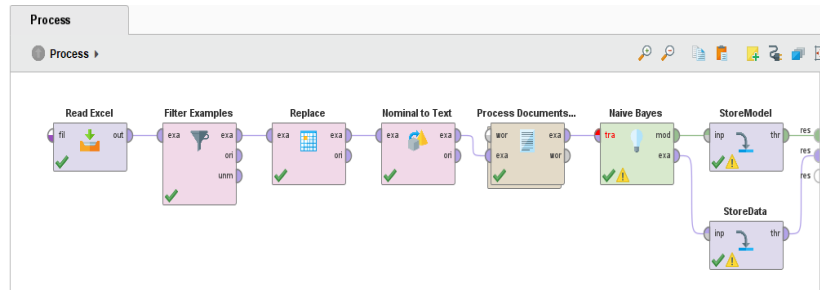
Setelah melewati tahap data preprocessing, dilanjutkan proses pengujian algoritma klasifikasi yaitu Naïve Bayes Classifier [28]. Namun sebelum melakukan proses pengujian algoritma, data yang sudah melewati tahap preprocessing tersebut perlu diberi label sentiment secara manual. Dari 300 data ulasan tersebut, diberikan label sentiment positif dan negative secara manual dan acak pada 100 data dan diberi nama data latih atau data training. Sedangkan sisanya yaitu sebanyak 200 data tidak diberi label sentiment dan diberi nama data testing [29]. Data training nantinya akan digunakan untuk melatih algoritma dalam mencari model yang sesuai, sedangkan data testing akan dipakai untuk menguji dan mengetahui performa model yang didapatkan pada tahapan testing. Berikut gambaran sebagian data yang sudah diberi label sentiment.

Row No.	Sentimen	text	abdul	adil	agysbxe	akar	akarnya	akbar
1	NEGATIF	kenaikan ang...	0	0	0	0	0	0
2	POSITIF	ganjar mahtu...	0	0	0	0	0	0
3	NEGATIF	tuyul anggara...	0	0	0	0	0	0
4	NETRAL	pendaftaran c...	0	0	0	0	0	0
5	NETRAL	memantapka...	0	0	0	0	0	0
6	POSITIF	ganjar mahtu...	0	0	0	0.277	0.277	0
7	POSITIF	menjelang ko...	0	0	0	0	0	0
8	POSITIF	ganjar mahtu...	0	0	0	0	0	0
9	NEGATIF	kampanye na...	0	0	0	0	0	0
10	NETRAL	personil pola...	0	0	0.216	0	0	0
11	POSITIF	anies baswe...	0.364	0	0	0	0	0
12	NETRAL	sahabatbawa...	0	0	0	0	0	0
13	NETRAL	coming soon ...	0	0	0	0	0	0
14	POSITIF	mengenal bo...	0	0	0	0	0	0
15	POSITIF	menjamin ka...	0	0	0	0	0	0

Gambar 16. Data yang Diberi Label

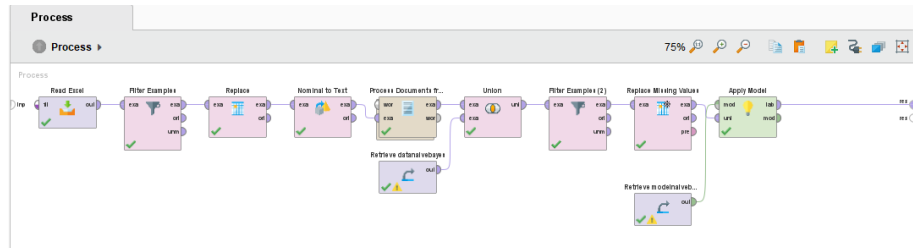
Analisis Sentimen Dengan Algoritma Naïve Bayes Classifier

Setelah melakukan proses labeling sentimen secara manual terhadap 100 data latih untuk mencari model yang sesuai, selanjutnya dilakukan pengujian terhadap 200 data testing untuk mengetahui performa model yang didapatkan menggunakan algoritma Naïve Bayes Classifier [30]. Pada Rapid Miner untuk melakukan proses analisis tersebut menggunakan operator Naïve Bayes dan menambahkan 2 operator lanjutan yaitu store model dan store data. Prosesnya dapat dilihat pada gambar berikut.



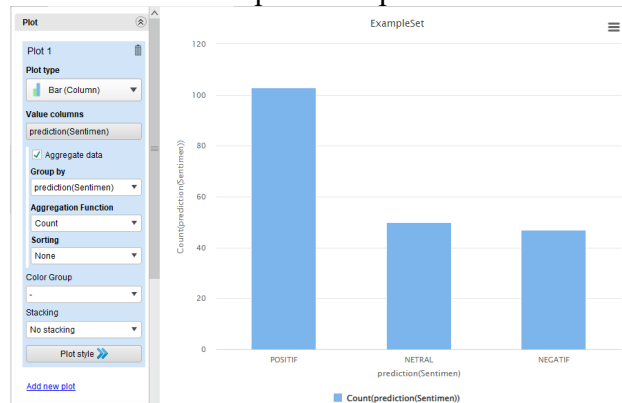
Gambar 17. Proses Pelatihan Algoritma Naïve Bayes

Selanjutnya untuk membuat prediksi sentiment pada 200 data testing dilakukan proses sesuai pada gambar berikut



Gambar 18. Proses Prediksi Sentimen

Dari permodelan diatas kemudian didapati hasil prediksi sentiment sebagai berikut



Gambar 19. Hasil Prediksi

Hasil analisis sentiment menggunakan algoritma Naïve Bayes menunjukkan prediksi bahwa terdapat 103 sentimen positif, 47 sentimen negative, dan 50 sentimen netral.

KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi yang dibuat menggunakan dataset dan algoritma yang diusulkan, maka dapat disimpulkan berdasarkan 300 data ulasan yang dibagi menjadi dua kategori yaitu 100 data latih dan 200 data uji kemudian dilakukan pengujian

menggunakan algoritma Naïve Bayes Classifier, didapati hasil prediksi dengan perolehan 103 sentimen positif, 47 sentimen negative, dan 50 sentimen netral. Dari hasil prediksi tersebut terlihat bahwa masyarakat Indonesia banyak yang menyambut peristiwa Pemilu 2024 secara positif. Banyak dari ulasan tersebut yang mengharapkan agar Calon Presiden dan Wakil Presiden yang terpilih nanti bisa memimpin dan membawa Negara Indonesia ke arah yang lebih baik.

SARAN

Sebuah saran penelitian yang menarik adalah untuk melakukan analisis lebih lanjut tentang perbedaan sentimen antara data latih dan data uji dalam konteks Pemilu 2024. Hal ini dapat memberikan pemahaman yang lebih dalam tentang bagaimana pandangan masyarakat terhadap peristiwa politik tersebut berubah seiring waktu. Selain itu, penting untuk meneliti faktor-faktor yang memengaruhi sentimen masyarakat terhadap Pemilu 2024, seperti isu-isu politik, sosial, ekonomi, dan budaya yang mungkin memainkan peran dalam membentuk opini mereka terhadap calon presiden dan wakil presiden. Selanjutnya, melakukan perbandingan kinerja algoritma Naïve Bayes Classifier dengan metode klasifikasi lainnya seperti Support Vector Machine (SVM), Decision Tree, atau algoritma Deep Learning seperti LSTM (Long Short-Term Memory) dapat memberikan wawasan tambahan tentang keefektifan model yang digunakan. Pengembangan model prediksi yang lebih kompleks menggunakan teknik-teknik NLP yang lebih mutakhir, seperti BERT atau Transformer, juga dapat meningkatkan akurasi dalam mengklasifikasikan sentimen dari ulasan masyarakat. Selain itu, penelitian yang menggali pengaruh media sosial dalam membentuk sentimen masyarakat terhadap Pemilu 2024, serta analisis regional untuk memahami perbedaan persepsi di berbagai wilayah geografis atau demografis, dapat memberikan wawasan yang berharga dalam memahami dinamika politik dan opini publik di Indonesia.

DAFTAR PUSTAKA

- [1] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *Jurnal Bumigora Information Technology (BITE)*, vol. 3, no. 1, pp. 19–27, 2021.
- [2] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," (*JurTI*) *Jurnal Teknologi Informasi*, vol. 3, no. 2, pp. 179–184, 2019.
- [3] W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, "Transferring naive bayes classifiers for text classification," in *AAAI*, 2007, pp. 540–545.
- [4] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis klasifikasi sentimen terhadap sekolah daring pada twitter menggunakan Supervised Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, 2021.
- [5] D. C. U. Lieharyani and R. Ambarwati, "Visualisasi Data Tweet di Sektor Pendidikan Tinggi Pada Saat Masa Pandemi," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 116–123, 2022.
- [6] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, 2022.
- [7] L. H. A. P. Prawira, A. F. Ummah, M. R. Aditiya, and D. W. Nugroho, "Knowledge Management: Efforts to Create an Excellent Digital Creative Industry," *Startuppreneur Business Digital (SABDA Journal)*, vol. 2, no. 2, pp. 172–181, 2023.

-
- [8] A. Felix, S. J. Salim, and J. M. Karsten, "Pemanfaatan Teknologi Layanan Fine Dining untuk Meningkatkan Customer Experience dan Influence Satisfaction," *Technomedia Journal*, vol. 8, no. 3, pp. 91–104, 2024.
 - [9] A. Saepulrohman, S. Saepudin, and D. Gustian, "Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine," *@ is The Best: Accounting Information Systems and Information Technology Business Enterprise*, vol. 6, no. 2, pp. 91–105, 2021.
 - [10] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan k-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)," 2018.
 - [11] J. Sihombing, "Klasifikasi Data Antropometri Individu Menggunakan Algoritma Naïve Bayes Classifier," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, pp. 1–10, 2021.
 - [12] D. S. Budi and H. Syahrial, "Optimizing Database Performance in the Data Transformation Process in SQL Server: Pengoptimalan Performa Database Pada Proses Transformasi Data Pada SQL Server," *Technomedia Journal*, vol. 8, no. 3 Februari, pp. 407–419, 2024.
 - [13] Y. Yang, H. Wang, R. Jiang, X. Guo, J. Cheng, and Y. Chen, "A review of IoT-enabled mobile healthcare: technologies, challenges, and future trends," *IEEE Internet Things J*, vol. 9, no. 12, pp. 9478–9502, 2022.
 - [14] H. Safitri, M. H. R. Chakim, and A. Adiwijaya, "Strategy Based Technology-Based Startups to Drive Digital Business Growth," *Startupreneur Business Digital (SABDA Journal)*, vol. 2, no. 2, pp. 207–220, 2023.
 - [15] D. Tuhenay, "Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM)," *JIKO (Jurnal Informatika dan Komputer)*, vol. 4, no. 2, pp. 105–111, 2021.
 - [16] N. Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *Jurnal Teknik Komputer AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022.
 - [17] F. Wulandari, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 88–96, 2023.
 - [18] B. Rawat and R. Bhandari, "Cloud Computing Applications In Business Development," *Startupreneur Business Digital (SABDA Journal)*, vol. 2, no. 2, pp. 143–154, 2023.
 - [19] L. South, D. Saffo, O. Vitek, C. Dunne, and M. A. Borkin, "Effective use of Likert scales in visualization evaluations: A systematic review," in *Computer Graphics Forum*, Wiley Online Library, 2022, pp. 43–55.
 - [20] F. F. Nugraha and C. Juliane, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4. 5," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, pp. 2862–2869, 2022.
 - [21] S. Purnama, R. Widayanti, P. Edastama, and A. R. S. Panjaitan, "Bantuan Sosial Untuk Masyarakat Yang Terkena Dampak Ekonomi Selama Covid-19 di Garut," *ADI Pengabdian Kepada Masyarakat*, vol. 2, no. 1, pp. 67–72, 2021.
 - [22] R. Salam, H. Nusantara, T. Ayuninggati, V. T. Devana, and A. P. Candra, "Peran Serta Dalam Melaksanakan Pembagian Makanan di Wilayah Bsd City," *ADI Pengabdian Kepada Masyarakat*, vol. 2, no. 1, pp. 62–66, 2021.
 - [23] N. Adiyanto, "Customer Relationship Management (CRM) Based On Web To Improve The Performance Of The Company," *ITSDI Journal Edition Vol. 1 No. 1 October 2019*, p. 32, 2019.
-

- [24] N. K. Purnamawati, A. M. Adiandari, N. D. A. Amrita, and L. P. V. I. Perdanawati, "The Effect Of Entrepreneurship Education And Family Environment On Interests Entrepreneurship In Student Of The Faculty Of Economics, University Of Ngurah Rai In Denpasar," *ADI Journal on Recent Innovation (AJRI)*, vol. 1, no. 2 Maret, pp. 158–166, 2020.
- [25] J. Sihombing, "Klasifikasi Data Antropometri Individu Menggunakan Algoritma Naïve Bayes Classifier," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, pp. 1–10, 2021.
- [26] S. H. Hardi and K. D. Hartomo, "Sentiment Analysis of Simobi Plus Mobile Application Using Naïve Bayes Classification," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, pp. 1117–1124, 2023.
- [27] B. Pernama and H. D. Purnomo, "Analisis Risiko Pinjaman dengan Metode Support Vector Machine, Artificial Neural Network dan Naïve Bayes," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 1, pp. 92–99, 2023.
- [28] S. Y. Nursyi'ah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," in *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, 2021, pp. 117–123.
- [29] K. A. Baihaqi *et al.*, "A Comparison Support Vector Machine, Logistic Regression And Naïve Bayes For Classification Sentimen Analisis user Mobile App," *International Journal of Artificial Intelligence Research*, vol. 7, no. 1, pp. 64–71, 2023.
- [30] H. S. Hopipah and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes," *Technomedia Journal*, vol. 6, no. 1, pp. 99–110, 2021.