

Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes

Ramadani Saputra^a, Firman Noor Hasan^b

^aTeknik Informatika, Fakultas Teknologi Industri Dan Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA, ramadanis6690@gmail.com

^bTeknik Informatika, Fakultas Teknologi Industri Dan Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA, firman.noorhasan@uhamka.ac.id

Submitted: 03-05-2024, Reviewed: 10-05-2024, Accepted 09-06-2024
<https://doi.org/10.47233/jteksis.v6i3.1378>

Abstract

The high prevalence of stunting in Indonesia, reaching 21.6% in 2022, has become an urgent concern requiring practical solutions. Stunting is a serious consequence of chronic malnutrition during early growth stages. It leads to physical, psychological, and cognitive impairments, hindering brain development, physical growth, and increasing the risk of metabolic disorders. To address this challenge, Candidate Pair number 02, Prabowo-Gibran, in the 2024 Presidential Election, proposed a Free Lunch and Milk Program. This research aims to investigate Indonesian society's views on the program through Twitter, analyzing opinions that have not been previously studied using the Naïve Bayes classification method via the RapidMiner application. Data collection was divided into two stages, before and after the presidential election. A total of 785 tweets were used as test data to identify public sentiment towards the program using the TF-IDF method. SMOTEUP sampling was employed to address class imbalance in the data, and model evaluation was conducted using 10-fold cross-validation. The research results indicate that approximately 81.7% of the analyzed Twitter data showed negative, 6.6% neutral, and 11.7% positive sentiment. The classification model demonstrated an accuracy rate 92.96%, recall 85.30%, and precision 94.57%, indicating a good performance in classifying sentiment from the test data.

Keywords: sentiment analysis, naïve bayes, free lunch and milk, rapidminer

Abstrak

Masalah tingginya angka stunting di Indonesia, mencapai 21,6% pada tahun 2022, telah menjadi perhatian mendesak yang membutuhkan solusi praktis. Stunting adalah dampak serius dari kurang gizi kronis pada masa pertumbuhan awal. Hal ini menyebabkan gangguan fisik, psikologis, dan kognitif serta menghambat perkembangan otak dan pertumbuhan fisik serta meningkatkan risiko gangguan metabolik. Untuk mengatasi tantangan ini, Pasangan Calon nomor urut 02 Prabowo-Gibran dalam Pemilihan Presiden 2024 mengusulkan Program Makan Siang dan Susu Gratis. Penelitian ini bertujuan untuk menyelidiki pandangan masyarakat Indonesia terhadap program tersebut pada media sosial Twitter, dengan menganalisis opini-opini yang belum pernah diteliti sebelumnya menggunakan metode klasifikasi Naïve Bayes melalui aplikasi RapidMiner. Pengumpulan data dibagi menjadi 2 tahap, sebelum dan sesudah pemilihan presiden. Data uji sebanyak 785 tweet Twitter digunakan untuk mengidentifikasi sentimen masyarakat terhadap program tersebut menggunakan metode TF-IDF. SMOTEUP sampling digunakan untuk mengatasi ketidakseimbangan kelas dalam data dan Evaluasi model dilakukan melalui teknik *cross-validation 10-fold*. Hasil penelitian menunjukkan bahwa sekitar 81,7% dari data Twitter yang dianalisis menunjukkan sentimen negatif, 6,6% bersifat netral, dan 11,7% menunjukkan sentimen positif. Model klasifikasi menunjukkan tingkat *accuracy* sebesar 92,96%, *recall* sebesar 85,30%, dan *precision* sebesar 94,57%, menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen dari data uji.

Keywords: analisis sentimen, naïve bayes, makan siang dan susu gratis, rapidminer

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Saat ini masyarakat dihadapkan dengan perkembangan teknologi informasi dan komunikasi yang cukup pesat, adanya hal ini mendorong masyarakat dalam menggunakan sosial media secara massif [1]. Pada Februari 2024 masyarakat Indonesia dihadapkan dengan pesta demokrasi yang diselenggarakan setiap 5 tahun sekali dimana diantara tiga pasangan calon terdapat satu pasangan calon dengan nomor urut 02 yang memiliki program unggulan yakni Program makan siang gratis dan susu gratis di sekolah dan pesantren, serta bantuan

gizi untuk anak balita dan ibu hamil [2]. Program makan siang dan susu gratis merupakan gagasan pasangan calon 02 yang bertujuan untuk mengatasi masalah stunting (tengkes) yang menjadi isu konkret dan mendesak di Indonesia. Stunting adalah dampak serius dari kurang gizi kronis pada masa pertumbuhan awal. Hal ini menyebabkan gangguan fisik, psikologis, dan kognitif serta menghambat perkembangan otak dan pertumbuhan fisik serta meningkatkan risiko gangguan metabolik [3]. Tingginya angka stunting di Indonesia, mencapai 21,6% di tahun 2022.[4].

Untuk mengatasi hal ini, Program ini menyediakan makan siang harian untuk siswa di tingkat prasekolah, sekolah dasar (SD), sekolah menengah pertama (SMP), sekolah menengah atas (SMA), dan pondok pesantren. Selain itu, program ini juga memberikan dukungan gizi bagi ibu hamil dan anak balita di seluruh Indonesia dengan tujuan untuk meningkatkan kesehatan serta membantu perekonomian keluarga. Diharapkan, program makan siang dan susu gratis ini dapat menjadi solusi dalam mengatasi masalah stunting dan meningkatkan kesejahteraan masyarakat Indonesia secara keseluruhan. Program ini direncanakan akan dilaksanakan secara merata pada tahun 2029, dengan alokasi anggaran sebesar 450 triliun rupiah per tahunnya. Program ini ditujukan untuk memberikan manfaat kepada sebanyak 82,9 juta penerima [5].

Pada tahun 2024, jumlah pengguna internet di Indonesia telah mencapai 221.563.479 jiwa dari total populasi sebanyak 278.696.200 jiwa [6]. Media sosial telah menjadi salah satu platform utama untuk penyebaran informasi, dengan Twitter menjadi salah satu yang paling populer. [7]. Kehadiran media sosial seperti twitter menjadi relevan dalam konteks penyebaran informasi. Dengan jumlah pengguna twitter di Indonesia yang mencapai 10.645.000, platform ini menjadi saluran penting dalam menyebarkan informasi di kalangan masyarakat Indonesia [8].

Analisis sentimen dengan algoritma Naïve Bayes merupakan pendekatan relevan untuk mengidentifikasi pandangan masyarakat terhadap suatu program [9]. Penggunaan data ulasan dari twitter mempermudah dalam memperoleh pemahaman secara mendalam apakah program ini diterima masyarakat. Metode ini dilakukan dengan memproses dan mengekstrak data teks secara otomatis untuk menentukan apakah ulasan cenderung positif, netral atau negatif. Kecepatan dan kesederhanaan algoritma naïve bayes ideal dipilih untuk menganalisis data dalam jumlah besar seperti yang dihasilkan oleh ulasan media sosial, dalam hal ini twitter [10].

Banyak penelitian sebelumnya telah dilakukan terkait analisis sentimen. Salah satu penelitian tersebut adalah analisis program kampus merdeka menggunakan metode Naïve Bayes [11]. Hasil penelitian menunjukkan bahwa sistem berhasil mengklasifikasikan 272 opini sebagai sentimen positif dan 229 opini sebagai sentimen negatif, dengan *accuracy* 60%, *precision* mencapai 64%, *recall* 58% dan *f1-score* 58%. Penelitian sentimen menggunakan metode naïve bayes terhadap pembangunan kereta cepat Jakarta-Bandung di media sosial twitter [12]. Dengan menggunakan 2007 data, menghasilkan temuan sentimen negatif 673, sentimen positif 668, dan hasil sentimen netral

665 hasil *accuracy* 71%, *precision* 73%, *recall* 89%. Penelitian sentimen terhadap Pinjaman Online di media sosial twitter menggunakan metode naïve bayes [13]. Menggunakan 2912 data, menghasilkan Sentimen negatif 68,61%, sentimen positif 31,39% Hasil analisis sentimen pengguna Twitter mengenai Pinjaman Online mencapai *accuracy* 80%.

Penelitian lain sentimen penggunaan *QRIS* (*Quick Respond Code Indonesian Standart*) menggunakan algoritma naïve bayes [14]. sebanyak 913 data menghasilkan 65% sentimen positif dan 35% negatif dengan hasil *accuracy* 99.89%, *precision* 99.83%, dan *recall* 99.68%. Penelitian sentimen pada review pengguna *E-Commerce* menggunakan algoritma naïve bayes [15]. mendapatkan hasil pengujian dengan *accuracy* sebesar 99,5%, *precision* sebesar 99,49%, *recall* sebesar 100%.

Penelitian ini bertujuan untuk mengeksplorasi pandangan masyarakat Indonesia terhadap Program pasangan calon 02 yang menawarkan makan siang dan minum susu gratis. Dengan menggunakan metode Naïve Bayes untuk klasifikasi data, penelitian ini akan menganalisis opini-opini yang belum pernah diteliti sebelumnya. Tujuan utamanya adalah untuk memahami respons masyarakat.

Pengumpulan data yang dilakukan dalam dua tahap penting, yaitu 10 hari sebelum dan 10 hari setelah pemilihan presiden, yang memberikan wawasan komprehensif tentang dinamika sentimen masyarakat dalam periode yang kritis pada media sosial twitter.

METODE PENELITIAN

2.1. Naïve Bayes

Algoritma Naïve Bayes adalah salah satu algoritma dalam teknik klasifikasi yang mengandalkan pada probabilitas dan statistik [16]. Algoritma ini, yang dikenal sebagai Teorema Bayes, berasal dari karya ilmiah Thomas Bayes, seorang ilmuwan Inggris yang merumuskan cara memprediksi probabilitas kejadian di masa depan berdasarkan pengalaman masa lalu [17]. Algoritma ini menggunakan teorema Bayes dengan asumsi bahwa fitur-fitur bersifat independen. Dalam analisis sentimen, Naïve Bayes menghitung kemungkinan suatu teks tergolong dalam kategori sentimen tertentu (positif, negatif, atau netral) berdasarkan kata-kata yang terdapat dalam teks tersebut.

Rumus yang digunakan dalam algoritma Naïve Bayes adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

- $P(C|X)$ adalah probabilitas suatu kelas C diberikan fitur-fitur X .
- $P(X|C)$ adalah probabilitas munculnya fitur-fitur X diberikan kelas C .
- $P(C)$ adalah probabilitas awal dari kelas
- $P(X)$ probabilitas total dari fitur-fitur X .

Dengan menggunakan pendekatan ini, Naïve Bayes dapat secara efektif mengklasifikasikan teks ke dalam berbagai kategori sentimen dengan mempertimbangkan distribusi kata-kata dalam teks yang dianalisis. Keunggulan utama dari algoritma Naïve Bayes adalah kemampuannya untuk menangani dataset besar dan kecepatan dalam proses pelatij dan prediksi. Selain itu, Naïve Bayes bekerja sangat baik dalam situasi di mana asumsi independensi fitur relatif valid, yang seringkali terjadi dalam analisis teks seperti analisis sentimen di media sosial.

2.2 RapidMiner

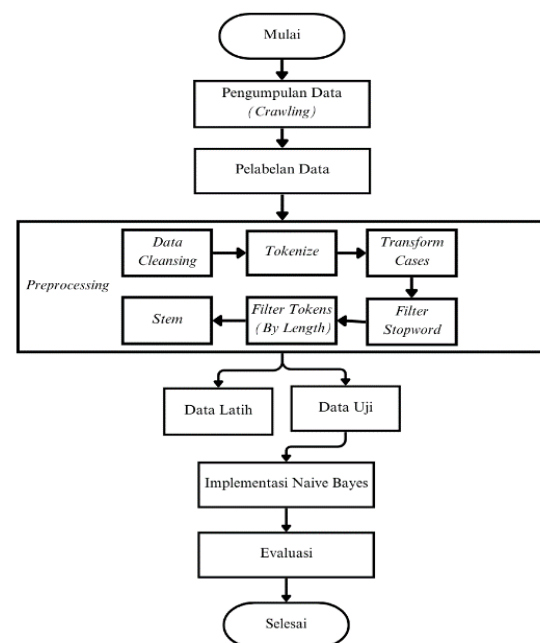
RapidMiner adalah alat data sains sumber terbuka untuk analisis data mining, text mining, dan analisis prediktif [18]. RapidMiner dapat menjadi alat yang berguna dalam menerapkan algoritma Naïve Bayes untuk analisis sentimen, peneliti dapat dengan mudah mengakses algoritma naïve bayes dan menerapkannya pada data dari media sosial twitter untuk menganalisis sentimen terkait program makan siang dan susu gratis. Fitur-fitur utama RapidMiner yang digunakan pada penelitian ini adalah:

- Operator *Process Documents from Data*: Pra-pemrosesan teks termasuk *tokenize*, *transform case*, *filter stopwords*, *filter token*, dan *stemming*
- Operator Naïve Bayes: Inti analisis sentimen, mengklasifikasi data ke dalam kategori positif, negatif, atau netral berdasarkan probabilitas Bayes.
- Operator *Store*: Menyimpan data latih dan uji yang sudah diproses untuk pelatihan dan pengujian model.
- Operator *Union*: Menggabungkan data latih dan uji sebelum pengujian model.
- Operator *Filter Example*: Menyaring data uji sesuai kriteria yang ditentukan.
- Operator *Replace Missing Value*: Menangani nilai data yang hilang.
- Operator *Apply Model*: Menerapkan model Naïve Bayes pada data uji untuk prediksi sentimen.
- Operator *SMOTE Upsampling*: Mengatasi ketidakseimbangan kelas dalam dataset dengan sampel sintetis.

- Operator *Cross-Validation*: Mengevaluasi kinerja model dengan pembagian dataset dan pengujian berulang.
- Operator *Performance*: Mengukur kinerja model dengan *accuracy*, *precision*, dan *recall*.
- Operator *Replace*: Menghilangkan tanda baca, link, dan karakter khusus lainnya dari teks untuk membersihkan data sebelum analisis sentimen.

2.3 Text Preprocessing

Text Preprocessing adalah proses penting dalam pengolahan dokumen teks yang tidak terstruktur menjadi terstruktur [19]. Tujuan dari *text preprocessing* adalah untuk membersihkan dan mempersiapkan data teks agar dapat diolah dengan lebih efektif. Langkah-langkah dalam *text preprocessing* teks meliputi pembersihan data untuk menghilangkan karakter yang tidak relevan atau mengganggu. *Tokenize* dilakukan untuk memecah teks menjadi unit-unit seperti kata atau frasa. *transform cases* menyamakan semua huruf menjadi huruf kecil atau besar. *filter stopwords* menghapus kata-kata umum yang tidak memberikan makna penting, *filter token* menghapus token yang tidak relevan dan *Stemming* mengubah kata-kata menjadi bentuk dasarnya. Dengan preprocessing ini, dokumen teks dipersiapkan untuk analisis lebih lanjut.



Gambar 1. Diagram Alir Penelitian

Diagram alir dalam (Gambar 1) menggambarkan langkah-langkah metodologi yang digunakan dalam penelitian ini. Langkah pertama adalah pengumpulan data melalui platform *Google Colab* dengan menggunakan bahasa pemrograman Python dan *library tweet-harvest*, yang menghasilkan total 1962 data. Selanjutnya, dilakukan pelabelan manual terhadap 1177 data untuk mengkategorikan mereka ke dalam sentimen negatif, netral, dan positif. Proses selanjutnya melibatkan *text preprocessing*, termasuk *data cleansing*, *tokenize*, *transform case*, *filter stopword*, *filter token*, dan *stemming*. Setelah proses *text processing* selesai, data dibagi menjadi dua bagian, yaitu data latih dan data uji, di mana sebanyak 785 data akan digunakan untuk pengujian. Selain itu, pembobotan kata menggunakan metode *TF-IDF* juga dilakukan untuk memberikan bobot pada kata-kata yang relevan dalam analisis sentimen [20]. Implementasi algoritma *naïve bayes* kemudian dilakukan, diikuti dengan tahap evaluasi yang melibatkan penggunaan *Synthetic Minority Over-sampling Technique (SMOTE) Upsampling* untuk menangani ketidakseimbangan kelas dalam dataset [21]. Proses evaluasi juga melibatkan *cross validation* dengan *number of folds 10* untuk memastikan keandalan model [22]. Selain itu, *confusion matrix* digunakan untuk mengevaluasi *accuracy*, *precision*, dan *recall* dari model yang dikembangkan.

HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengumpulan data menggunakan *Google Colab* dengan menggunakan Bahasa pemrograman python serta memanfaatkan *library tweet-harvest*, peneliti membagi menjadi dua tahap pengumpulan data. Pada tahap pertama (Gambar 2).

```
# Pengumpulan Data
filename = 'MSSG_sebelum.csv'
search_keyword = 'makan siang dan susu gratis until:2024-02-13 since:2024-02-04'
limit = 1000
```

Gambar 2. Pengumpulan Data Tahap Pertama

Menggunakan kata kunci "makan siang dan susu gratis *until:2024-02-13 since:2024-02-04*", data Twitter yang tersedia dalam rentang waktu 10 hari sebelum pencoblosan pemilihan presiden dikumpulkan dan disimpan dalam *file CSV* dengan nama *MSSG_sebelum.csv*.

```
# Pengumpulan Data
filename = 'MSSG_setelah.csv'
search_keyword = 'makan siang dan susu gratis until:2024-02-24 since:2024-02-15'
limit = 1000
```

Gambar 3. Pengumpulan Data Tahap Pertama

Selanjutnya pada tahap kedua (Gambar 3). Pengumpulan data dilakukan dengan kata kunci "makan siang dan susu gratis *until:2024-02-24 since:2024-02-15*", mengambil data dalam rentang waktu 10 hari setelah pencoblosan, dan disimpan dalam file CSV dengan nama *MSSG_setelah.csv*. Penggabungan data dari kedua tahap tersebut menghasilkan total 1962 data tweet. Pengumpulan data merupakan langkah krusial dalam melakukan analisis sentimen terhadap program makan siang dan susu gratis Prabowo-Gibran di Twitter. Dengan membagi proses pengumpulan data menjadi dua tahap yang terpisah, dapat mengamati dan memperoleh data dari dua periode waktu yang berbeda, yaitu sebelum dan setelah pemilihan presiden. Ini memberikan gambaran yang komprehensif tentang respons dan sentimen masyarakat terhadap program tersebut dalam rentang waktu yang signifikan sebelum dan sesudah peristiwa politik penting tersebut. Dengan demikian, dataset yang dihasilkan memberikan dasar yang kuat untuk analisis sentimen yang mendalam dan akurat terhadap program makan siang dan susu gratis Prabowo-Gibran di platform twitter.

3.2. Pelabelan Data

Proses pelabelan data dari total 1962 tweet yang telah terkumpul, sebanyak 1177 tweet dipilih untuk dilabeli secara manual berdasarkan opini yang terdapat di Twitter terkait program makan siang dan susu gratis Prabowo-Gibran. Proses pelabelan dilakukan dengan mengidentifikasi apakah setiap tweet mengekspresikan sentimen positif, netral, atau negatif terhadap program tersebut. Hasil pelabelan tersebut kemudian dijadikan sebagai data latih yang digunakan dalam proses pengembangan model analisis sentimen. Informasi mengenai sentimen dari masing-masing tweet ini dicatat dan disusun dalam (Tabel 1) untuk memberikan gambaran yang jelas dan terstruktur tentang persepsi masyarakat terhadap program tersebut.

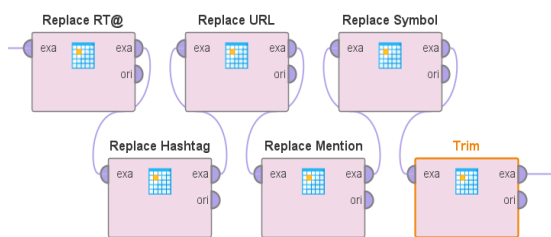
Tabel 1. Contoh Proses Pelabelan

Twitt	Sentimen
Preeetttt bangeeet ini Anak Raja" dari Jokowi... Cawapres hasil dari pelanggaran etika MKMK dan DKPP KPU sok-sok an menawarkan program, padahal program unggulan susu & makan siang gratis hanya menguntungkan pengusaha impor & pendukung	Negatif

internal sendiri... #AsalBukan02 @kesitt https://t.co/ggis9presW"	
Dampak program makan siang dan minum susu gratis. 1) menambah utang 2) meningkatkan impor pangan 3) menaikkan harga pangan 4) mengurangi biaya subsidi pupuk, BBM, gas dll. 5) meningkatkan kekayaan konco-konco sbg pelaksana proyek.	Negatif
@_riverheaven Makan siang dan susu gratis itu sangat penting dan bagus khusus sekolah yg di daerah ekonomi yg kurang tp implementasinya sulit dan mudah menjadi proyek korupsi	Netral
@sosmedkeras anggaran makan siang dan susu gratis mungkin bagus, tapi selama mental pejabat2 masih korup, ya gitu deh. Anggap aja anggaran per anak sekali makan siang 15 ribu, siapa yg jamin kl sampai ke anak 15rbu, bisa jadi malah cuma 3 ribu doang.	Netral
@Nab1002 @Robbie_prasetya @ainunnajib Sebenarnya itu program berkesinambungan, utk pencegahannya memang banyak hal, program makan siang dan susu gratis itu memang di prioritaskan utk mereka yg berekonomi kekurangan utk pencegahan stunting dan gizi buruk. Pemberdayaan SDM itu sangat penting utk pembangunan 2045.	Positif
@JatmikoRifiawan @prabowo Ini namanya action nyata dari Prabowo-Gibran, makan siang dan susu gratis buat sekolah itu meningkatkan kualitas anak-anak kita biar hasil belajarnya maksimal #WaktunyaPrabowoGibran #BersamaIndonesiaMaju	Positif

3.3. Text Preprocessing

Data cleansing adalah proses penting dalam text preprocessing untuk analisis lebih lanjut. Dalam proses ini, data yang terkumpul akan dibersihkan dari berbagai kesalahan atau anomali yang mungkin ada. Dalam penelitian ini, data dibersihkan menggunakan suboperator di dalamnya, di mana terdapat operator *replace* dan *trim*.



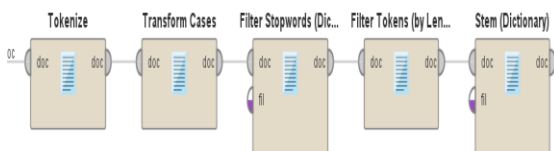
Gambar 4. Proses Data Cleansing

Pada (Gambar 4), dapat terlihat langkah-langkah yang dilakukan dalam proses cleansing data:

- Replace RT@*, yaitu langkah untuk menghapus kalimat yang dimulai dengan "@RT" dengan menggunakan kode "RT @.*". Hal ini dilakukan untuk menghilangkan kalimat yang diawali dengan "RT" dan diikuti oleh username tertentu.

- Replace Hashtag*, yaitu langkah untuk menghapus kalimat yang dimulai dengan tanda pagar "#" dengan menggunakan kode "#.*?". Tujuannya adalah untuk menghilangkan seluruh teks setelah tanda pagar.
- Replace URL*, yaitu langkah untuk menghapus kalimat yang dimulai dengan "https" dengan menggunakan kode "https.*?". Ini dilakukan untuk menghilangkan tautan URL dari teks.
- Replace Mention*, yaitu langkah untuk menghapus kalimat yang dimulai dengan "@" dengan menggunakan kode "@.*?". Tujuannya adalah untuk menghilangkan username dari teks.
- Replace Symbol*, yaitu langkah untuk menghapus simbol-simbol khusus menggunakan kode "[!\"#\$%&'()*+,-.\\:;<=>?@[\\|_`{}~]". Hal ini dilakukan untuk menghilangkan simbol-simbol dari teks yang tidak relevan.
- Operator Trim*, yaitu langkah untuk menghilangkan spasi tambahan di awal dan akhir teks, memastikan bahwa teks bebas dari karakter kosong yang tidak diinginkan.

Setelah melalui tahap cleansing data, langkah selanjutnya dalam proses persiapan data adalah melakukan serangkaian langkah untuk mempersiapkan teks untuk analisis lebih lanjut. Tahap ini termasuk dalam proses *text preprocessing* data teks dan melibatkan beberapa langkah penting (Gambar 5), yaitu *tokenize*, *transform cases*, *filter stopwords*, *filter token*, dan *stemming*.



Gambar 5. Proses Text Preprocessing Data

- Proses *Tokenize* membagi teks menjadi token atau kata-kata yang lebih kecil sehingga setiap kata dapat diolah secara individual. Dalam contoh yang tercantum pada (Tabel 2), setiap kalimat telah dipecah menjadi token-token yang terdiri dari kata-kata individu.

Tabel 2. Contoh Proses *Tokenize*

Sebelum <i>Tokenize</i>	Setelah <i>Tokenize</i>
Preeetttt bangeeet ini	"Preeetttt" "bangeeet" "ini"
Anak Raja dari Jokowi	"Anak" "Raja" "dari"
Cawapres hasil dari	"Jokowi" "Cawapres" "hasil"
pelanggaran etika MKMK	"dari" "pelanggaran" "etika"

dan DKPP KPU sokso	"MKMK"	"dan"	"DKPP"
an menawarkan program,	"KPU"	"sokso"	"an"
padahal program unggulan	"menawarkan"	"program"	
susu makan siang gratis	"padahal"	"program"	
hanya menguntungkan	"unggulan"	"susu"	"makan"
pengusaha impor	"siang"	"gratis"	"hanya"
pendukung internal sendiri	"menguntungkan"		
	"pengusaha"	"impor"	
	"pendukung"	"internal"	
	"sendiri"		

b. *Transform Cases* proses mengubah bentuk huruf dalam teks menjadi bentuk tertentu, seperti huruf kecil atau huruf besar, untuk memastikan konsistensi dalam analisis. Seperti yang ditunjukkan pada (Tabel 3), semua kata-kata telah diubah menjadi huruf kecil untuk konsistensi.

Tabel 3. Contoh Proses *Transform Cases*

Sebelum <i>Transform Cases</i>	Sesudah <i>Transform Cases</i>
Preeetttt bangeeet ini Anak Raja dari Jokowi Cawapres hasil dari pelanggaran etika MKMK dan DKPP KPU sokso an menawarkan program, padahal program unggulan susu makan siang gratis hanya menguntungkan pengusaha impor pendukung internal sendiri	preeetttt bangeeet ini anak raja dari jokowi cawapres hasil dari pelanggaran etika mkmk dan dkpp kpu sokso an menawarkan program padahal program unggulan susu makan siang gratis hanya menguntungkan pengusaha impor pendukung internal sendiri

c. *Filter Stopword* adalah proses eliminasi kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis, seperti "dan", "atau", atau "dimana". Dalam contoh pada (Tabel 4), kata-kata tersebut telah dihapus dari teks untuk meningkatkan relevansi dan kebermaknaan.

Tabel 4. Contoh Proses *Filter Stopword*

Sebelum <i>Filter Stopword</i>	Sesudah <i>Filter Stopword</i>
preeetttt bangeeet ini anak raja dari jokowi cawapres hasil dari pelanggaran etika mkmk dan dkpp kpu sokso an menawarkan program padahal program unggulan susu makan siang gratis hanya menguntungkan pengusaha impor pendukung internal sendiri	preeetttt bangeeet anak raja jokowi cawapres pelanggaran etika mkmk dkpp kpu sokso menawarkan program program unggulan susu makan siang gratis menguntungkan pengusaha impor pendukung internal

d. *Filter Token* adalah proses Proses ini bertujuan untuk menyaring kata-kata berdasarkan panjangnya, di mana kata-kata yang terlalu pendek atau terlalu panjang akan dihapus. Contohnya, dalam (Tabel 5), kata-kata yang memiliki panjang antara 4 hingga 25 karakter dipertahankan untuk analisis lebih lanjut.

Tabel 5. Contoh Proses *Filter Token*

Sebelum <i>Filter Token</i>	Sesudah <i>Filter Token</i>
preeetttt bangeeet anak raja jokowi cawapres pelanggaran etika mkmk dkpp kpu sokso menawarkan program program unggulan susu makan siang gratis menguntungkan pengusaha impor pendukung internal	preeetttt bangeeet anak raja jokowi cawapres pelanggaran etika mkmk dkpp sokso menawarkan program program unggulan susu makan siang gratis menguntungkan pengusaha impor pendukung internal

e. *Stemming* adalah proses mengonversi kata-kata ke bentuk dasarnya, sehingga kata-kata dengan variasi bentuk dapat diidentifikasi sebagai kata yang sama. Dalam (Tabel 6).

Tabel 6. Contoh Proses *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
preeetttt bangeeet anak raja jokowi cawapres pelanggaran etika mkmk dkpp kpu sokso menawarkan program program unggulan susu makan siang gratis menguntungkan pengusaha impor pendukung internal	preeetttt bangeeet anak raja jokowi cawapres pelanggaran etika mkmk dkpp sokso menawarkan program program unggulan susu makan siang gratis menguntungkan pengusaha impor pendukung internal

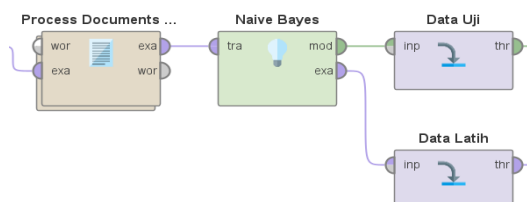
3.4. Pembobotan Kata

Setelah data telah melalui serangkaian proses *text preprocessing*, langkah selanjutnya adalah pembobotan kata menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) pada operator *Process Documents from Data*. Metode ini dipilih karena sebagai metode yang mudah, efisien, dan memiliki akurasi tinggi dalam mengukur pentingnya sebuah kata dalam sebuah dokumen. *TF-IDF* memperhitungkan frekuensi kemunculan kata dalam sebuah dokumen (*TF*) serta seberapa umum kata tersebut di seluruh koleksi dokumen (*IDF*) [23]. Dengan *TF-IDF*, kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul dalam dokumen lain akan diberi bobot yang lebih tinggi, sehingga berpengaruh lebih besar dalam hasil analisis sentimen. Hasil dari pembobotan kata ini akan digunakan untuk melatih model analisis sentimen yang memungkinkan klasifikasi data ke dalam kategori negatif, netral, dan positif berdasarkan kemunculan kata-kata yang relevan dalam teks.

3.5. Implementasi Algoritma Naïve Bayes

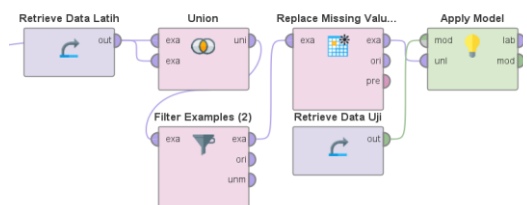
Implementasi algoritma Naïve Bayes sebagai metode analisis sentimen dilakukan dengan menghubungkan operator *Process Document from*

Data ke operator Naïve Bayes. Setelahnya, operator Naïve Bayes disambungkan kembali ke dua operator Store. Fungsi dari operator Store pertama adalah menyimpan hasil dari proses yang dilakukan oleh operator Naïve Bayes, yang kemudian dinamai ulang menjadi Data Latih. Sedangkan, operator Store lainnya dinamai Data Uji dan bertujuan untuk menyimpan hasil dari proses pada data Uji (Gambar 6).



Gambar 6. Proses Implementasi Naive Bayes

Setelah itu, data uji digabungkan dengan data latih yang sudah disiapkan menggunakan operator *Union* (Gambar 7). Kemudian, data tersebut diproses lebih lanjut menggunakan operator *Filter Example* untuk mengatasi teks dengan sentimen kosong. Untuk menangani nilai yang hilang, digunakan operator *Replace Missing Value* dengan pengaturan default "zero", memungkinkan sistem untuk mengelola data yang tidak terisi. Terakhir, operator *Apply Model* digunakan untuk menyimpan dan menampilkan hasil analisis sentimen yang dihasilkan oleh model Naïve Bayes.



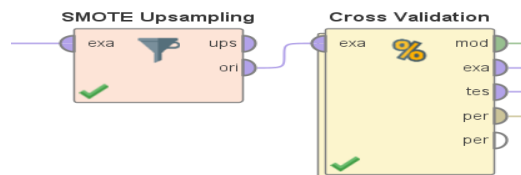
Gambar 7. Implementasi Data Uji

Setelah proses pengujian dilakukan, ini memberikan gambaran yang lebih jelas tentang persebaran sentimen dalam data uji, yang penting untuk dievaluasi lebih lanjut dalam analisis sentimen.

3.6. Evaluasi

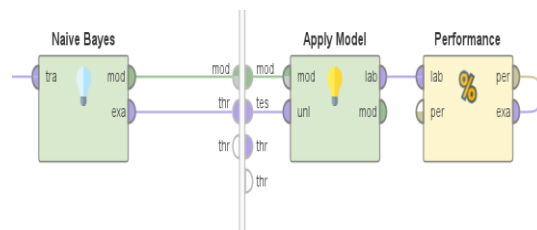
langkah selanjutnya adalah untuk mengevaluasi yaitu menggunakan operator *SMOTE Up Sampling*. Langkah ini bertujuan untuk menangani ketidakseimbangan kelas dalam dataset dengan menghasilkan sampel sintetis dari kelas minoritas. Setelah itu, operator ini dihubungkan dengan

operator *cross-validation* dengan jumlah *fold* sebanyak 10, seperti yang ditunjukkan dalam (Gambar 8). *Cross-validation* digunakan untuk membagi dataset menjadi beberapa subset untuk mengukur performa mode secara lebih akurat.



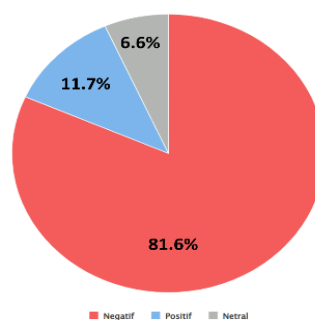
Gambar 8. Menghubungkan Operator *SMOTE Upsampling* dan *Cross Validation*

Dalam operator *cross-validation*, terdapat operator lain yang penting untuk proses evaluasi model (Gambar 9), yaitu Naïve Bayes. Operator Naïve Bayes dihubungkan dengan operator *Apply Model*, yang bertujuan untuk menerapkan model. Setelah itu, hasil dari operator *Apply Model* dihubungkan dengan operator *Performance*. Operator *Performance* digunakan untuk mengukur kinerja model dalam melakukan klasifikasi, dengan menyajikan metrik-metrik seperti *accuracy*, *precision*, dan *recall*.



Gambar 9. Proses Evaluasi Operator *Performance* guna mengukur kinerja model

Setelah proses evaluasi dilakukan, dari total data uji sebanyak 785 ditemukan bahwa 641 (81,7%) di antaranya memiliki sentimen negatif, 92 (11,7%) memiliki sentimen positif, dan 52 (6,6%) data memiliki sentimen netral, yang divisualisasikan pada (Gambar 10) dalam bentuk *Pie Chart*.



Gambar 10. Visualisasi *Pie Chart* Sebaran Sentimen

Hasil *confusion matrix* dari data uji, seperti yang ditampilkan dalam (Gambar 11), memberikan gambaran yang jelas tentang kinerja model Naïve Bayes dalam mengklasifikasikan sentimen.

accuracy: 90.96% +/- 2.51% (micro average: 90.96%)

	true Positif	true Negatif	true Netral	class precision
pred. Positif	87	5	0	94.57%
pred. Negatif	15	614	12	95.79%
pred. Netral	33	6	13	25.00%
class recall	64.44%	98.24%	52.00%	

Gambar 11. Hasil *Confusion Matrix*

Dengan perhitungan rumus sebagai berikut:

$$Accuracy = \frac{\text{(Data Benar)}}{\text{(Jumlah Data)}} = \frac{714}{785} = 0,909554 \text{ atau } 92,96\%$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{87}{92} = 0,945652 \text{ atau } 94,57\%$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{87}{102} = 0,852941 \text{ atau } 85,30\%$$

SIMPULAN

Berdasarkan hasil pengujian data analisis sentimen di Twitter menggunakan metode Naïve Bayes terhadap program makan siang dan susu gratis Prabowo-Gibran dari total 785 jumlah data uji dan telah dievaluasi menunjukkan adanya dominasi sentimen negatif dengan jumlah 641 data twitter atau sekitar 81,7%. Sentimen negatif ini muncul atas berbagai faktor, contohnya seperti adanya sikap keraguan masyarakat terhadap efektivitas program, adanya kekhawatiran tentang transparansi dan akuntabilitas penyaluran dana, ataupun adanya ketidaksetujuan terhadap target atau mekanisme program. Disisi lain, ditemukan juga 52 jumlah data pada twitter yang bersifat netral, atau sekitar 6,6%. Hal ini menunjukkan ketidakjelasan atau netralitas dalam tanggapan masyarakat. Selanjutnya, data positif sebanyak 92 jumlah data pada twitter, atau sekitar 11,7%, menunjukkan bahwa masih ada sikap optimisme atau dukungan terhadap program tersebut. Selain itu, *accuracy* sebesar 92,96%, *recall* sebesar 85,30%, dan *precision* sebesar 94,57% menunjukkan bahwa model analisis sentimen memiliki performa yang baik dalam mengklasifikasikan sentimen dari data uji. Meskipun mayoritas tanggapan masyarakat terhadap program makan siang dan susu gratis Prabowo-Gibran cenderung negatif, namun temuan analisis sentimen ini memberikan pemahaman

berharga terkait pandangan dan respons yang beragam di *platform* Twitter.

DAFTAR PUSTAKA

- [1] D. Wiryany, S. Natasha, and R. Kurniawan, "PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI TERHADAP PERUBAHAN SISTEM KOMUNIKASI INDONESIA," *Jurnal Nomosleca*, vol. 8, no. 2, pp. 242–252, 2022.
- [2] "Visi, Misi dan Program Calon Presiden dan Wakil Presiden 2024-2029 H. Prabowo Subianto, Gibran Rakabuming Raka," *medcom.id*. Accessed: May 18, 2024. [Online]. Available: https://va.medcom.id/2023/pemilu/others/PRABOWOGIBRA_N_VISI_MISI.pdf
- [3] V. E. Togatorop, L. Rahayuwati, and R. D. Susanti, "Predictor of Stunting Among Children 0-24 Months Old in Indonesia: A Scoping Review," *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, vol. 7, no. 5, pp. 5654–5674, 2023, doi: 10.31004/obsesi.v7i5.5222.
- [4] N. Aurora Prameswari *et al.*, "Public Search Interest in Stunting Interventions as an Effort to Reduce the Incidence of Stunting in Indonesia during 2018-2022: A Google Trends Analysis," *Amerta Nutrition*, vol. 7, no. 3SP, pp. 41–49, 2023, doi: 10.20473/amnt.v7i3SP.2023.41-49.
- [5] R. N. Chatarine, "TKN: Dana Program Makan Siang Gratis Rp 450 Triliun, Tak Gunakan Dana Bansos," *nasional.kompas.com*. Accessed: May 18, 2024. [Online]. Available: <https://nasional.kompas.com/read/2023/12/20/19132411/tkn-dana-program-makan-siang-gratis-rp-450-triliun-tak-gunakan-dana-bansos>
- [6] R. Mulyana, H. Pramadya, and N. R. Vindiazhari, "Pelatihan Literasi Digital Lembaga Untuk Lanjut Usia Indonesia (LLI) Kota Bandung Digital Literacy Training For Lembaga Lanjut Usia (LLI) Bandung City," *Solusi Bersama : Jurnal Pengabdian dan Kesejahteraan Masyarakat*, vol. 1, no. 2, pp. 01–07, 2024.
- [7] A. Kartino, M. K. Anam, Rahmadden, and Junadhi, "Analisis Akun Twitter Berpengaruh terkait Covid-19 menggunakan Social Network Analysis," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 697–704, Aug. 2021, doi: 10.29207/resti.v5i4.3160.
- [8] D. Setiyawati and N. Cahyono, "Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia," *Indonesian Journal of Computer Science*, vol. 12, no. 1, pp. 262–272, 2023.
- [9] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 504–510, Jun. 2021, doi: 10.29207/resti.v5i3.3067.
- [10] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 8–14, 2023.
- [11] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," *Jurnal TEKNO KOMPAK*, vol. 17, no. 1, pp. 25–38, 2023.
- [12] J. Florensius Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: KAJIAN ILMIAH INFORMATIKA DAN KOMPUTER*, vol. 4, no. 1, pp. 360–367, 2023, doi: 10.30865/klik.v4i1.1033.
- [13] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan*

- Komputer*, vol. 3, no. 6, pp. 1340–1348, 2023, doi: 10.30865/klik.v3i6.936.
- [14] P. Paramita and A. Ibrahim, “ANALISIS SENTIMEN TERHADAP PENGGUNA QRIS (QUICK RESPOND CODE INDONESIAN STANDART) PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” *JOISIE Journal Of Information System And Informatics Engineering*, vol. 7, no. 1, pp. 1–6, 2023, [Online]. Available: <https://t.co/lJemg7TbKb>
- [15] Sulindawaty, E. Laia, and M. Yamin, “Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 305–316, 2023, doi: 10.30865/klik.v4i1.1186.
- [16] R. A. Firsttama, A. A. Arifiyanti, and D. S. Y. Kartika, “Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes,” *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 2, pp. 282–285, Apr. 2024, doi: 10.47233/jteksis.v6i2.1263.
- [17] W. Gunawan, R. A. Wiradiputra, A. P. Sari, D. Prayama, and E. R. Nainggolan, “Prediction of Cross-Platform and Native Apps Technology Opportunities for Beginner Developers Using C 4.5 and Naive Bayes Algorithms,” *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2145–2153, Dec. 2023, doi: <http://dx.doi.org/10.30630/joiv.7.4.01514>.
- [18] N. T. S. Saptadi, P. Chyan, and J. M. Leda, “Analysis of Supermarket Product Purchase Transactions With the Association Data Mining Method,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 618–627, Jun. 2023, doi: 10.29207/resti.v7i3.4844.
- [19] A. Algiffary and T. Sutabri, “Implementasi Machine Learning dengan Algoritma Naive Bayes Terhadap Sistem Informasi Pelayanan Pemberkasan Kepegawaian di BKPSDM Kota Palembang,” *Indonesian Journal of Computer Science*, vol. 12, no. 3, pp. 1272–1281, 2023.
- [20] B. Bayu Baskoro, I. Susanto, and S. Khomsah, “Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR),” *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 3, no. 2, pp. 21–029, 2021, doi: 10.20895/INISTA.V3I2.
- [21] M. Hayaty, S. Muthmainah, and S. M. Ghufran, “Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification,” *International Journal of Artificial Intelligence Research*, vol. 4, no. 2, pp. 86–94, Dec. 2020, doi: 10.29099/ijair.v4i2.152.
- [22] R. Aryanti, T. Misriati, and R. Hidayat, “Klasifikasi Risiko Kesehatan Ibu Hamil Menggunakan Random Oversampling Untuk Mengatasi Ketidakseimbangan Data,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 5, pp. 409–416, Apr. 2023, [Online]. Available: <https://djournal.com/klik>
- [23] C. H. Yutika, Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, pp. 422–430, Apr. 2021, doi: 10.30865/mib.v5i2.2845.