

Penerapan Algoritma Naïve Bayes Pada Analisa Sentimen Twitter Terhadap Opini Publik Badan Pangan Nasional

Andhika Prima*, Afiyati

Teknik Informatika, Universitas Mercu Buana, Jakarta
*41520010078@mercubuana.ac.id

Abstrak— Penelitian ini fokus pada penerapan algoritma Naïve Bayes untuk menganalisis sentimen Twitter terhadap Badan Pangan Nasional (BAPANAS). Dalam era digital, media sosial, khususnya Twitter, menjadi saluran utama masyarakat untuk menyampaikan opini terkait instansi pemerintah. Algoritma Naïve Bayes digunakan untuk mengklasifikasikan sentimen menjadi positif atau negatif. Dengan langkah-langkah yang melibatkan crawling data, preprocessing, pelabelan data otomatis menggunakan InSet Lexicon, pembobotan kata dengan TF-IDF, data splitting, dan klasifikasi dengan algoritma Naïve Bayes. Hasil klasifikasi menunjukkan akurasi sebesar 79.7%, dengan presisi 78.6%, recall 74.0%, dan F1 score 76.2%. Algoritma Naïve Bayes mengklasifikasikan sebanyak 1.093 data. Dari hasil tersebut, 453 sentimen positif (41.4%) sementara 640 sentimen negatif (58.6%) berdasarkan data testing sebanyak 20%.

Article History:

Received: June 16, 2023

Revised: July 24, 2024

Accepted: Sept 15, 2024

Published: Jan 31, 2025

Kata Kunci— Analisa Sentimen; Naïve Bayes; Badan Pangan Nasional; Klasifikasi; Twitter

DOI: 10.22441/jitkom.v9i1.007

I. PENDAHULUAN

Badan Pangan Nasional (BAPANAS) sebagai lembaga pemerintah mempunyai tanggung jawab yang besar dalam menjaga ketahanan pangan nasional. Keberhasilan BAPANAS dalam menjalankan fungsinya sangat penting dalam menjamin kecukupan pangan bagi masyarakat.

Di era digital dan kemajuan teknologi, Internet menjadi bahan referensi utama yang menunjang berbagai aktivitas. Misalnya, penggunaan media sosial, yang merupakan bagian integral dari Internet, telah menjadi hal yang lumrah dalam kehidupan masyarakat. Dalam konteks ini, media sosial digunakan masyarakat untuk mengungkapkan opini pribadi, perasaan, dan hal-hal terkait lainnya. Penggunaan media sosial oleh masyarakat tidak hanya memfasilitasi ekspresi diri, namun juga membuat opini yang disampaikan kepada lembaga pemerintah lebih mudah diakses dan tersebar luas melalui berbagai jejaring sosial [1], [2].

Jejaring sosial tidak hanya berfungsi sebagai media penyampaian pendapat, namun juga sebagai sumber informasi yang mempengaruhi persepsi masyarakat. Opini publik yang dibentuk melalui platform Twitter dapat memberikan dampak yang signifikan terhadap citra dan kredibilitas lembaga seperti BAPANAS [3], [4], [5].

Twitter merupakan salah satu jejaring sosial yang menjadi saluran utama masyarakat dalam menyampaikan pandangan, kritik, dan pendapatnya terhadap kegiatan instansi pemerintah. Dalam konteks ini, penting untuk memahami opini dan persepsi masyarakat terhadap BAPANAS untuk meningkatkan efektivitas kebijakan dan komunikasi lembaga. Oleh karena itu, menganalisis sentimen publik di jejaring sosial dapat

memberikan informasi berharga untuk kebijakan publik, manajemen reputasi, dan pengambilan keputusan [6], [7], [8].

Dalam konteks ini, algoritma Naïve Bayes menonjol sebagai metode analisis sentimen yang dapat diterapkan pada data teks seperti yang ditemukan di Twitter. Algoritma Naïve Bayes adalah algoritma klasifikasi yang banyak digunakan dalam Data Mining atau Text Mining [9], [10], [11], [12], [13]. Penelitian ini diawali dengan keinginan untuk memberikan kontribusi yang diharapkan dapat memberikan wawasan yang bermanfaat bagi BAPANAS dalam berinteraksi dengan masyarakat.

Sebelum menerapkan model atau algoritma, dilakukan proses pelabelan data dan pembobotan kata. Untuk meningkatkan efisiensi waktu dan tenaga, pendekatan pelabelan otomatis menggunakan InSet Lexicon digunakan sebagai metode leksikon bahasa Indonesia [14], [15], serta TF-IDF sebagai pembobotan kata. Pada penelitian ini, algoritma Naïve Bayes digunakan untuk menganalisis tren opini masyarakat terhadap BAPANAS di Indonesia dari sudut pandang opini positif dan negatif [16]–[22].

II. LITERATURE REVIEW

Dalam sebuah penelitian yang dilakukan oleh [7], analisis sentimen diterapkan pada tweet terkait Dewan Perwakilan Rakyat (DPR). Hasil analisis menunjukkan bahwa dari 1.546 tweet, 95 diklasifikasikan sebagai positif, 693 sebagai netral, dan 758 sebagai negatif. Algoritma Naïve Bayes mencapai skor akurasi sebesar 80%, menunjukkan kemampuannya untuk memprediksi sentimen dengan tingkat akurasi yang baik, khususnya dalam membedakan antara sentimen positif, netral, dan negatif.

Penelitian lain, seperti yang disajikan oleh [11], menggunakan Naive Bayes Classifier untuk klasifikasi sentiment terhadap Kementerian Perdagangan Republik Indonesia dilakukan dengan menganalisis 1000 tweet di jejaring sosial Twitter. Hasilnya menunjukkan tingkat analisis sentimen sebesar 84,02%, yang mayoritas merupakan opini positif, sementara 15,98% mengungkapkan pendapat negatif. Evaluasi dilakukan melalui perhitungan nilai akurasi dan konfusi matriks, menghasilkan tingkat akurasi sebesar 89,24%.

Selain itu, penelitian oleh [8] mengeksplorasi analisis sentimen terkait kekurangan minyak goreng melalui penggunaan 600 tweet. Hasil klasifikasi menunjukkan dominasi sentimen negatif dengan 363 tweet, sementara sentimen positif tercatat sebanyak 237 tweet. Dampak dari temuan ini adalah perubahan sentimen masyarakat menjadi negatif terkait masalah kekurangan minyak goreng. Pengujian presisi pada penelitian ini menghasilkan nilai akurasi sebesar 85,8%, presisi sebesar 64,7%, dan recall sebesar 94,2%, memberikan pemahaman yang mendalam terkait efektivitas Naive Bayes dalam menghadapi tantangan analisis sentimen pada konteks tertentu.

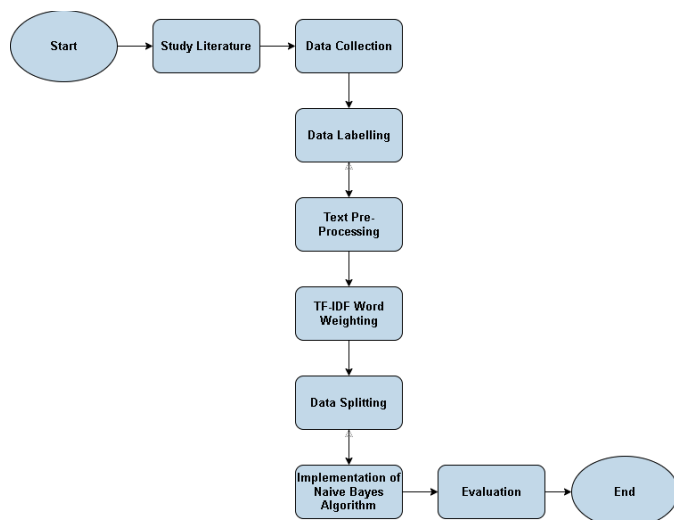
III. METODOLOGI PENELITIAN

A. Penelitian Kualitatif

Penelitian ini merupakan penelitian kuantitatif yang menerapkan algoritma Naive Bayes untuk menganalisis sentimen opini publik terhadap Badan Pangan Nasional (BAPANAS) di platform Twitter. Dengan menggunakan algoritma klasifikasi ini, penelitian ini bertujuan untuk mengukur sentimen tweet dan mengklasifikasikannya ke dalam kategori positif dan negatif.

B. Tahapan Penelitian

Berikut ini adalah alur tahapan penelitian yang dituangkan dalam bentuk diagram alir. Diagram alir digunakan untuk menggambarkan urutan langkah-langkah yang akan dilakukan dalam penelitian ini.



Gambar 1. Tahapan Penelitian

Rincian Alur Tahap Penelitian pada gambar 1:

1) Study Literature

Studi Literatur merupakan langkah awal yang penting dalam penelitian ini untuk memahami kerangka dan konteks analisis sentimen, khususnya pada platform Twitter. Sumber yang relevan antara lain teori analisis sentimen, teknik pengumpulan data dari twitter, dan penerapan algoritma Naïve Bayes pada analisis sentimen. Penelitian sebelumnya yang menerapkan algoritma Naïve Bayes untuk analisis sentimen media sosial menjadi fokus tinjauan literatur ini. Keberhasilan dan kelemahan studi-studi ini diidentifikasi dan landasan konseptual yang kuat disediakan

2) Data Collection

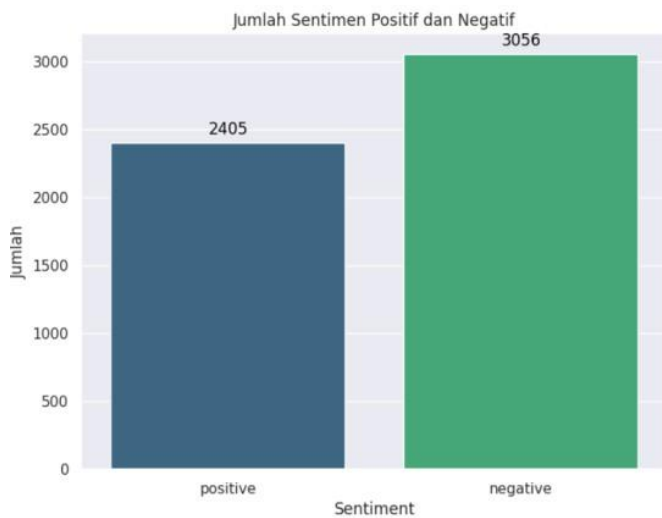
Pada tahap ini, dikumpulkan melalui proses crawling menggunakan library Python yang bernama "twitter-harvest". Pengumpulan data dilakukan menggunakan parameter pencarian berupa kata kunci "badan pangan nasional" " dalam rentang waktu 03 November 2021 hingga 23 November 2023 dan total 5.952 data. Setelah itu penghapusan duplikasi data dan menghapus atribut yang tidak perlu, jadi penulis hanya memfokuskan atribut "full text" yang mana berisikan hasil sentiment dari twitter. Jadi Total data yang sudah crawling dan menghapus duplikasi data adalah 5.747 tweets.

3) Data Labelling

Pengumpulan data dilakukan Pada tahap pelabelan data, metode pelabelan manual dirasa tidak efisien dari segi tenaga dan waktu, karena jumlah data yang diperoleh melalui crawling cukup banyak. Oleh karena itu, penulis memutuskan untuk menggunakan kamus InSet Lexicon untuk pelabelan otomatis. Kamus ini terdiri dari 3.609 kata positif dan 6.609 kata negatif yang masing- masing memiliki nilai bobot antara -5 dan +5.

Pendekatan pelabelan ini menghitung bobot setiap kata dalam kamus untuk menentukan apakah kalimat tersebut mempunyai sentimen positif atau negatif. Jumlahkan bobot kata dan dibagi dengan jumlah total kata dalam kalimat untuk mendapatkan nilai sentimen rata-rata kalimat tersebut. dari hasil rata-rata nya di ketahui kalimat itu negatif atau positif.

Hasil pelabelan data menunjukkan dari total 5.748 tweets, terdapat 2.405 tweets dengan sentimen positif, 3.056 tweets dengan sentimen negatif, dan 286 tweets dengan sentimen netral. Dengan memfokuskan pada sentimen positif dan negatif, maka total dataset nya adalah 5.462 tweets dengan hasil normalisasi persentase menunjukkan bahwa sentimen positif mencakup 44.04% % dan sentimen negatif mencakup 55.96%. Kemudian berikut adalah gambar 4 untuk menunjukkan hasil visualisasi pelabelan data.



Gambar 2. Visualisasi Data Labelling

4) Text Pre-Processing

Tahap ini merupakan proses pengolahan data yang mengubah data yang diperoleh dari penambangan menjadi data yang lebih bersih.

Text pre-processing adalah tahapan proses awal dalam pemrosesan teks yang mengubah teks menjadi data untuk diproses lebih lanjut. Sekumpulan karakter (teks) yang berurutan harus dipecah menjadi komponen yang lebih bermakna, dan ini dapat dilakukan pada berbagai tingkatan [18]. Data yang telah melalui tahap preprocessing siap untuk diolah pada tahap selanjutnya. Pada tahap ini, beberapa teknik yang digunakan untuk processing seperti *case folding*, *tokenizing*, *cleansing*, dan *stopword*.

5) TF-IDF Word Weighting

Pembobotan fitur dilakukan untuk mengubah fitur kata menjadi data numerik. Pada tahap ini, setiap fitur yang diperoleh dari langkah preprocessing teks diberi bobot dengan istilah *Term Frequency Inverse Document Frequency (TF-IDF)* yang menghasilkan bobot berdasarkan frekuensi kemunculan kata tertentu dalam dokumen [2] TF-IDF memberikan bobot kata yang lebih tinggi untuk kata-kata yang sering muncul dalam suatu dokumen, namun jarang di seluruh kumpulan dokumen. Ini membantu Anda mengenali kata-kata yang memberikan informasi lebih spesifik tentang dokumen.

6) Data Splitting

Dalam proses data *splitting*, digunakan metode pembagian data dengan perbandingan 80% untuk data *training* dan 20% untuk data *testing*. Atribut yang digunakan atribut "*full_text*" (berisi teks lengkap dari tweet) dan atribut hasil pelabelan dilakukan dengan menggunakan fungsi "train test split" dari modul "skeleton.model selection" dalam *library* sklearn.

7) Implementation of Naïve Bayes Algorithm

Algoritma Naïve Bayes merupakan algoritma klasifikasi yang biasa digunakan pada *data mining* atau *text mining*. Naive Bayes adalah metode klasifikasi *text mining* yang digunakan dalam analisis sentimen. Metode ini mungkin cocok untuk klasifikasi dalam presisi dan komputasi data [21]. Asumsi

mendasar dari pendekatan *Naïve Bayes* adalah bahwa semua atribut bersifat independen. Hal ini membuat klasifikasi menjadi sangat sederhana dan efektif. Di bawah ini persamaan teorema Bayes [14] :

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)} \tag{1}$$

yang mana :

X = Data dengan memiliki kelas yang belum diketahui

H = Hipotesis data suatu kelas tertentu

P(H|X) = Peluang Hipotesis H berdasarkan kondisi X

P(H) = Kemungkinan peluang Hipotesis (prior probability)

P(X|H) = Peluang X berdasarkan kondisi pada dugaan A

P(X) = Kemungkinan peluang X

Naive Bayes banyak digunakan dalam teknik klasifikasi khususnya Twitter dengan menggunakan berbagai metode seperti *Unigram Naive Bayes*, *Multinomial Naive Bayes*, dan *Maximum Entropy Classification* [21].

8) Evaluation

Pada tahap ini, kinerja algoritma Naive Bayes dievaluasi menggunakan confusion matrix sebagai tool utama, Confusion matrix adalah tool untuk mengevaluasi model klasifikasi untuk memperkirakan objek mana yang benar atau salah. Sebuah Matriks untuk membandingkan hasil klasifikasi dengan data asli [10]. Tabel confusion matrix dapat dilihat pada tabel 1.

Tabel 1. Table Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

Pada tabel 1. TP adalah True Positive, TN adalah True Negative, FP adalah False Positive dan FN adalah False Negative. Pada penelitian ini performa klasifikasi yang akan dihitung adalah accuracy, precision, recall dan f1 score. Pada persamaan (1), (2), (3), dan (4) berturut-turut merupakan rumus perhitungan accuracy, precision dan recall [16, 1].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{5}$$

IV. HASIL DAN ANALISA

A. Dataset

Dataset penelitian terdiri dari 5.461 tweet yang telah melalui proses pelabelan, deduplikasi, dan penghapusan atribut kolom yang tidak perlu. Analisis ini berfokus pada sentimen positif dan negatif dengan membuang data yang memiliki sentimen netral. Pengumpulan data dilakukan menggunakan parameter pencarian berupa kata kunci “badan pangan nasional” " dalam rentang waktu 03 November 2021 hingga 23 November 2023.

	full_text	sentiment
0	Direktur Utama Perum BULOG Budi Waseso bersama...	positive
1	Harga Beras Naik, Badan Pangan Nasional Dimint...	positive
2	Cah enom ini sepertinya bakal calon menteri KK...	negative
3	Anggota Komisi IV DPR RI dari Fraksi Partai Na...	negative
4	PT Sang Hyang Seri anggota BUMN Pangan ID FOOD...	positive

Gambar 3. Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5461 entries, 0 to 5460
Data columns (total 2 columns):
# Column Non-Null Count Dtype
---  ---
0 full_text 5461 non-null object
1 sentiment 5461 non-null object
dtypes: object(2)
memory usage: 85.5+ KB
```

Gambar 4. Jumlah Dataset

B. Pelabelan Data

Proses pelabelan data pada penelitian ini dilakukan secara otomatis menggunakan InSet Lexicon (*Indonesian Sentiment Lexicon*) atau bisa disebut kamus sentimen bahasa Indonesia. Kamus ini terdiri dari 3.609 kata positif dan 6.609 kata negatif dengan nilai -5 hingga +5 [12]. Kamus ini menyediakan daftar kata-kata dengan sentimen yang telah dikategorikan dan di bobot kan sebelumnya, berikut adalah tabel 2 contoh daftar kata negatif dan positif beserta bobot nya dalam InSet Lexicon.

Tabel 2. Contoh Daftar Kata Pada InSet Lexicon

Kata	Bobot
Sakit	-5
Soalnya	-1
Purba	-3
Gaada	-3
Cenderung	-2
Hendaknya	1
Menarik	4
Semboyan	2
Silahkan	3
Terimakasih	5

Dan Berikut tabel 3 ini adalah hasil pelabelan otomatis dengan menggunakan Inset Lexicon.

Tabel 3. Hasil Pelabelan Otomatis

	full_text	sentiment
0	Direktur Utama Perum BULOG Budi Waseso bersama Kepala Badan Pangan Nasional Arief Prasetyo Adi melakukan pengecekan langsung dengan mendatangi LotteMart Wholesale Pasar Rebo pada Jum'at (08/09). Selengkapnya klik : https://t.co/0utDc046Kp https://t.co/A00pzo6M3n	positive
1	Harga Beras Naik, Badan Pangan Nasional Diminta Menggelar Operasi Pasar di Kota Bekasi - https://t.co/J7mZ0zCt7g https://t.co/1Vm0INSReO	positive
2	Cah enom ini sepertinya bakal calon menteri KKP atau kepala Badan Pangan Nasional. Siapa yg setuju?	negative
3	Anggota Komisi IV DPR RI dari Fraksi Partai NasDem, Sulaeman L Hamzah, meminta Badan Pangan Nasional (Bapanas) mengantisipasi dampak El Nino terhadap stabilitas pasokan dan ketersediaan pangan Tanah Air. https://t.co/01L1ru0K8D	negative
4	PT Sang Hyang Seri anggota BUMN Pangan ID FOOD, Sub Klaster Pertanian bersama Badan Pangan Nasional serta Badan Riset dan Inovasi Nasional (BRIN), Kementerian BUMN dan Pelaku Usaha sektor Pangan menjalankan kerja sama budidaya pertanian berbasis bibit unggul. #SHScadanganberas https://t.co/pzFHDh3SV	positive

C. Pre-Processing

Tujuan Pre-processing adalah untuk membersihkan, mengubah, atau menyesuaikan data agar sesuai dengan kebutuhan analisis dan pemodelan yang akan dilakukan. Pada tahap Pre-processing, langkah-langkah dilakukan meliputi :

1) Case Folding

Langkah ini bertujuan untuk konsistensi dan mempermudah proses analisis, menghindari perbedaan kapitalisasi yang tidak diperlukan. Berikut adalah hasil case folding pada tabel 4.

Tabel 4. Hasil Case Folding

	full_text	Case folding
0	Direktur Utama Perum BULOG Budi Waseso bersama...	direktur utama perum bulog budi waseso bersama...
1	Harga Beras Naik, Badan Pangan Nasional Diminta...	harga beras naik, badan pangan nasional diminta...
2	Cah enom ini sepertinya bakal calon menteri KK...	cah enom ini sepertinya bakal calon menteri kk...
3	Anggota Komisi IV DPR RI dari Fraksi Partai Na...	anggota komisi iv dpr ri dari fraksi partai na...
4	PT Sang Hyang Seri anggota BUMN Pangan ID FOOD...	pt sang hyang seri anggota bumn pangan id food...

2) Tokenizing

Langkah tokenisasi ini memecah teks menjadi token atau kata-kata individual, Berikut adalah hasil tokenisasi pada tabel 5.

Tabel 5. Hasil Tokenisasi

	Case folding	Tokenizing
0	direktur utama perum bulog budi waseso bersama...	[direktur, utama, perum, bulog, budi, waseso, ...
1	harga beras naik, badan pangan nasional diminta...	[harga, beras, naik, ,, badan, pangan, nasiona...
2	cah enom ini sepertinya bakal calon menteri kk...	[cah, enom, ini, sepertinya, bakal, calon, men...
3	anggota komisi iv dpr ri dari fraksi partai na...	[anggota, komisi, iv, dpr, ri, dari, fraksi, p...
4	pt sang hyang seri anggota bumn pangan id food...	[pt, sang, hyang, seri, anggota, bumn, pangan,...

3) *Stopword*

Langkah ini bertujuan untuk mempersiapkan data agar lebih fokus pada kata-kata yang lebih bermakna, meninggalkan stopwords yang mungkin mengganggu proses analisis sentiment. Berikut adalah hasil *Stopword* pada tabel 6.

Table 6. Hasil *Stopword*

	<i>Tokenizing</i>	<i>Stopword</i>
0	[direktur, utama, perum, bulog, budi, waseso, ...	['direktur', 'utama', 'perum', 'bulog',
1	[harga, beras, naik, , badan, pangan, nasiona...	['harga', 'beras', 'naik', ',', 'bad...
2	[cah, enom, ini, sepertinya, bakal, calon, men...	['cah', 'enom', 'ini', 'sepertinya', ...
3	[anggota, komisi, iv, dpr, ri, dari, fraksi, p...	['anggota', 'komisi', 'iv', 'dpr', 'r...
4	[pt, sang, hyang, seri, anggota, bumh, pangan...	['pt', 'sang', 'hyang', 'seri', 'angg...

4) *Cleansing*

Tahap selanjutnya dalam preprocessing data adalah cleansing, yang bertujuan untuk membersihkan teks dari karakter-karakter khusus, tanda baca, dan elemen-elemen lain yang mungkin tidak memberikan kontribusi pada analisis sentiment. Berikut adalah hasil *Stopword* pada tabel 7.

Tabel 7. Hasil *Cleansing*

	<i>Cleansing</i>
0	direktur utama perum bulog budi waseso bersama...
1	harga beras naik badan pangan nasional diminta...
2	cah enom ini sepertinya bakal calon menteri kk...
3	anggota komisi iv dpr ri dari fraksi partai na...
4	pt sang hyang seri anggota bumh pangan id food...

D. *Modelling dan Pengujian*

1) *Case Folding*

Pada tahap ini, proses proses pemisahan dataset menjadi dua bagian utama, yaitu data *training* dan data *testing*, 20% dari dataset akan diambil sebagai data testing, sementara 80% sisanya akan menjadi data training. Hasil Data *Splitting* 4.368 data *training* sementara 1.093 data *testing*.

2) *Pembobotan TF-IDF*

Dalam rangka meningkatkan representasi fitur teks pada tahap analisis sentimen, penelitian ini menerapkan metode pembobotan TF-IDF (Term Frequency-Inverse Document Frequency). Untuk mengimplementasikan pembobotan TF-IDF secara efisien, penelitian ini menggunakan "Pipeline" dari library "sklearn".

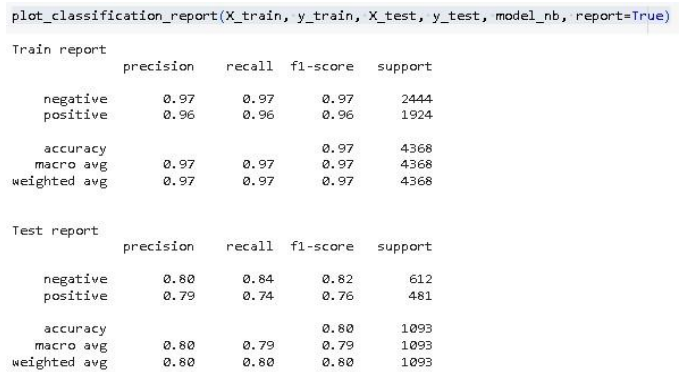
Pipeline digunakan sebagai alat untuk mengatur serangkaian langkah-langkah pemrosesan data secara berurutan, fungsi "TfidfVectorizer()" dari library "sklearn" digunakan untuk mengubah teks menjadi representasi TF-IDF.

Pipeline ini terdiri dari TF-IDF Vectorizer dan penerapan algoritma klasifikasi Naïve Bayes, keduanya disatukan dalam satu pipeline.

3) *Pengujian*

Pada tahap Pengujian, dilakukan evaluasi kinerja model klasifikasi sentimen yang telah dibuat dengan menggunakan algoritma Naïve Bayes. Proses pengujian ini melibatkan data training dan data testing yang telah dipisahkan sebelumnya.

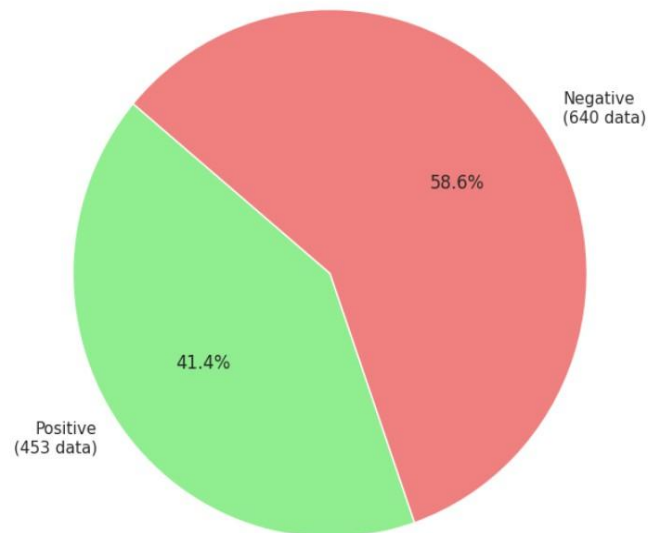
Kemudian, model yang telah ditraining dievaluasi akurasi pada data training dan data testing. Hasil akurasi data training dan testing kemudian dicetak untuk memberikan gambaran seberapa baik model ini dalam mengklasifikasikan sentimen pada data yang belum pernah dilihat sebelumnya. Berikut Adalah repot klasifikasi di gambar 5 berikut.



Gambar 5. Hasil Akurasi Pada Naïve Bayes

Kemudian berikut ini adalah hasil Visualisasi klasifikasi Naïve Bayes pada data *testing*.

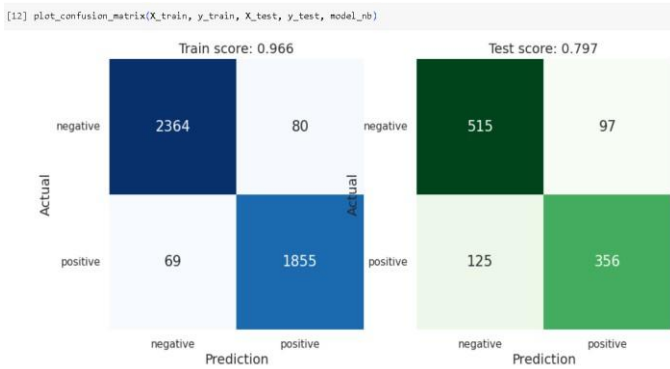
Klasifikasi Sentimen pada Data Uji - Naïve Bayes



Gambar 6. Klasifikasi *Naïve Bayes*

4) *Analisa Hasil*

Analisis hasil dilakukan dengan mengukur akurasi, presisi, recall, dan F1 Score menggunakan metrik yang di hasilkan dari confusion matrix. Confusion matrix adalah tabel yang membandingkan data aktual dengan hasil prediksi dan mencakup berbagai variabel, seperti True Positive, True Negative, False Positive, dan False Negative.



Gambar 7. Hasil Confusion Matrix

Berdasarkan hasil visualisasi confusion matrix pada data testing, model mengklasifikasikan sentimen positif dengan tepat sebanyak 356 data (True Positive). Namun, terdapat 125 data (False Negative) yang seharusnya diklasifikasikan sebagai sentimen positif, tetapi diprediksi sebagai sentimen negatif. Sementara itu, untuk sentimen negatif, model berhasil mengklasifikasikan 515 data (True Negative) dengan benar, namun terdapat 97 (False Positive) data yang seharusnya diklasifikasikan sebagai sentimen negatif, tetapi diprediksi sebagai sentimen positif. Kemudian berikut ini adalah perhitungan dan accuracy, precision, recall dan f1 score :

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Accuracy = \frac{515+356}{515+97+125+356} \times 100\% = \frac{871}{1093} \times 100\%$$

$$Accuracy = 79.7\%$$

2. precision

$$precision = \frac{TP}{TP+FP} \times 100\%$$

$$precision = \frac{356}{356+97} \times 100\%$$

$$precision = 78.6\%$$

3. recall = $\frac{TP}{TP+FN} \times 100\%$

$$recall = \frac{356}{356+125} \times 100\%$$

$$recall = 74.0\%$$

4. f1 score = $2 \times \frac{recall \times precision}{recall+precision}$

$$f1\ score = 2 \times \frac{74.0 \times 78.6}{74.0+78.6}$$

$$f1\ score = 76.2\%$$

Hasil evaluasi performa algoritma Naive Bayes pada dataset menggunakan confusion matrix menunjukkan accuracy sebesar 79.7%. accuracy ini mencerminkan sejauh mana model mampu mengklasifikasikan dengan benar antara sentimen positif dan negatif dari total data testing.

Selain accuracy, kita juga melihat metrik evaluasi lainnya seperti presisi, recall, dan F1-score. Presisi mengukur ketepatan model dalam mengklasifikasikan sentimen positif, yang dalam

kasus ini sebesar 78.6%. Recall, mengukur sejauh mana model dapat mendeteksi keseluruhan sentimen positif, mendapatkan nilai sebesar 74.0%. F1-score, sebagai rata-rata antara presisi dan recall, menghasilkan nilai sebesar 76.2%.

V. KESIMPULAN

Berdasarkan penerapan algoritma Naïve Bayes pada analisis sentimen Twitter terhadap opini publik Badan Pangan Nasional, langkah-langkah yang melibatkan crawling data, preprocessing, pelabelan data otomatis menggunakan InSet Lexicon, pembobotan kata dengan TF-IDF, data splitting, dan klasifikasi dengan algoritma Naïve Bayes telah berhasil dijalankan. Hasil klasifikasi menunjukkan akurasi sebesar 79.7%, dengan presisi 78.6%, recall 74.0%, dan F1 score 76.2%.

Dalam analisis sentimen terhadap data testing, algoritma Naïve Bayes mengklasifikasikan sebanyak 1.093 data. Dari hasil tersebut, 453 sentimen (41.4%) dikategorikan sebagai positif, sementara 640 sentimen (58.6%) dikategorikan sebagai negatif. Persentase positif dan negatif ini memberikan gambaran yang cukup jelas terkait dengan opini publik terhadap Badan Pangan Nasional di platform Twitter

DAFTAR PUSTAKA

- [1] M. Al Khadafi, K. P. Kartika, and F. Febrinita, "Penerapan metode naïve Bayes classifier Dan Lexicon based Untuk Analisis Sentimen cyberbullying Pada BPJS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 725–733, 2022. DOI: 10.36040/jati.v6i2.5633.
- [2] M. R. Amly, Y. Yusra, and M. Fikry, "Penerapan Metode Naïve Bayes Classifier Pada Klasifikasi Sentimen Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024," *Jurnal Sistem Komputer Dan Informatika (JSON)*, vol. 4, no. 4, p. 621, July 2, 2023. DOI: 10.30865/json.v4i4.6214.
- [3] M. R. F. S. Bani, U. Enri, and T. N. Padilah, "Analisis sentimen terhadap bakal calon presiden 2024 dengan algoritme Naïve Bayes," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 265, 2022. DOI: 10.30865/jurikom.v9i2.3989.
- [4] R. S. Buana et al., "Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika dengan Support Vector Machine (SVM) dan Naïve Bayes," *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, vol. 7, no. 2, pp. 194–200, April 1, 2023. DOI: 10.35870/jtik.v7i2.705.
- [5] M. H. Casandy and D. Mahdiana, "Penerapan algoritma Naïve Bayes untuk melakukan analisis sentimen pada PT Pos Indonesia (Persero)," *KRESNA Jurnal Riset Dan Pengabdian Masyarakat*, vol. 2, no. 2, pp. 221–229, 2022. DOI: 10.36080/jk.v2i2.51.
- [6] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *Jurnal Tekno Kompak*, vol. 15, no. 1, p. 131, February 17, 2021. DOI: 10.33365/jtk.v15i1.744.
- [7] D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika Dan Teknik Elektro Terapan*, vol. 10, no. 1, January 7, 2022. DOI: 10.23960/jitet.v10i1.2262.
- [8] M. N. Fadillah and R. N. Sukmana, "Penerapan algoritma naïve Bayes classifier untuk analisis sentimen Kelangkaan Minyak Goreng Pada media sosial twitter," *Infotronik: Jurnal Teknologi Informasi Dan Elektronika*, vol. 7, no. 2, p. 82, 2022. DOI: 10.32897/infotronik.2022.7.2.1716.
- [9] A. Fauzi, M. F. Akbar, and Y. F. A. Asmawan, "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *Jurnal Informatika*, vol. 6, no. 1, pp. 77–83, April 14, 2019. DOI: 10.31311/ji.v6i1.5437.
- [10] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan

- Metode Naive Bayes," *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, vol. 5, no. 3, 2019. DOI: 10.26418/jp.v5i3.34368.
- [11] F. Hasibuan, W. Priatna, and T. S. Lestari, "Analisis sentimen terhadap kementerian perdagangan pada sosial media Twitter menggunakan metode Naïve Bayes," *Techno.COM Jurnal*, vol. 21, no. 4, pp. 741–752, 2022. DOI: 10.33633/tc.v21i4.6565.
- [12] F. Koto and G. Rahmaningtyas, "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," 2017. DOI: 10.1109/IALP.2017.8300625.
- [13] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Classifiers," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, vol. 7, no. 1, p. 32, July 2, 2021. DOI: 10.24014/coreit.v7i1.12945.
- [14] A. Kusuma and A. Nugroho, "Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 15, no. 2, p. 137, December 3, 2021. DOI: 10.32815/jitika.v15i2.557.
- [15] F. A. Maulana and I. Ernawati, "Analisa Sentimen Cyberbullying Di Jejaring Sosial Twitter Dengan Algoritma Naïve Bayes," *Senamika*, vol. 1, no. 2, pp. 529–538, 2020. DOI: <https://conference.upnvj.ac.id/index.php/senamika/article/download/619/461>.
- [16] S. Mujahidin, M. N. Hasyim, and B. M. Pratama, "Implementasi analisis sentimen opini publik mengenai Sirkuit Internasional Mandalika pada Twitter menggunakan metode multinomial naïve Bayes Classifier," *Bianglala Informatika (E-journal)*, vol. 10, no. 2, 2022. DOI: 10.31294/bi.v10i2.13544.
- [17] D. Musfiroh et al., "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, March 6, 2021. DOI: 10.57152/malcom.v1i1.20.
- [18] R. Noviana and I. Rasal, "PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER," *Jurnal Teknik Dan Sains*, vol. 2, no. 2, pp. 51–60, 2023. DOI: 10.56127/jts.v2i2.791.
- [19] A. A. Permana et al., "Sentimen Analisis Opini Masyarakat Pada Media Sosial Twitter Terhadap Vaksin Berbayar Menggunakan Metode Naïve Bayes Classifier (NBC)," *Jurnal Teknik*, vol. 10, no. 2, August 6, 2021. DOI: 10.31000/jt.v10i2.5471.
- [20] N. Ruhyana, "ANALISIS SENTIMEN TERHADAP PENERAPAN SISTEM PLAT NOMOR GANJIL/GENAP PADA TWITTER DENGAN METODE KLASIFIKASI NAIVE BAYES," *ikraith-informatika*, vol. 3, no. 1, pp. 94–99, Nov. 2018.
- [21] S. Samsir et al., "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, p. 157, 2021. DOI: 10.30865/mib.v5i1.2580.
- [22] A. Turmudi Zy et al., "Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes," *Jurnal Teknologi Informatika Dan Komputer*, vol. 8, no. 2, pp. 202–213, September 30, 2022. DOI: 10.37012/jtik.v8i2.1240.