

Analisis Sentimen Masyarakat Terhadap Kebijakan Kenaikan Umk 6,5% Menggunakan Metode Naive Bayes

Farid Fitriyadi¹, Arsaela Astikasari²

^{1,2} Universitas Sahid Surakarta (Prodi Informatika, Fakultas Sains Teknologi dan Kesehatan, Universitas Sahid Surakarta, Surakarta, Indonesia)

E-mail: faridfitriyadi@gmail.com, arsaela7@gmail.com

Article History

Submitted : Dec 03, 2024
Revised : Dec 14, 2024
Accepted : Dec 23, 2024
Available Online : Jan 02, 2025
Published Regularly : Jan 02, 2025

Kata Kunci: Analisis Sentimen; Naive Bayes; UMK; Text Bloob; Twitter; KDD

Keywords: Sentiment Analysis; Naive Bayes; UMK; Text Bloob; Twitter; KDD

Contact



Author

faridfitriyadi@gmail.com

ABSTRAK

Kenaikan Upah Minimum Kabupaten/Kota (UMK) sebesar 6,5% telah memicu beragam tanggapan pro dan kontra dari masyarakat di semua sektor baik dari sisi pengusaha, buruh maupun karyawan. Maka dari itu, peneliti bermaksud melakukan analisis sentimen untuk memahami respons publik terhadap kebijakan tersebut dengan menggunakan algoritma Naive Bayes. Data penelitian dikumpulkan dari media sosial seperti twitter sejumlah 395 data tweet dengan hashtag #kenaikanumk, kemudian dilakukan pemrosesan dengan metodologi KDD yang selanjutnya dilakukan pelabelan otomatis menggunakan libray text bloob dan dikategorikan ke dalam tiga jenis sentimen: positif, negatif, dan netral. Berdasarkan hasil performance penelitian, analisis sentiment terhadap kenaikan UMK mendapatkan tingkat akurasi mencapai sebesar 95% yang mengindikasikan bahwa naïve bayes sangat efektif untuk menganalisis sentimen berbasis teks, presisi sebesar 32%, recall sebesar 33%, dan f-measure sebesar 32%. Sedangkan, hasil analisis terhadap sentimen menunjukkan bahwa sebagian besar sentimen masyarakat terhadap kebijakan ini bersifat netral, namun terdapat perbedaan signifikan antara kelompok masyarakat yang mendukung dan menentang kebijakan tersebut.

ABSTRACT

The increase in the Regency/City Minimum Wage (UMK) by 6.5% has triggered various pros and cons from the community in all sectors, both from the side of employers, workers and employees. Therefore, the researcher intends to conduct a sentiment analysis to understand the public response to the policy using the Naive Bayes algorithm. The research data was collected from social media such as Twitter, amounting to 395 tweet data with the hashtag #kenaikanumk, then processed using the KDD methodology which was then automatically labeled using the bloob text library and categorized into three types of sentiment: positive, negative, and neutral. Based on the results of the research performance, the sentiment analysis of the increase in UMK obtained an accuracy level of 95% which indicates that naïve bayes is very effective for analyzing text-based

sentiment, precision of 32%, recall of 33%, and f-measure of 32%. Meanwhile, the results of the sentiment analysis show that most of the public sentiment towards this policy is neutral, but there are significant differences between community groups who support and oppose the policy.

1. Pendahuluan

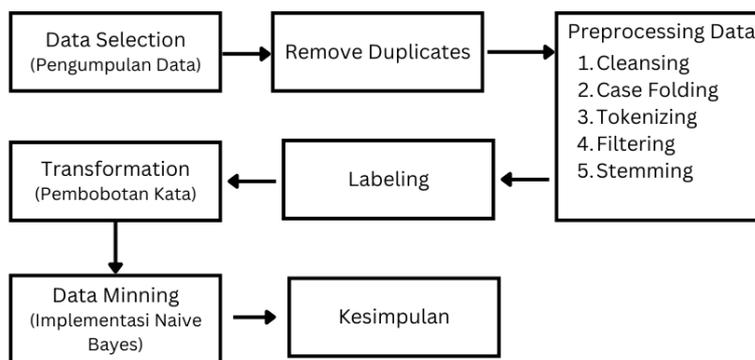
Kenaikan UMK adalah salah satu kebijakan strategis yang diambil pemerintah dalam meningkatkan kesejahteraan pekerja. Namun, kebijakan ini sering kali menuai pro dan kontra di kalangan masyarakat karena dampaknya yang signifikan terhadap berbagai aspek ekonomi. Dari sudut pandang pekerja, kenaikan UMK sering kali dipandang sebagai upaya untuk meningkatkan daya beli dan mengimbangi tekanan inflasi. Namun, bagi pengusaha, terutama di sektor UKM, kenaikan ini dianggap sebagai beban tambahan yang dapat mengurangi margin keuntungan atau bahkan mengancam keberlanjutan usaha. Situasi ini menimbulkan polarisasi opini yang menarik untuk dikaji lebih dalam, terutama melalui analisis sentimen di platform digital.

Maka dari itu, analisis sentimen sangat diperlukan untuk menggali persepsi publik secara real-time. Selain itu, penelitian ini tidak hanya bertujuan untuk memahami persepsi publik terhadap kebijakan kenaikan UMK tetapi juga memberikan wawasan yang dapat membantu pembuat kebijakan dalam menyusun langkah-langkah yang lebih responsif dan komunikatif.

Penelitian ini menggunakan metode Naive Bayes, sebuah algoritma machine learning yang sederhana namun efektif untuk mengklasifikasikan sentimen. Sentimen akan dibagi menjadi 3 kategori yaitu kategori positif, negatif, dan netral. Data penelitian didapatkan dari media sosial seperti twitter dengan hastag #kenaikanumk dalam rentang waktu 1 januari 2024 sampai dengan 31 desember 2024 sejumlah 395 data tweet, kemudian dilakukan pemrosesan dengan metodologi KDD yang selanjutnya dilakukan pelabelan otomatis menggunakan libray text blob dan dikategorikan ke dalam tiga jenis sentimen: positif, negatif, dan netral.

2. Metode Penelitian

Tahapan penelitian yang dilakukan pada penelitian ini dapat digambarkan dengan diagram alur metodologi penelitian sebagai berikut :



Gambar 1 Tahapan Penelitian

a. Data Selection atau Pengumpulan Data

Tahap awal pada penelitian ini dimulai dari proses pengumpulan data. Data didapatkan melalui proses crawling pada media sosial twitter dengan memanfaatkan library tweepy yang terdapat pada python dan fitur search Application Program Interface (API) yang di

sediakan oleh twitter. Setelah mendapatkan akses API Key maka selanjutnya kita melakukan pencarian data yang relevan dengan penelitian.

b. Remove Duplicates

Melakukan pengecekan atau eliminasi data hasil crawling yang bersifat duplikasi. Selanjutnya data tersebut akan dijadikan dataset yang akan digunakan untuk penelitian.

c. Preprocessing

Dataset yang sudah ada perlu dilakukan preprocessing untuk mengurangi noise atau pembersihan data sebelum dilakukan proses pengolahan data lebih lanjut. Dalam proses preprocessing ada beberapa tahapan yang harus dilakukan diantaranya :

1) Cleansing

Cleansing adalah tahapan dimana karakter dan tanda baca yang tidak dibutuhkan dihilangkan dari teks. Hal tersebut berfungsi untuk mengurangi noise pada dataset. Contoh karakter yang dihilangkan dalam proses ini seperti URL, tag (#), tanda baca seperti titik (.), koma (,) dan tanda baca lainnya. Tidak hanya tanda baca, cleansing juga dapat menghilangkan angka, whitespace, dan emoji yang terdapat di sebuah tweet [8].

2) Case Folding

Case folding adalah proses penyeragaman / perubahan huruf baik huruf kecil maupun huruf besar [8].

3) Tokenizing

Tokenizing adalah proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Data tweet yang diambil akan dipotong menjadi sebuah token yang dipisahkan dengan tanda petik dan koma.

4) Filtering atau Stopword Removal

Stopword removal adalah proses penghapusan pada sebuah kalimat jika ada kata / term yang tidak dianggap tidak terlalu penting. Oleh karena itu tujuan stopwords removal ini untuk menghilangkan / menghapus kata / term yang dianggap tidak penting seperti waktu, penghubung, dan lain sebagainya agar memudahkan dalam proses klasifikasi [9].

5) Stemming

Stemming adalah proses untuk menghilangkan imbuhan pada masing-masing kata yang telah diproses sehingga menjadi kata dasar, dengan tujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat [10].

d. Labeling

Text blob merupakan salah satu library dari python untuk melakukan pemrosesan bahasa alami (NLP). Library text blob akan menghitung nilai polaritas dari sebuah data untuk mendapatkan label atau sentimen pada data tersebut. Nilai polaritas merupakan suatu fungsi untuk melihat kecenderungan sentiment pada sebuah data. Hasil dari nilai polaritas akan digolongkan menjadi kelas positif, kelas netral, dan kelas negative [11].

e. Transformation atau Pembobotan Kata

Setelah data melalui tahap preprocessing dan labeling, selanjutnya akan dibuat model agar data yang mentah atau masih berupa kata-kata dapat diubah dan dihitung. Setelah diubah data akan menjadi vector kemudian diolah menggunakan algoritma naive bayes agar dapat diberi nilai dan pembobotan untuk setiap kata [12].

Salah satu metode pembobotan yang digunakan adalah Term Frequency Inverse Document Frequency (TF-IDF), yang memberikan nilai bagi setiap kata yang mengindikasikan seberapa sering kata kunci atau istilah lain muncul dalam dokumen tersebut [13].

f. Data Mining

Data training yang telah melalui tahap preprocessing, labeling serta pembobotan kata akan menjadi input pada proses training dalam menerapkan algoritma naive bayes. Setelah melalui tahap perhitungan frekuensi pada tiap-tiap dokumen, maka dilakukan tahap pengklasifikasian menggunakan algoritma naive bayes [12].

g. Kesimpulan

Mendapatkan hasil dari perhitungan dengan algoritma naive bayes dan memberikan kesimpulan dari penelitian ini.

3. Hasil dan Pembahasan

a. Pengumpulan Data

Pada tahap pengumpulan data peneliti menggunakan metode crawling twitter dengan memanfaatkan library tweepy pada python. Kata kunci pencarian yang digunakan pada penelitian ini ialah “kenaikan umk”. Berikut ini adalah hasil crawling data twitter :

	conversation_id_str	created_at	favorite_count	full_text	id_str
0	1872423130930765933	Fri Dec 27 11:38:39 +0000 2024	3	@rdmakbar @dimarcotop @irwndfry Eh koreksi sa...	1872608025355407678
1	1872586076730351766	Fri Dec 27 10:11:26 +0000 2024	1	UMK di Jogja naik 6.5% pada tahun 2025 luur. K...	1872586076730351766 https://pbs.twimg.com/ext_tw_video
2	1872416647300202857	Thu Dec 26 22:58:11 +0000 2024	0	Ada yg bilang UMR naik tax ratio meningkat. Pd...	1872416647300202857
3	1872375885677944980	Thu Dec 26 20:16:12 +0000 2024	0	Kenaikan UMK Kota Sukabumi Tahun 2025 Mulai Di...	1872375885677944980
4	1872274572323754076	Thu Dec 26 13:33:37 +0000 2024	0	Pemerintah Provinsi Riau telah menerima rekome...	1872274572323754076
...
390	1743077061114429739	Fri Jan 05 01:08:50 +0000 2024	0	umk naik sejalan dg kenaikan harga sari roti n...	1743077061114429739

Gambar 2 Hasil Crawling Data Twitter

```
[ ] import pandas as pd
import numpy as np
import re

# Membaca data dari file CSV
df = pd.read_csv('/content/tweets-data/kenaikan_umk.csv')

# Menghapus duplikasi
df = df.drop_duplicates()
```

Gambar 3 Proses Remove Duplicates

Data yang diambil dalam penelitian ini adalah data tweet dengan rentang waktu 1 januari 2024 sampai dengan 31 desember 2024. Dari hasil crawling data tersebut di dapatkan sejumlah 395 data. Data tersebut kemudian disimpan dalam bentuk .csv untuk dilakukan proses selanjutnya. Data hasil crawling ini masih dalam bentuk yang tidak terstruktur dan mengandung banyak noise, seperti duplikasi data, tanda baca, angka, simbol, dan kata-kata tidak baku yang tidak diperlukan dalam proses klasifikasi. Oleh karena itu, data tersebut harus melalui remove duplicates dan tahap Pre-Processing untuk pengolahan lebih lanjut.

b. Remove Duplicates

Pada tahap ini peneliti mengeliminasi tweet duplikat agar data bersifat unik, mencegah penambahan bobot sampel serta menghindari sifat bias pada data.

Setelah dilakukan pemindaian terhadap duplicates data tweet yang ada, tidak ditemukan data duplikasi.

Preprocessing

Tahapan yang dilakukan setelah pengumpulan data adalah tahapan Preprocessing. Pada tahap ini dilakukan pembersihan data yang diharapkan dapat mengurangi noise di dalam dataset. Sehingga proses minning akan berjalan dengan maksimal.

1. Cleansing

Proses cleansing merupakan proses pembersihan data dari komponen yang tidak diperlukan. Misalnya karakter, tanda baca, tautan, tagar, mention, dan angka. Agar data yang akan digunakan dalam proses minning dalam keadaan bersih dan data menjadi efektif. Selanjutnya, sistem akan mencari dan mengeliminasi karakter khusus yang telah ditentukan. Hasil proses cleansing dapat dilihat pada tabel dibawah ini :

Tabel 1 Hasil Proses Cleansing

No	Tweet asli	Tweet hasil cleansing
1	Menakar Dampak Kenaikan UMK. Ini Kata Pengusaha https://t.co/gqDI8axKit	Menakar Dampak Kenaikan UMK. Ini Kata Pengusaha
2	Harga beras udah naik sekitar 80ribu per karung sementara kenaikan UMK Kab. Tangerang cuma 73ribu tahun ini. Awal tahun yg berat buat keluarga buruh single income yg gajinya UMR sehat-sehat ya kalian.	Harga beras udah naik sekitar 80ribu per karung sementara kenaikan UMK Kab Tangerang cuma 73ribu tahun ini Awal tahun yg berat buat keluarga buruh single income yg gajinya UMR sehat-sehat ya kalian
3	Menko Airlangga Jadi Plt Menaker Segera Bahas Kenaikan UMP dan UMK Buruh https://t.co/R8eNMgbb1C	Menko Airlangga Jadi Plt Menaker Segera Bahas Kenaikan UMP dan UMK Buruh

2. Case Folding

Pada proses case folding dilakukan penyeragaman huruf menjadi huruf kecil semua (lower case) agar pengolahan data lebih mudah dilakukan. Pada proses ini sistem akan menelusuri semua data tweet dari awal sampai akhir dan jika ditemukan huruf kapital akan diubah menjadi huruf kecil (lower case). Hasil proses case folding dapat dilihat pada tabel dibawah ini:

Tabel 2 Hasil Proses Case Folding

No	Tweet hasil cleansing	Tweet hasil case folding
1	Menakar Dampak Kenaikan UMK. Ini Kata Pengusaha	menakar dampak kenaikan umk. Ini kata pengusaha
2	Harga beras udah naik sekitar 80ribu per karung sementara kenaikan UMK Kab Tangerang Cuma 73ribu tahun ini Awal tahun yg berat buat keluarga buruh single income yg gajinya UMR sehatsehat ya kalian	harga beras udah naik sekitar 80ribu per karung sementara kenaikan umk kab 30angerang Cuma 73ribu tahun ini awal tahun yg berat buat keluarga buruh single income yg gajinya umr sehatsehat ya kalian
3	Menko Airlangga Jadi Plt Menaker Segera Bahas Kenaikan UMP dan UMK Buruh	menko airlangga jadi plt menaker segera bahas kenaikan ump dan umk buruh

3. Tokenizing

Pada tahap ini dilakukan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Data tweet yang diambil akan dipotong menjadi sebuah token yang dipisahkan dengan tanda petik dan koma. Hasil proses tokenizing, dapat dilihat pada tabel dibawah ini :

Tabel 3 Hasil Proses Tokenizing

No	Tweet hasil case folding	Tweet hasil tokenizing
1	menakar dampak kenaikan umk. Ini kata pengusaha	['menakar', 'dampak', 'kenaikan', 'umk', 'ini', 'kata', 'pengusaha']
2	harga beras udah naik sekitar 80ribu per karung sementara kenaikan umk kab 31angerang Cuma 73ribu tahun ini awal tahun yg berat buat keluarga buruh single income yg gajinya umr sehatsehat ya kalian	['harga', 'beras', 'udah', 'naik', 'sekitar', '80ribu', 'per', 'karung', 'sementara', 'kenaikan', 'umk', 'kab', 'tangerang', 'Cuma', '73ribu', 'tahun', 'ini', 'awal', 'tahun', 'yg', 'berat', 'buat', 'keluarga', 'buruh', 'single', 'income', 'yg', 'gajinya', 'umr', 'sehatsehat', 'ya', 'kalian']
3	menko airlangga jadi plt menaker segera bahas kenaikan ump dan umk buruh	['menko', 'airlangga', 'jadi', 'plt', 'menaker', 'segera', 'bahas', 'kenaikan', 'ump', 'dan', 'umk', 'buruh']

4. Stopword removal atau filtering

Pada tahap ini akan dilakukan pemindaian semua data tweet dari awal sampai akhir dan jika ada kata / term yang tidak dianggap tidak terlalu penting akan dilakukan penghapusan atau dihilangkan misalnya kata penghubung, waktu dan lain sebagainya agar memudahkan dalam proses klasifikasi. Hasil proses stopwords removal atau filtering, dapat dilihat pada tabel dibawah ini :

Tabel 4 Hasil proses stopwords removal atau filtering

No	Tweet hasil tokenizing	Tweet hasil Stopword removal / filtering
1	['menakar', 'dampak', 'kenaikan', 'umk', 'ini', 'kata', 'pengusaha']	['menakar', 'dampak', 'kenaikan', 'umk', 'pengusaha']
2	['harga', 'beras', 'udah', 'naik', 'sekitar', '80ribu', 'per', 'karung', 'sementara', 'kenaikan', 'umk', 'kab', 'tangerang', 'cuma', '73ribu', 'tahun', 'ini', 'awal', 'tahun', 'yg', 'berat', 'buat', 'keluarga', 'buruh', 'single', 'income', 'yg', 'gajinya', 'umr', 'sehatsehat', 'ya', 'kalian']	['harga', 'beras', 'udah', '80ribu', 'karung', 'kenaikan', 'umk', 'kab', 'tangerang', '73ribu', 'berat', 'keluarga', 'buruh', 'single', 'income', 'gajinya', 'umr', 'sehatsehat']
3	['menko', 'airlangga', 'jadi', 'plt', 'menaker', 'segera', 'bahas', 'kenaikan', 'ump', 'dan', 'umk', 'buruh']	['menko', 'airlangga', 'plt', 'menaker', 'bahas', 'kenaikan', 'ump', 'umk', 'buruh']

5. Stemming

Pada tahap ini akan dilakukan pemindaian semua data tweet dengan menghilangkan kata yang berimbuhan pada masing-masing kata sehingga menjadi kata dasar. Hasil proses stemming, dapat dilihat pada tabel dibawah ini :

Tabel 5 Hasil Proses Stemming

No	Tweet hasil tokenizing	Tweet hasil Stemming
1	['menakar', 'dampak', 'kenaikan', 'umk', 'pengusaha']	['takar', 'dampak', 'naik', 'umk', 'usaha']
2	['harga', 'beras', 'udah', '80ribu', 'karung', 'kenaikan', 'umk', 'kab', 'tangerang', '73ribu', 'berat', 'keluarga', 'buruh', 'single', 'income', 'gajinya', 'umr', 'sehatsehat']	['harga', 'beras', 'udah', '80ribu', 'karung', 'naik', 'umk', 'kab', 'tangerang', '73ribu', 'berat', 'keluarga', 'buruh', 'single', 'income', 'gaji', 'umr', 'sehatsehat']
3	['menko', 'airlangga', 'plt', 'menaker', 'bahas', 'kenaikan', 'ump', 'umk', 'buruh']	['menko', 'airlangga', 'plt', 'menaker', 'bahas', 'naik', 'ump', 'umk', 'buruh']

c. Labelling

Proses labelling merupakan proses untuk mengetahui nilai sentimen dari sebuah tweet yang dapat diberikan label positif, negatif, dan netral. Pada tahap ini peneliti memanfaatkan library dari python yaitu “**text blob**” untuk proses pelabelan sentimen. Library text blob akan menghitung nilai polaritas dari sebuah data untuk mendapatkan label atau sentimen pada data tersebut. Nilai polaritas merupakan suatu fungsi untuk melihat kecenderungan sentiment pada sebuah data. Hasil dari nilai polaritas akan digolongkan menjadi kelas positif, kelas netral, dan kelas negative. Contoh hasil dari proses labelling dapat dilihat pada tabel 7.

Tabel 6 Hasil Proses Labeling

NO	Tweet yang sudah dicleansing	Hasil Labeling
1	Menakar Dampak Kenaikan UMK. Ini Kata Pengusaha	Neutral
2	Harga beras udah naik sekitar 80ribu per karung sementara kenaikan UMK Kab Tangerang cuma 73ribu tahun ini Awal tahun yg berat buat keluarga buruh single income yg gajinya UMR sehat-sehat ya kalian	Negative
3	Menko Airlangga Jadi Plt Menaker Segera Bahas Kenaikan UMP dan UMK Buruh	Neutral

Setelah dilakukan labeling. Hasil labeling dengan **text blob** menunjukkan bahwa data tweet yang mengandung sentiment negative sejumlah 11 label, sentiment neutral sejumlah 376 label, sentiment positive sejumlah 8 label. Dengan total data sejumlah 395 data.

d. Transformation atau Pembobotan Kata

Setelah dilakukan tahapan preprocessing dan labeling, maka dataset siap untuk dilakukan klasifikasi. Langkah selanjutnya adalah Transformation atau Pembobotan Kata. Salah satu metode pembobotan yang akan digunakan adalah Term Frequency Inverse Document Frequency (TF-IDF). TF-IDF merepresentasikan teks sebagai vektor numerik berdasarkan bobot pentingnya kata-kata dalam dokumen. TF-IDF mempertimbangkan frekuensi kata dalam dokumen (TF) dan seberapa jarang kata tersebut muncul di seluruh korpus (IDF).

Berikut ini adalah Hasil dari Pembobotan Kata dengan TF-IDF :

ukur	ulas	umk	umk2025	umkbadung	umkharus	umkjatim2025	umkjogja	umkkabmojokerto2025	umkm
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.04337021618933565	0.0	0.0	0.0	0.0	0.2534481684204311	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.1163191814350891	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0855746113763872	0.21360706960750034	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.07654861220402727	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.12139410907300274	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.03663683999161032	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.05566107384428145	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.09741714483169302	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 4 Hasil Pembobotan Kata dengan TF-IDF

Berikut adalah nominasi kata dengan bobot terbanyak :

word	total_tfidf
umk	30.861384510915375
naik	29.834286800870085
2025	20.263989802453235
persen	15.334496471311816
upah	15.086133859432968
gaji	13.707686907777463
buruh	13.408716149051852
kerja	10.200627073299266
minimum	10.084173439570618
ump	10.020871787790686

Gambar 5 Nominasi Pembobotan Kata Terbanyak

e. Implementasi Naïve Bayes

Langkah selanjutnya adalah Data Mining (proses pengolahan data) dengan pengimplementasian naïve bayes model Multinomial Naive Bayes yang merupakan salah satu algoritma Naive Bayes yang diimplementasikan menggunakan library sklearn pada Python untuk dilakukan pengklasifikasian. Algoritma naïve bayes merupakan algoritma untuk metode klasifikasi dengan membagi dataset menjadi data training dan data testing.

1) Pengujian Klasifikasi

Pada penelitian ini, data diproses menggunakan algoritma Naive Bayes dengan skenario pembagian data 80:20 (80% data training dan 20% data testing).

Berikut adalah hasil dari pengujian klasifikasi menggunakan algoritma naïve bayes :

```
Laporan Klasifikasi:
              precision    recall  f1-score   support

   negative         0.00         0.00         0.00         2
    neutral         0.95         1.00         0.97        75
   positive         0.00         0.00         0.00         2

 accuracy                   0.95         79
 macro avg         0.32         0.33         0.32         79
 weighted avg         0.90         0.95         0.92         79
```

Gambar 6 Hasil Pengujian Klasifikasi

Berdasarkan Gambar 7, hasil klasifikasi menggunakan Naïve Bayes pada perbandingan 80:20 menunjukkan nilai akurasi yang tinggi sebesar 95%, presisi sebesar 32%, recall sebesar 33%, dan f-measure sebesar 32%.

2) Visualisasi Data

Hasil klasifikasi analisis sentimen terhadap ulasan tentang Analisis Sentimen Masyarakat Terhadap Kebijakan Kenaikan Umk 6,5% dapat divisualisasikan menggunakan word cloud untuk memberikan gambaran atau informasi umum tentang data sentimen tersebut

DAFTAR PUSTAKA

- [1] R. N. Fahmi, N. and A. Primajaya, "PENEMBAKAN LASKAR FPI OLEH POLRI DENGAN METODE NAIVE BAYES CLASSIFIER," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, <https://ejournal.akakom.ac.id/index.php/jiko/article/view/437>, p. 2, 2021.
- [2] P. Arsi, B. A. Kusuma and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier," *JURNAL INFORMATIKA UPGRIS*, vol. 7 no.1, <https://journal.upgris.ac.id/index.php/JIU/article/view/7636/4304>, JUNI 2021.
- [3] W. Yulita, E. D. Nugroho and M. H. Algifari, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, Vols. 2, No. 2, <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344/672>, pp. 1-9, 2021.
- [4] A. Safira and F. N. Hasan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Sistem Informasi*, Vols. 5, No.1, <https://journal.unilak.ac.id/index.php/zn/article/view/12856>, pp. 59-70, Januari 2023.
- [5] S. Puad, G. and A. S. Y. Irawan, "ANALISIS SENTIMEN MASYARAKAT PADA TWITTER TERHADAP PEMILIHAN UMUM 2024 MENGGUNAKAN ALGORITMA NAÏVE BAYES," *Jurnal Mahasiswa Teknik Informatika*, Vols. 7, No.3, <https://ejournal.itn.ac.id/index.php/jati/article/view/6920>, Juni 2023.
- [6] M. W. A. Putra, S. E. and H. , "Analisis Sentimen Dompot Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier," *IT Journal Research and Development (ITJRD)*, Vols. 5, No.1, 10.25299/itjrd.2020.vol5(1).5159, p. 72 – 86, Agustus 2020.
- [7] R. W. Utami, A. Jazuli and T. Khotimah, "ANALISIS SENTIMEN TERHADAP XIAOMI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES," *Indonesian Journal of Technology, Informatics and Science (IJTIS)*, Vols. 3, No. 1, 10.24176/ijtis.v3i1.7514, pp. 21-30, Desember 2021.
- [8] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes," *Jurnal Data Mining dan Sistem Informasi*, vol. 2(2), pp. 1-9, 2021.
- [9] F. Fatayat and R. Nugroho, "Analisa Penentuan Dosen Pembimbing Tugas Akhir Mahasiswa Menggunakan Naive Bayes Classifier," *Simtika*, vol. 4(3), pp. 1-7, 2021.
- [10] N. Wardani and A. Erfina, "Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes," *In Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, Vols. 1, No. 01, pp. 11-18, 2021.
- [11] A. Baita, Y. Pristyanto and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," *Information System Journal*, vol. 4(2), pp. 42-46, 2021.
- [12] K. D. Indarwati and H. Februariyanti, "ANALISIS SENTIMEN TERHADAP KUALITAS PELAYANAN APLIKASI GO-JEK MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *JATISI*, vol. 1, no. <https://doi.org/10.35957/jatisi.v10i1.2643>, p. 1, 2023.
- [13] I. Rozi, R. Ardiansyah and N. Rebeka, "Penerapan Normalisasi Kata Tidak Baku Menggunakan Levenshtein Distance pada Analisa Sentimen Layanan PT. KAI di Twitter.," *Seminar Informatika Aplikatif*, no. <http://jurnalti.polinema.ac.id/index.php/SIAP/art>, pp. 106-112, 2019.