1413 PENINGKATAN MODEL ANALISIS SENTIMEN MELALUI ALGORITMA NAIVE BAYES BERDASARKAN DATA KOMENTAR YOUTUBE

By Deden Syarif Hidayat

WORD COUNT 3906 TIME SUBMITTED 31-DEC-2024 12:13PM

PAPER ID 113798188

PENINGKATAN MODEL ANALISIS SENTIMEN MELALUI ALGORITMA *NAIVE BAYES* BERDASARKAN DATA KOMENTAR *YOUTUBE*

DedenSyarifHidayat[1],

Abstract

The implementation of the QR Code-based Fuel Oil subsidy policy for four-wheeled vehicles, which began on October 1, 2024, has drawn various responses from the public. This policy aims to ensure that the distribution of subsidized fuel is more targeted, but has caus mixed reactions, especially regarding the registration process and the specified vehicle capacity limits. This study aims to analyze public sentiment towards the policy using comment data from a YouTube video entitled "Listen! New Rules Criteria for Using Subsidized Fuel - SIP 02/09" uploaded to the Seputar iNews RCTI YouTube channel. The method used is sentiment analysis based on the Naive Bayes algorithm. data preprocessing process includes text cleaning, tokenization, removing stopwords, and stemming to ensure that the analyzed data is clean and structured. The data is divided into ning and test data to build and evaluate the model. Model performance is measured using accuracy, precision, and recall metrics. The results of the analysis show that the majority comments show negative sentiment, reflecting public dissatisfaction with this policy. Negative sentiment is mainly caused by the complexity of the QR Code registration process and the limited criteria for eligible vehicles. This study highlights the importance of more effective communication strategies from the government to increase public understanding and acceptance. In addition, this study opens up opportunities for further development in the use of computational sentiment analysis in public policy studies.

Keywords : *sentiment analysis, naive bayes, subsidized fuel policy, QR code, YouTube*

Abstrak

Penerapan kebijakan subsidi Bahan Bakar Minyak (BBM) berbasis *QR Code* untuk kendaraa da empat, yang dimulai pada 1 Oktober 2024, memunculkan berbagai tanggapan dari masyarakat. Kebijakan ini bertujuan untuk memastikan distribusi BBM bersubsidi lebih tepat sasaran, namun menimbulkan reaksi agam, terutama terkait proses pendaftaran dan batasan kapasitas kendaraan yang ditentukan. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan tersebut dengan menggunakan data komentar dari video *YouTube* berjudul "Simak! Aturan Baru K eria Penggunaan BBM Bersubsidi - *SIP* 02/09" yang diunggah pada kanal *YouTube* Seputar *iNews RCTI*. Metode yang digunakan adalah analisis sentimen berbasis algoritma *Naive Bayes*. Proses *preprocessing* data meliputi pembersihan teks, tokenisasi, penghilangan *stopwords*, dan *stemming* untuk memastikan data yang dianalisis bersih dan terstruktur. Data dibagi menjadi data pelatihan dan uji untuk membangun serta mengevaluasi model. Kinerja model diukur menggunakan metrik *accuracy*, *precision*, dan *recall*. Hasil analisis menunjukkan bahwa mayoritas komentar menunjukkan sentimen negatif, mencerminkan ketidakpuasan masyarakat terhadap kebijakan ini. Sentimen negatif utamanya disebabkan oleh kompleksitas proses pendaftaran *QR Code* dan keterbatasan kriteria kendaraan yang memenuhi syarat. Penelitian ini menyoroti pentingnya strategi komunikasi yang ih efektif dari pemerintah untuk meningkatkan pemahaman dan penerimaan masyarakat. Selain itu, penelitian ini membuka peluang untuk pengembangan lebih lanjut dalam pemanfaatan analisis sentimen berbasis komputasi pada studi kebijakan publik.

Kata kunci : Analisis Sentimen, *Naïve Bayes*, Kebijakan BBM Bersubsidi, *QR Code* BBM, Komentar Youtube.

---

1. PENDAHULUAN

Kebutuhan energi yang terus meningkat di Indonesia telah mendor g pemerintah untuk memperbarui kebijakan terkait distribusi bahan bakar minyak (BBM) bersubsidi guna memastikan alokasi yang lebih tepat sasaran. Salah satu kebijakan baru yang diterapkan mulai 1 Oktober 2024 adalah penggunaan sistem QR Code untuk kendaraan roda empat, di mana pembeli BBM bersubsidi diwajibkan mendaftar melalui platform digital resmi Pertamina. Kebijakan ini bertujuan mengoptimalkan anggaran negara, mengurangi konsumsi BBM yang tidak terkendali, dan meningkatkan

efisiensi distribusi sebagaimana tercermin dalam penurunan kuota subsidi dalam RAPBN 2025 menjadi 19,41 juta kiloliter (CNN Indonesia, 2024).

Namun, kebijakan ini menimbulkan beragam reaksi di masyarakat, mulai dari dukungan hingga kritik, terutama terkait dengan proses pendaftaran yang dinilai rumit dan pembatasan kapasitas kendaraan. Beberapa kelompok masyarakat menganggap kebijakan ini tidak adil karena berpotensi membatasi akses subsidi bagi kelompok tertentu. Di sisi lain, kebijakan ini juga mendapat dukungan dari sebagian pihak yang melihatnya sebagai langkah strategis dalam mengatasi ketidakefisienan subsidi BBM. Setiap perubahan, apapun struktur dan alasannya, akan selalu menghadapi upaya oposisi dari beberapa kelompok yang kurang stabil terhadap perubahan atau juga pihak-pihak yang kurang berharap dalam hal tercapainya suatu perubahan [1].

Dengan memanfaatkan analisis sentimen, penelitian ini bertujuan mengevaluasi respons masyarakat terhadap kebijakan tersebut. Data komentar dari video *YouTube* berjudul "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - *SIP* 02/09" digunakan untuk memahami pola sentimen positif dan negatif. Teknik analisis sentimen berbasis algoritma *Naive Bayes* dipilih karena keandalannya dalam klasifikasi data teks.

Tujuan dari penelitian ini adalah untuk mengidentifikasi sentimen masyarakat terhadap kebijakan tersebut serta mengukur *accuracy* model *Naive Bayes* dalam mengklasifikasikan sentimen positif dan negatif. Penelitian ini diharapkan memberikan kontribusi praktis berupa rekomendasi kepada pemerintah dalam merancang kebijakan yang lebih efektif serta mendukung formulasi strategi komunikasi untuk meningkatkan penerimaan publik.

2. TINJAUAN PUSTAKA DAN TEORI

2.1 Tinjauan Pustaka

Berbagai penelitian telah dilakukan terkait analisis sentimen pada kebijakan publik menggunakan algoritma *Naive Bay* Penelitian oleh Ardiansyah dan Nur'aini [2] menganalisis sentimen masyarakat terhadap kenaikan harga BBM melalui platform Twitter. Dengan *dataset* sebanyak 2312 data yang telah diproses, algoritma *Naive Bayes* menunjukkan *accuracy* sebesar 92%, menjadikannya andal dalam menganalisis respons masyarakat terhadap kebijakan.

Penelitian lainnya, Base et al. [3], membandingkan efektivitas tiga algoritma *Naive Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbo (KNN)* dalam menganalisis sentimen publik terhadap kenaikan harga BBM. Hasilnya menunjukkan bahwa *Naive Bayes* memiliki *accuracy* tertinggi sebesar 80,28%, mengungguli algoritma lainnya.

Sementara itu, Rahmadani et al [4] menggunakan *Naive Bayes* dan metode *TF-IDF* untuk menganalisis sentimen ulasan pengguna *game* di Google Play Store. Hasil penelitian menunjukkan *accuracy* sebesar 84,12%, menegaskan efektivitas kombinasi algoritma *Naive Bayes* dengan teknik ekstraksi fitur *TF-IDF* dalam analisis sentimen.

Di sisi lain, penelitian oleh Muliawaan Dazki [5] menggunakan Twitter untuk menganalisis sentimen publik terhadap kebijakan pemindahan ibu kota negara. Algoritma *Naive Bayes* menunjukkan *accuracy* sebesar 65,26%, yang lebih tinggi dibandingkan *KNN* dan *Random Forest*. Penelitian ini menyoroti distribusi sentimen publik, yaitu 36,8% positif, 25% netral, dan 38,1% negatif, memberikan gambaran mendalam tentang persepsi masyarakat.

Dari studi-studi sebelumnya, analisis sentimen telah terbukti memberikan wawasan penting dalam memahami respons masyarakat terhadap kebijakan publik. Namun, penelitian terdahulu sebagian besar berfokus pada platform seperti Twitter atau ulasan aplikasi, sementara analisis sentimen berbasis komentar *YouTube* belum dieksplorasi secara mendalam. Selain itu, kebijakan terkait penggunaan *QR Code* untuk subsidi BBM belum menjadi subjek penelitian yang signifikan.

Oleh karena itu, penelitian ini menawarkan kebaharuan dengan memanfaatkan data komentar *YouTube* untuk mengevaluasi sentimen masyarakat terhadap kebijakan BBM bersubsidi berbasis *QR Code*. Penelitian ini juga mengeksplorasi kata-kata dominan yang muncul dalam sentimen positif dan negatif, sehingga memberikan kontribusi baru dalam studi kebijakan publik berbasis analisis sentimen digital.

2.1 Analisis Sentimen

Analisis sentimen atau disebut juga *opinion mining* adalah bidang studi yang bertujuan untuk menganalisis sentimen, opini, sikap, emosi, evaluasi, dan penilaian masyarakat terhadap suatu layanan, produk, individu, organisasi, peristiwa, topik, atribut, dan masalah [6].

2.2 Text Mining

Text mining atau penambangan teks diartikan sebagai proses ekstraksi pola berupa informasi sebuah data besar yang tidak berstruktur berstruktur dan bentuknya berupa teks [7].

2.3 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF adalah metode untuk menentukan sejauh mana suatu term mencerminkan konten dalam suatu dokumen dengan memberi skor pada masing-masing kata yang , nilai *TF-IDF* diperoleh melalui perkalian nilai *TF (Term Frequency)* dengan nilai *IDF (Inverse Document Frequency)* [8].

2.4 Naive Bayes

Naive Bayes adalah suatu pendekatan tanpa aturan yang mengandalkan cabang matematika yang dikenal sebagai teori probabilitas untuk memperoleh probabilitas seoptimal mungkin dengan memeriksa frekuensi atau jumlah kemunculan setiap klasifikasi dalam data pelatihan [9].

3. METODOLOGI PENELITIAN

3.1. TAHAPAN PENELITIAN

Penelitian ini terdiri dari beberapa tahapan utama berdasarkan pendekatan KDD, yang meliputi *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation* [10]. Tabel 1 menunjukkan setiap tahapan penelitian beserta aktivitas dan deskripsi yang dilakukan.

Tahapan-tahapan ini dijelaskan sebagai berikut:

1. *Data Selection*
   Memilih data yang relevan untuk dianalisis, dalam hal ini berupa komentar *YouTube* terkait kebijakan.
2. *Preprocessing*
   Pre-processing merupakan tahapan yang terfokus pada pembersihan data, dengan tujuan mengatasi gangguan, duplikasi, inkonsistensi, dan kesalahan tipografi dalam konteks proses data mining [11].
3. *Transformation*
   Mengubah teks menjadi representasi numerik menggunakan metode *TF-IDF* agar dapat diproses oleh algoritma.
4. *Data Mining*
   Proses *Data Mining* merupakan proses di mana pengolahan data dengan menerapkan algoritma [12]. Peneliti menerapkan algoritma *Naive Bayes* untuk mengklasifikasikan data menjadi sentimen positif dan negatif.
5. *Evaluation*
   Mengevaluasi performa model menggunakan metrik seperti *accuracy*, *precision*, dan *recall*.
6. *Reporting*
   Menampilkan pengetahuan.

Tabel 1. Tahapan Penelitian

Tahapan	Aktivitas	Deskripsi
Data Selection	Pengumpulan data	Mengambil data komentar YouTube dari video terkait.
Preprocessing	Pelabelan Data	Melabeli komentar sebagai positif, netral, atau negatif.
	Penyaringan Data	Menghapus komentar netral, fokus pada positif dan negatif
	Pembersihan Data	Menghapus elemen seperti link, HTML, dan simbol dari komentar.
	Peran Dan Seleksi Atribut Serta Pembersihan Data	Melakukan peran terhadap atribut label untuk analisis tahap selanjutnya, menyeleksi atribut yang dibutuhkan dan
Transformation	Representasi Data	Mengubah teks komentar menjadi vektor numerik dengan TF-IDF.
	Pemrosesan Teks	Melakukan case folding, tokenisasi, stemming, dan filter stopwords.
Data Mining	Klasifikasi sentimen	Menerapkan algoritma *Naive Bayes* untuk klasifikasi positif dan negatif.
Data Mining	Pembagian Dataset	Membagi data menjadi 70% untuk latih dan 30% untuk uji.
	Penerapan Algoritma	Menggunakan Naïve Bayes untuk klasifikasi sentimen..
Evaluation	Pengukuran Performa	Mengevaluasi kinerja model dengan *accuracy*, *precision*, dan *recall*.
Reporting	Penyajian Hasil	Menampilkan kata-kata yang sering muncul dalam kategori positif dan negatif dalam tabel frekuensi dan *word cloud*.
	Analisis dan Rekomendasi	Menyimpulkan hasil untuk memberikan wawasan kepada pengambil kebijakan.
Reporting	Penyajian Hasil	Menampilkan kata-kata yang sering muncul dalam kategori

		positif dan negatif dalam tabel frekuensi dan *word cloud*.

3.2. Pengumpulan Data

Sumber data untuk penelitian ini yaitu data sekunder. Data sekunder ialah sumber data yang tidak diperoleh langsung oleh peneliti [13]. Data diperoleh dari komentar pada video *YouTube* berjudul "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - SIP 02/09". Komentar dikumpulkan menggunakan teknik *crawling* dengan total 1.416 komentar.

3.3. Analisa Data

Analisis data dilakukan untuk memahami opini masyarakat terhadap kebijakan subsidi BBM berbasis *QR Code*. Data diambil dari komentar pengguna *YouTube* pada video "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - *SIP* 02/09" di kanal Seputar *iNews RCTI*. Komentar dianalisis menggunakan algoritma *Naive Bayes* dengan tahapan *preprocessing*, *transformation* menggunakan *TF-IDF*, dan evaluasi model. Data komentar dipilih berdasarkan relevansi, kemudian dibersihkan dari elemen tidak penting, diberi label sentimen positif atau negatif, dan dikonversi menjadi representasi numerik. Hasil analisis memberikan wawasan mengenai persepsi publik terhadap kebijakan ini, yang mayoritas menunjukkan sentimen negatif. Pendekatan ini memungkinkan identifikasi isu utama yang memengaruhi penerimaan masyarakat terhadap kebijakan tersebut.

4. HASIL DAN PEMBAHASAN

Memuat hasil, pengujian dan pembahasan terkait analisis sentimen yang telah dilakukan pada Bab 3.

4.1 HASIL PENELITIAN

4.1.1 DATA SELECTION

Peneliti mengumpulkan data komentar dari video "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - *SIP* 02/09" di kanal Seputar *iNews RCTI* menggunakan *API YouTube* dengan skrip *Python* yang memanfaatkan pustaka *googleapiclient*. Data yang dikumpulkan meliputi waktu publikasi, nama pengguna, dan isi komentar sebanyak 1.416 baris, yang disimpan dalam file Excel *(.xlsx)* untuk analisis lebih lanjut.

```python
# Block 1: Import library yang dibutuhkan
import pandas as pd
from googleapiclient.discovery import build

# Block 2: Fungsi untuk mengambil komentar dari video
def video_comments(video_id, api_key):
    replies = []
    youtube = build('youtube', 'v3', developerKey=api_key)

    # Mengambil komentar utama dan balasan
    video_response = youtube.commentThreads().list(
        part='snippet,replies',
        videoId=video_id
    ).execute()

    # Iterasi untuk mendapatkan semua komentar
    while video_response:
        for item in video_response['items']:
            # Informasi komentar utama
            published = item['snippet']['topLevelComment']['snippet']['publishedAt']
            user = item['snippet']['topLevelComment']['snippet']['authorDisplayName']
            comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
            replies.append([published, user, comment])

            # Jika ada balasan
            replycount = item['snippet']['totalReplyCount']
            if replycount > 0:
                for reply in item['replies']['comments']:
                    published = reply['snippet']['publishedAt']
                    user = reply['snippet']['authorDisplayName']
                    repl = reply['snippet']['textDisplay']
                    replies.append([published, user, repl])

        # Mengambil halaman berikutnya jika ada
        if 'nextPageToken' in video_response:
            video_response = youtube.commentThreads().list(
                part='snippet,replies',
                pageToken=video_response['nextPageToken'],
                videoId=video_id
            ).execute()
        else:
            break
    return replies

# Block 3: Informasi API Key dan Video ID
api_key = 'XXXXXXXXXXXXXXXXXXXXXXX'  # Ganti dengan API Key Anda
video_id = "GrqynETeqns"  # Ganti dengan ID video Anda
comments = video_comments(video_id, api_key)

# Block 4: Konversi hasil ke DataFrame dan simpan ke file Excel
df = pd.DataFrame(comments, columns=['publishedAt', 'authorDisplayName', 'textDisplay'])
df.to_excel('qrcode-pertamina-subsidi.xlsx', index=False)

print("Data berhasil disimpan ke 'qrcode-pertamina-subsidi.xlsx'")
```

Gambar 1. *Crawling* Data

Gambar 2. Data Hasil *Crawling*

Setelah tahap *crawling* didapatkan data berformat excel yang berisi waktu publikasi *(publishedAt)*, nama pengguna *(authorDisplayName)*, dan isi komentar *(textDisplay)* diperoleh sebanyak 1.416 baris pada periode dari komentar pengguna yang dipublikasikan hingga tanggal 8 November 2024 pukul 22:33.

4.1.2 PREPROCESSING

Tahap *preprocessing* merupakan kegiatan mempersiapkan data mentah sebelum memulai proses selanjutnya, berikut tahapan-tahapannya:

1. Penambahan Nomor

   Menambahkan atribut "No." pada data excel hasil crawling pada tahap Data Selection untuk mempermudah identifikasi data selama analisis.

2. Pelabelan Data

   Data dilabeli secara manual menjadi 3 kategori: positif, netral dan negatif sesuai dengan arahan pakar.

Setelah pelabelan terhadap komentar tersebut diperoleh distribusi sentimen sebagai berikut:

a. Positif: 135 komentar

Tabel 2. Contoh Komentar Positif

No	*textDisplay*	label
1	Setuju	positif
2	Lanjutkn	positif
3	Bagusss.. Yg penting pertalite jangan di hilang kan...	positif

b. Netral: 412 komentar

Tabel 3. Contoh Komentar Netral

No	*textDisplay*	label
1	Silahkan berkomentar para netizen	netral
2	Lebih sehat jln kaki pak😁	netral
3	Mw tnya..kalau pertalit ngk supsdi hrga ny brpa y	netral

c. Negatif: 869 komentar

Tabel 4. Contoh Komentar Negatif

No	*textDisplay*	label
1	Mencekik rakyat	negatif
2	Jannn ribetttttttt	negatif
3	Ruwet semua aturran di rubah	negatif
4	Ribet poll...,,😂😂😂	negatif

3. Penyaringan Data

Data berlabel netral dihapus karena sesuai dengan tujuan penelitian yang berfokus pada kategori positif dan negatif. Total data berjumlah 1.004.

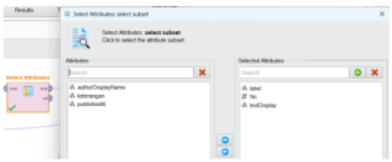
4. Peran Atribut

Dalam langkah ini, peneliti menggunakan aplikasi AI Studio 2024.1.0 untuk memberikan peran atribut sebagai label dengan operator Set Role, agar data siap digunakan dalam proses selanjutnya. Langkah ini juga melibatkan penggunaan operator Read Excel untuk mengambil data mentah dari file yang telah disimpan.

Gambar 3. Peran Atribut

5. Seleksi Atribut

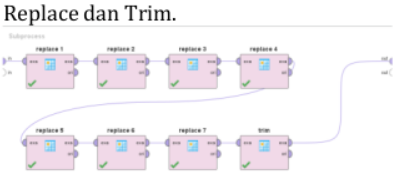
Langkah ini digunakan untuk memilih atribut untuk kebutuhan analisis, atribut yang digunakan yaitu No., label dan textDisplay.



Gambar 4. Seleksi Atribut

6. Pembersihan Data

Pembersihan data dilakukan pada atribut textDisplay dengan menghapus elemen-elemen seperti link, sintaks HTML, balasan nama pengguna, simbol, angka, hashtag, karakter non-alfabet, dan spasi berlebih menggunakan AI Studio 2024.1.0, dengan memanfaatkan operator Replace dan Trim.



Gambar 5. Operator Pembersihan Data

Informasi mengenai operator pembersihan data:

1. *replace 1*: Pembersihan *link* menggunakan pola "[^<]*".
2. *replace 2*: Pembersihan *sintaks HTML*
 dengan menggunakan pola "<br>".
3. *replace 3*: Menghapus nama balasan pengguna menggunakan pola "@@\w+".
4. *replace 4*: Menghapus simbol berlebih menggunakan pola "[\.,/]+".
5. *replace 5*: Menghapus semua angka menggunakan pola "\d+".
6. *replace 6*: Menghapus *hashtag* menggunakan pola "#\w+".

7. *replace 7*: Menghapus semua karakter *non*-alfabet menggunakan pola "[^a-zA-Z\s]+".
8. *trim*: Menghilangkan spasi berlebih.

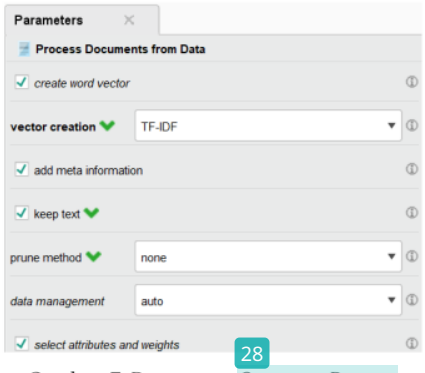
Tabel 5. Contoh Hasil Setelah Pembersihan

No	Sebelum	Sesudah
1	Setuju	Setuju
2	Lanjutkn	Lanjutkn
3	Bagusss.. Yg penting pertalite jangan di hilang kan...	Bagusss Yg penting pertalite jangan di hilang kan
4	Mencekik rakyat	Mencekik rakyat
5	Jannn ribetttttttt	Jannn ribetttttttt
6	Ruwet semua aturran di rubah	Ruwet semua aturran di rubah
7	Ribet poll...,😂😂😂	Ribet poll

4.1.3 *TRANSFORMATION*

Pada tahap *Transformation* digunakan operator *Process Documents from Data* menggunakan parameter *vector creation TF-IDF* untuk merubah data teks menjadi representasi numerik, di dalam operator tersebut terdiri dari beberapa langkah yaitu:

Gambar 6. Operator *Process Documents From Data*



Gambar 7. Parameter *Operator Process Documents From Data*

Gambar 8. Operator Di Dalam *Process Documents From Data*

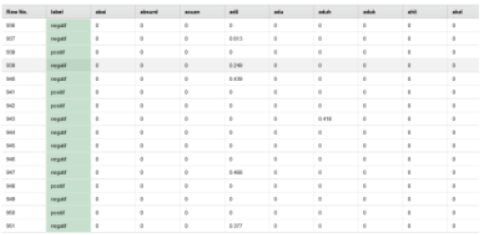
Informasi Mengenai Operator Di Dalam *Process Documents From Data*:

a. *Transform Cases*
   Proses mengubah teks menjadi huruf kecil (lowercase) dari dokumen yang awalan berupa (uppercase) huruf besar [14].
   Contoh:
   Sebelum: Ribet poll
   Sesudah: ribet poll
b. *Tokenize*
   Proses tokenisasi adalah proses memecahkan kalimat menjadi potongan kata atau token untuk mengetahui asal munculnya kata [9].
   Contoh:
   Sebelum: ribet poll
   Sesudah: ['ribet','poll']
c. *Stem*
   Prosedur ini kemudian dilanjutkan dengan menghapus afiks dari setiap kata hingga menjadi kata dasar, dengan tujuan untuk membersihkan kata yang salah ejaan [15].
   Contoh:
   Sebelum: ['jann','ribetttttttt']
   Sesudah: ['jann','ribet']
d. *Filter Stopwords*
   Adalah kosa-kata yang tidak termasuk kata unik ataupun ciri khas terhadap suatu dokumen/data atau tidak memiliki pesan apapun secara signifikan pada teks ataupun kalimat [16]. Dalam proses ini dilakukan penghapusan kata-kata yang tidak

memiliki makna signifikan dalam analisis sentimen seperti “akan”, “kan”, “yang”, “untuk”, ”saya”. Proses ini juga mencakup penghapusan kata-kata yang tidak ditemukan di dalam Kamus Besar Bahasa Indonesia.
Contoh:
Sebelum: jann ribet
Sesudah: ribet

e. Filter Token By Length: Menyaring kata-kata atau token dengan panjang tertentu (3-25 karakter).
Contoh:
Sebelum: bagus pertalite jangan di hilang
Sesudah: bagus pertalite jangan hilang

Setelah semua langkah selesai, data diubah menjadi vektor numerik menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)*. Agar data yang digunakan dapat diolah oleh mesin dan diklasifikasikan dengan algoritma, maka dilakukan pembobotan dengan TF-IDF [17]. Proses ini memberikan bobot pada kata-kata berdasarkan frekuensi kemunculannya dalam dokumen dan dataset, menghasilkan representasi teks numerik yang memungkinkan algoritma pembelajaran mesin untuk melakukan analisis sentimen secara efektif.



Gambar 9. Proses Pembobotan *TDF-IDF*

Tabel 6. Contoh Hasil *Word List*

Word	*Total Occurences*	*Document Occurences*	Negatif	Positif
mobil	320	210	254	66
subsidi	279	204	246	33
rakyat	230	169	221	9
aturan	193	163	184	9
pertalite	129	109	102	27
pemerintah	107	89	100	7
pertamax	88	78	70	18
ribet	65	60	61	4
susah	65	57	65	0

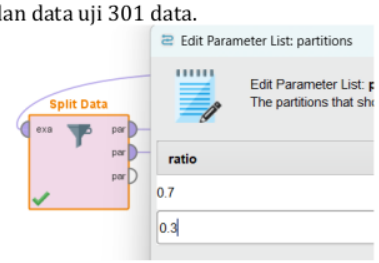
Tabel ini menunjukkan hasil kata-kata yang diekstrak dari komentar menggunakan operator Process Documents from Data. Analisis ini memberikan informasi mengenai kata-kata yang terdapat dalam data komentar.

a. *Word* menunjukkan kata yang dianalisis
b. *Total occurences* menunjukkan jumlah total kata dalam seluruh data
c. *Document occurences* mengindikasi jumlah komentar yang mengandung kata tersebut
d. Kolom negatif dan positif memperlihatkan distribusi kemunculan kata dalam masing-masing kategori sentimen.

4.1.4 *Data Mining*

Pada tahap Data Mining, data yang sudah melalui proses *Transformation* akan digunakan untuk membangun model klasifikasi sentimen. Tahapan ini bertujuan untuk mengidentifikasi pola dalam data komentar apakah positif atau negatif. Dalam penelitian ini, langkah-langkah yang dilakukan pada Data Mining sebagai berikut:

a. Pembagian Dataset
Dataset dibagi menjadi dua bagian yaitu data latih dan data uji dengan rasio 0.7 dan 0.3 menggunakan operator *split data* serta parameter *sampling type* menggunakan *linear sampling* dihasilkan data latih berjumlah 703 data dan data uji 301 data.



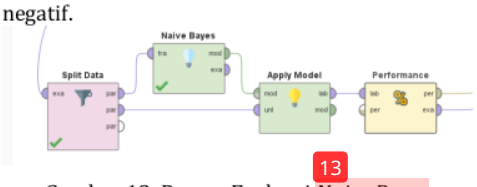
Gambar 10. Operator *Split Data*

b. Penerapan Algoritma *Naive Bayes*
Setelah data dibagi, model klasifikasi diterapkan menggunakan algoritma Naïve Bayes dengan operator. Algoritma Naïve Bayes mengkategorikan setiap komentar berdasarkan kemungkinan bahwa setiap baris komentar akan termasuk ke dalam kategori positif atau negatif

Gambar 11. Algoritma *Naïve Bayes*

4.1.5 *Evaluation*

Setelah model Naïve Bayes dilatih dan diuji, evaluasi perfoma model dilakukan dengan menggunakan operator *Apply Model* dan *Perfomance* pada AI Studio 2024.1.0. Operator Apply Model digunakan untuk menerapkan model yang telah dilatih pada data uji yang telah melalui proses *preprocessing* dan *transformation*. Prediksi sentimen untuk data uji dihasilkan oleh operator ini. Selanjutnya operator *Perfomance* digunakan untuk mengevaluasi kemampuan algoritma *Naïve Bayes* untuk secara akurat mengkategorikan sentimen komentar menjadi kategori positif atau negatif.



Gambar 12. Proses Evaluasi *Naïve Bayes*

Hasil evaluasi menunjukkan performa algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen komentar sebagai berikut:

Gambar 13. Hasil Evaluasi *Naïve Bayes*

Berdasarkan *confusion matrix* pada Gambar 4. 13 Hasil Evaluasi Model Sentimen Naïve Bayes, kita dapat menghitung metrik-metrik lain untuk menilai kinerja model lebih lanjut:

a. *Accuracy*
   Secara keseluruhan, *accuracy* model adalah 79,40%, yang menunjukkan bahwa model berhasil mengklasifikasikan 79,40% dari semua komentar yang diuji, baik positif maupun negatif.
b. *Precision*
   Untuk kelas negatif adalah 90,99%, yang menunjukkan bahwa 90,99% dari semua prediksi negatif benar-benar merupakan komentar negatif, namun untuk kelas positif adalah 39,71% yang menunjukkan model mengalami kesulitan untuk menemukan komentar positif dengan benar.
c. *Recall*
   Untuk kelas negatif adalah 83,79% yang menunjukkan bahwa model berhasil menemukan 83,79% dari komentar negatif sebenarnya, untuk kelas positif 56,25% yang menunjukkan bahwa model berhasil menemukan 56,25% dari komentar positif yang sebenarnya.

Secara keseluruhan, *accuracy* model Naïve Bayes adalah 79,40% yang menunjukkan bahwa model berhasil mengklasifikasikan 79,40% dari semua komentar yang diuji, baik positif atau negatif. Namun hasil evaluasi menunjukkan bahwa model lebih baik dalam mengidentifikasi komentar negatif, dengan tingkat keakuratan *precision* 90,99% dan *recall* 83,79%. Sebaliknya, model menunjukkan kesulitan dalam mengidentifikasi komentar positif, dengan *precision* 39,71% dan *recall* 56,25%. Salah satu faktor yang mempengaruhi hasil ini adalah ketidakseimbangan jumlah data, di mana data dengan kategori negatif berjumlah 869 jauh lebih banyak dibandingkan data dengan kategori positif yang berjumlah 135 data. Ketidakseimbangan data ini menyebabkan model lebih berfokus dalam mengklasifikasikan komentar negatif dengan baik, sementara perfomanya dalam mengidentifikasi komentar positif cenderung lebih rendah. Meskipun demikian, Naïve Bayes masih dapat mengklasifikasikan komentar negatif dengan cukup baik. Penelitian ini hanya mengevaluasi kinerja algoritma Naïve Bayes tanpa melakukan optimasi lebih lanjut pada model, sehingga masih ada ruang untuk perbaikan, teruama dalam meningkatkan prediksi terhadap komentar positif.

4.1.6 *Reporting*

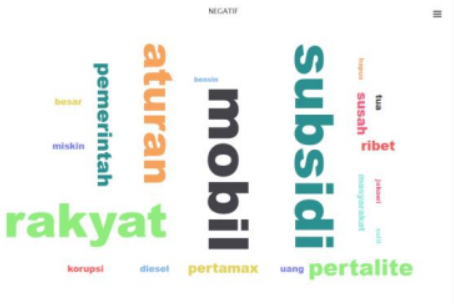
Hasil analisis sentimen ditampilkan untuk menunjukkan gambaran mengenai distribusi sentimen dan kata-kata yang sering muncul dalam kategori positif dan negatif. Pendistribusian dilakukan dalam format tabel berikut:

Tabel 7. 10 Kata Teratas Kategori Negatif

Word	*Total Occurencec*	*Document occurences*	Negatif	Positif
mobil	mobil	320	210	254
subsidi	subsidi	279	204	246
rakyat	rakyat	230	169	221
aturan	aturan	193	163	184
pertalite	pertalite	129	109	102
pemerintah	pemerintah	107	89	100
pertamax	pertamax	88	78	70
susah	susah	65	57	65
ribet	ribet	65	60	61
masyarakat	masyarakat	49	43	47

Tabel ini menunjukkan kata-kata yang sering muncul pada kategori negatif, diurutkan berdasarkan jumlah atribut negatif tertinggi. Tujuan penyajian ini untuk memberi gambaran tentang kata-kata yang paling berkontribusi

terhadap sentimen negatif di antara komentar yang dianalisis.

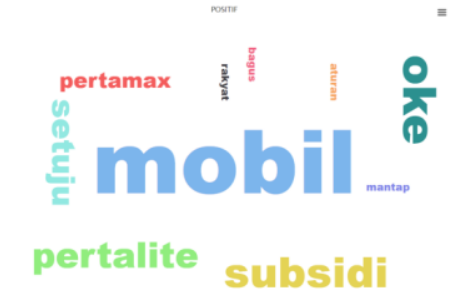


Gambar 14. *Word Cloud* Kata Kategori Negatif

Tabel 8. 10 Kata Teratas Kategori Positif

Word	*Total Occurencec*	*Document occurences*	Negatif	Positif
mobil	mobil	320	210	254
oke	oke	43	26	9
subsidi	subsidi	279	204	246
pertalite	pertalite	129	109	102
setuju	setuju	31	31	6
pertamax	pertamax	88	78	70
aturan	aturan	193	163	184
mantap	mantap	16	16	7
rakyat	rakyat	230	169	221
bagus	bagus	16	16	8

Tabel ini menunjukkan kata-kata yang sering muncul pada kategori positif, diurutkan berdasarkan jumlah atribut positif tertinggi. Tujuan penyajian ini untuk memberi gambaran tentang kata-kata yang paling berkontribusi terhadap sentimen positif di antara komentar yang dianalisis.



Gambar 15. *Word Cloud* Kata Kategori Positif

4.2 PEMBAHASAN

Analisis ini menunjukkan bahwa sentimen negatif lebih dominan dibandingkan sentimen positif. Namun, adanya apresiasi di kelompok positif menandakan bahwa kebijakan tersebut memiliki potensi untuk diterima lebih banyak orang jika ditangani dengan baik.

Selain itu, analisis yang dilakukan menggunakan algoritma Naïve Bayes menunjukkan hasil *accuracy* 79,40% dengan tingkat *precision* yang lebih tinggi pada kategori negatif daripada kategori positif.

Secara keseluruhan, hasil ini menunjukkan betapa pentingnya meningkatkan transparansi, penyederhanaan akses subsidi, dan komunikasi yang lebih baik untuk mengurangi persepsi negatif dan meningkatkan penerimaan masyarakat terhadap kebijakan tersebut.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan subsidi BBM berdasarkan komentar pada video *YouTube* "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - *SIP* 02/09", menggunakan algoritma *Naïve Bayes*. Hasil penelitian ini menunjukkan bahwa:

1. Sentimen negatif lebih dominan, dengan keluhan tentang akses yang sulit, susah dan ribet terhadap penerapan kebijakan *QR Code* BBM bersubsidi dan Sentimen positif muncul seperti oke, setuju, mantap dan bagus meski tidak mendominasi, dengan dukungan terhadap penerapan kebijakan BBM.
2. Dengan *accuracy* 79,40%, algoritma *Naïve Bayes* menunjukkan mengklasifikasikan komentar, meskipun hasilnya lebih baik pada sentimen negatif.

Dalam kesimpulan ini, transparansi kebijakan, penyederhanaan prosedur, dan komunikasi yang lebih efektif untuk mengurangi sentimen negatif dan meningkatkan dukungan terhadap kebijakan subsidi BBM.

Hasil penelitian ini menunjukkan beberapa hal yang dapat dijadikan masukan untuk pengembangan penelitian di masa depan serta upaya untuk meningkatkan kualitas analisis sentimen. Berikut adalah saran yang dapat diajukan:

1. Pengembangan penelitian selanjutnya
   Penelitian ini menggunakan data komentar dari video YouTube "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - SIP 02/09" sebagai sumber data. Untuk hasil yang lebih representatif, penelitian berikutnya menggunakan lebih banyak sumber dari platform media sosial lainnya seperti Twitter atau sekarang yang telah berubah menjadi X.
2. Peningkatan model analisis sentimen

Model Naïve Bayes dalam penelitian ini menunjukkan performa dengan *accuracy* 79,40% masih perlu untuk dilakukan peningkatan. Penelitian mendatang disarankan untuk mengeksplorasi algoritma lain seperti Random Forest, SVM atau penggunaan penyeimbang data untuk mengevaluasi *accuracy*, *precision* dan *recall*.

3. Integrasi Analisis Sentimen dengan Kebijakan sentimen

   Penelitian ini dapat dijadikan masukan untuk pihak terkait dalam menyusun atau menerapkan kebijakan subsidi BBM yang lebih baik. Penelitian mendatang dapat berfokus pada integrasi analisis sentimen dengan data wawancara atau survei untuk memberikan konteks lebih mendalam tentang pendapat masyarakat.

4. Peningkatan Strategi Komunikasi Pemerintah

   Berdasarkan temuan bahwa sentimen negatif lebih dominan, pihak terkait disarankan untuk meningkatkan komunikasi kepada masyarakat terhadap kebijakan. Penelitian mendatang dapat berfokus pada analisis sentimen terhadap strategi komunikasi pemerintah untuk membangun pandangan masyarakat terhadap kebijakan dengan baik.

6. UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih banyak kepada pihak-pihak yang telah mendukung serta membantu dalam penelitian yang telah saya lakukan ini, pertama saya ucapkan kepada Tuhan yang Maha Esa, kepada dosen pembimbing, keluarga dan teman-teman atas keterlibatannya dalam memberikan dukungan untuk penyusunan penelitian ini. Dan saya juga berterimakasih kepada pihak jurnal yang telah memberikan saya kesempatan untuk mempublikasikan jurnal yang telah saya buat.

1413 PENINGKATAN MODEL ANALISIS SENTIMEN MELALUI ALGORITMA NAIVE BAYES BERDASARKAN DATA KOMENTAR YOUTUBE

ORIGINALITY REPORT

20%
SIMILARITY INDEX

PRIMARY SOURCES

1	ejournal.itn.ac.id Internet	141 words — 4%
2	e-journal.stmiklombok.ac.id Internet	91 words — 2%
3	repository.uin-suska.ac.id Internet	35 words — 1%
4	jurnal.umj.ac.id Internet	29 words — 1%
5	djournals.com Internet	27 words — 1%
6	docplayer.info Internet	26 words — 1%
7	ejournal.upbatam.ac.id Internet	24 words — 1%
8	journal.untar.ac.id Internet	21 words — 1%
9	repository.lppm.unila.ac.id Internet	20 words — < 1%

10	www.ejurnal.stmik-budidarma.ac.id Internet	20 words — < 1%
11	Rafi Rahmadani, Abdul Rahim, Rudiman Rudiman. "ANALISIS SENTIMEN ULASAN “OJOL THE GAME” DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES DAN MODEL EKSTRAKSI FITUR TF-IDF UNTUK MENINGKATKAN KUALITAS GAME", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Crossref	19 words — < 1%
12	Reza Zulfiqri, Betha Nurina Sari, Tesa Nur Padilah. "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MEDIA SOSIAL INSTAGRAM PADA SITUS GOOGLE PLAY STORE MENGGUNAKAN NAÏVE BAYES CLASSIFIER", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Crossref	19 words — < 1%
13	ojs.unikom.ac.id Internet	16 words — < 1%
14	Ristyani Slamet, Windu Gata, Annisa Novtariany, Khairunisa Hilyati, Febri Ainun Jariyah. "Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal", INTECOMS: Journal of Information Technology and Computer Science, 2022 Crossref	15 words — < 1%
15	ojs.unud.ac.id Internet	15 words — < 1%
16	Tiara Danirmala, Yusuf Sulistyo Nugroho. "Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter", The Indonesian Journal of Computer Science, 2023	14 words — < 1%

Crossref

17 Bayu Ferdiansyah, Leonard Goeirmanto. "Prediksi Loyalitas dalam Keterikatan Karyawan terhadap Perusahaan Menggunakan Algoritma C4.5* (Studi Kasus PT.XYZ)", Jurnal Sistem dan Teknologi Informasi (JUSTIN), 2020 — 13 words — < 1%
Crossref

18 Muhammad Galih Ramaputra, Hendri Purnomo. "Analisis Sentimen Opini Masyarakat Terhadap Penggunaan ChatGPT di Bidang Pendidikan Berbasis Twitter", Jurnal Pepadun, 2024 — 13 words — < 1%
Crossref

19 Nur Aisyah Wahyuni, Dinda Putri Ayu, Hafidz Irsyad. "Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine", Arcitech: Journal of Computer Science and Artificial Intelligence, 2024 — 13 words — < 1%
Crossref

20 ejournal.akprind.ac.id — 13 words — < 1%
Internet

21 www.esdm.go.id — 13 words — < 1%
Internet

22 www.researchgate.net — 13 words — < 1%
Internet

23 publikasiilmiah.unwahas.ac.id — 12 words — < 1%
Internet

24 Andra Setiawan, Ryan Randy Suryono. "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes", Edumatic: Jurnal Pendidikan Informatika, 2024 — 11 words — < 1%
Crossref

25	Elly Indrayuni, Achmad Nurhadi. "Sentiment Analysis About COVID-19 Booster Vaccine on Twitter Using Deep Learning", SinkrOn, 2022 Crossref	11 words — < 1%
26	ojs.uho.ac.id Internet	11 words — < 1%
27	ejournal-binainsani.ac.id Internet	10 words — < 1%
28	ejournal.nusamandiri.ac.id Internet	10 words — < 1%
29	file.techscience.com Internet	10 words — < 1%
30	ojs.stmik-banjarbaru.ac.id Internet	10 words — < 1%
31	repositori.stikes-ppni.ac.id Internet	10 words — < 1%
32	www.beritane.com Internet	10 words — < 1%
33	Linda Jayanti, Steven R. Sentinuwo, Oktavian A. Lantang, Agustinus Jacobus. "Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking Menggunakan Data Mining", Jurnal Teknik Informatika, 2016 Crossref	9 words — < 1%
34	www.ilmudata.org Internet	9 words — < 1%

35	Minardi Minardi, Ranita Lasepa, Santoso Riyadi, Syahrur Ramadhan, Dedi Dwi Saputra. "Sentiment Analysis Terhadap Perspektif Warganet Atas Tragedi Kanjuruhan Malang di Twitter Menggunakan Naïve Bayes Classifier", Jurnal Informatika, 2023 Crossref	8 words — < 1%
36	Rahma Sarita, Dedy Iswanto, Handoyo Wirastomo. "PERANAN BADAN USAHA MILIK DESA (BUMDES) TERHADAP PENGEMBANGAN EKONOMI DESA MANEMENG KECAMATAN BRANG ENE KABUPATEN SUMBAWA BARAT", JOURNAL of APPLIED BUSINESS and BANKING (JABB), 2023 Crossref	8 words — < 1%
37	dspace.uii.ac.id Internet	8 words — < 1%
38	ejurnal.seminar-id.com Internet	8 words — < 1%
39	ojs.upi-yai.ac.id Internet	8 words — < 1%
40	www.hindawi.com Internet	8 words — < 1%
41	www.nafiriz.com Internet	8 words — < 1%
42	www.slideshare.net Internet	8 words — < 1%
43	Teresia Ardika Dewi, Evangs Mailoa. "PERBANDINGAN IMPLEMENTASI METODE SMOTE PADA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM ANALISIS SENTIMEN OPINI MASYARAKAT TENTANG MIXUE",	6 words — < 1%

Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023

Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF