



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

Deden Syarif Hidayat¹, Odi Nurdiawan², Fadhil M.Basysyar³, Muhamad Sulaeman⁴

¹Program Studi Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, ^{2,3,4}STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

¹dedensyarifhidayat@gmail.com, ²odinurdiawan2020@gmail.com, ³fadhil.m.basysyar@gmail.com,

⁴sulaeman.ikmi@gmail.com

Abstract

The implementation of the QR Code-based Fuel Oil (BBM) 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 faced criticism, especially regarding the complexity of the QR Code registration process and the limitations on the criteria for eligible vehicles. This study aims to analyze public sentiment towards the policy using comment data from a YouTube video entitled "Simak! New Rules for Criteria for Using Subsidized BBM - SIP 02/09" uploaded to the Seputar iNews RCTI YouTube channel. The method used is sentiment analysis based on the Naive Bayes algorithm. The data preprocessing process includes text cleaning, tokenization, removing stopwords, and stemming to ensure that the analyzed data is clean and structured. The dataset is divided into training data (70%) and test data (30%) to build and evaluate the model. The model shows an accuracy of 79.40%, with better performance in recognizing negative sentiment than positive. The results show that the majority of comments have negative sentiment, reflecting public dissatisfaction with this policy. This study highlights the importance of more effective communication strategies from the government to improve public understanding and acceptance of implemented policies. In addition, the results of this study also open up opportunities for further development in the use of computational sentiment analysis to support decision making 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 kendaraan roda empat, yang dimulai pada 1 Oktober 2024, telah memunculkan berbagai tanggapan dari masyarakat. Kebijakan ini bertujuan untuk memastikan distribusi BBM bersubsidi lebih tepat sasaran, namun menghadapi kritik terutama terkait kompleksitas proses pendaftaran *QR Code* dan pembatasan kriteria kendaraan yang memenuhi syarat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan tersebut dengan menggunakan data komentar dari video *YouTube* berjudul "Simak! Aturan Baru Kriteria 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 mencakup pembersihan teks, *tokenisasi*, penghilangan *stopwords*, dan *stemming* untuk memastikan data yang dianalisis bersih dan terstruktur. Dataset dibagi menjadi data pelatihan (70%) dan data uji (30%) untuk membangun serta mengevaluasi model. Model menunjukkan akurasi sebesar 79,40%, dengan performa yang lebih baik dalam mengenali sentimen negatif dibandingkan positif. Hasil penelitian menunjukkan bahwa mayoritas komentar memiliki sentimen negatif, mencerminkan ketidakpuasan masyarakat terhadap kebijakan ini. Penelitian ini menyoroti pentingnya strategi komunikasi yang lebih efektif dari pemerintah untuk meningkatkan pemahaman dan penerimaan masyarakat terhadap kebijakan yang diimplementasikan. Selain itu, hasil penelitian ini juga membuka



peluang untuk pengembangan lebih lanjut dalam pemanfaatan analisis sentimen berbasis komputasi untuk mendukung pengambilan keputusan dalam studi kebijakan publik.

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

1. PENDAHULUAN

Kebutuhan energi yang terus meningkat di Indonesia telah mendorong 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. Beberapa pihak mendukung kebijakan ini sebagai langkah strategis untuk mengatasi ketidakefisienan subsidi BBM, sementara kritik muncul terutama terkait dengan proses pendaftaran yang dianggap rumit dan pembatasan kapasitas kendaraan. Kritik ini mencerminkan tantangan dalam menyosialisasikan kebijakan baru dan memastikan aksesibilitas yang adil.

Dengan semakin berkembangnya teknologi, analisis sentimen masyarakat melalui platform media sosial telah menjadi alat penting untuk memahami opini kebijakan publik [1]. Penelitian ini memanfaatkan analisis sentimen untuk mengevaluasi respons masyarakat terhadap kebijakan BBM bersubsidi. Data komentar dari video *YouTube* berjudul "Simak! Aturan Baru Kriteria Penggunaan BBM Bersubsidi - SIP 02/09" digunakan untuk mengidentifikasi pola sentimen positif dan negatif. Teknik analisis sentimen berbasis algoritma *Naive Bayes* dipilih karena keandalannya dalam mengklasifikasikan data teks [2].

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 Bayes*. Penelitian oleh Ardiansyah dan Nur'aini [3] menganalisis sentimen masyarakat terhadap kenaikan harga BBM di Twitter, dengan dataset sebanyak 2312 data, menghasilkan akurasi 92%. Penelitian oleh Basedt et al. [4] membandingkan *Naive Bayes* dengan algoritma lain dalam menganalisis sentimen terhadap kenaikan harga BBM, dan *Naive Bayes* menunjukkan akurasi tertinggi sebesar 80,28%. Rahmadani et al. [5] menggunakan *Naive Bayes* dan *TF-IDF* untuk menganalisis sentimen ulasan game di Google Play Store, dengan akurasi 94,12%, menunjukkan efektivitas kombinasi kedua teknik tersebut.

Di sisi lain, Muliawan dan Dazki [6] menganalisis sentimen terhadap kebijakan pemindahan ibu kota di Twitter dengan akurasi 65,26%, lebih tinggi dibandingkan *KNN* dan *Random Forest*. Penelitian ini menyoroti distribusi sentimen, dengan 36,8% positif, 25% netral, dan 38,1% negatif. Meskipun penelitian ini memberikan wawasan penting, sebagian besar berfokus pada Twitter atau ulasan aplikasi, sementara analisis sentimen berbasis komentar *YouTube* masih kurang dieksplorasi.

Penelitian terbaru oleh Widia et al. (2024) [7] berjudul "Optimasi Algoritma *Naive Bayes* Untuk Menganalisis Sentimen Pada Konten Pemindahan Ibu Kota di *YouTube*" mengklasifikasikan sentimen komentar *YouTube* dengan akurasi 93,71%. Temuan ini relevan dengan penelitian ini yang juga menggunakan komentar *YouTube* untuk menganalisis sentimen



terhadap kebijakan BBM bersubsidi berbasis *QR Code*. Penelitian ini memberikan kontribusi baru dengan menggali kata-kata dominan dalam sentimen positif dan negatif, serta memperluas pemahaman penggunaan *YouTube* dalam analisis sentimen kebijakan publik.

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 [8].

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 [9].

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 ada, nilai *TF-IDF* diperoleh melalui perkalian nilai *TF (Term Frequency)* dengan nilai *IDF (Inverse Document Frequency)* [10].

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 [11].

3. METODOLOGI PENELITIAN

3.1. PEMILIHAN TF-IDF

Metode *TF-IDF* dipilih karena kemampuannya dalam memberikan bobot pada kata-kata yang relevan, sehingga meningkatkan kualitas representasi data teks untuk analisis [5].

3.2. TAHAPAN PENELITIAN

Penelitian ini terdiri dari beberapa tahapan utama berdasarkan pendekatan KDD, yang meliputi *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation* [12]. 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 [13].
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 [14]. 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
<i>Data Selection</i>	Pengumpulan data	Mengambil data komentar <i>YouTube</i> dari video terkait.
<i>Preprocessing</i>	Pelabelan Data	Melabeli komentar sebagai positif, netral, atau negatif.
	Penyaringan Data	Menghapus komentar netral, fokus pada positif dan negatif
	Pembersihan	Menghapus elemen



	Data	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
<i>Transformation</i>	Representasi Data	Mengubah teks komentar menjadi vektor numerik dengan TF-IDF.
	Pemrosesan Teks	Melakukan case folding, tokenisasi, stemming, dan filter stopwords.
<i>Data Mining</i>	Klasifikasi sentimen	Menerapkan algoritma <i>Naive Bayes</i> untuk klasifikasi positif dan negatif.
Data Mining	Pembagian Dataset	Membagi data menjadi 70% untuk latih dan 30% untuk uji.
	Penerapan Algoritma	Menggunakan <i>Naive Bayes</i> untuk klasifikasi sentimen..
Evaluation	Pengukuran Performa	Mengevaluasi kinerja model dengan <i>accuracy, precision, dan recall</i> .
<i>Reporting</i>	Penyajian Hasil	Menampilkan kata-kata yang sering muncul dalam kategori positif dan negatif dalam tabel frekuensi dan <i>word cloud</i> .
	Analisis dan Rekomendasi	Menyimpulkan hasil untuk memberikan wawasan kepada pembuat kebijakan.
<i>Reporting</i>	Penyajian Hasil	Menampilkan kata-kata yang sering muncul dalam kategori positif dan negatif dalam tabel frekuensi dan <i>word cloud</i> .

3.3. PENGUMPULAN DATA

Sumber data untuk penelitian ini yaitu data sekunder. Data sekunder ialah sumber data yang

tidak diperoleh langsung oleh peneliti [15]. 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.

3.4. KETERBATASAN DATASET

Dataset yang digunakan dalam penelitian ini memiliki ketidakseimbangan jumlah komentar antara sentimen positif dan negatif. Setelah dilakukan pelabelan, ditemukan bahwa hanya 135 komentar yang memiliki sentimen positif, sementara 869 komentar lainnya memiliki sentimen negatif. Ketidakseimbangan ini dapat memengaruhi performa model dalam mengklasifikasikan sentimen positif.

3.5. 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.

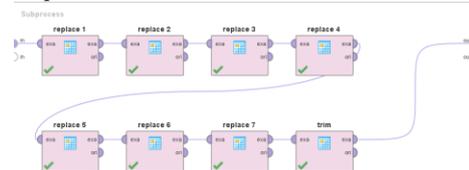
4. HASIL DAN PEMBAHASAN

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



1	Mencekik rakyat	negatif
2	Jannn ribettttttt	negatif
3	Ruwet semua aturran di rubah	negatif
4	Ribet poll..., 😞 😞 😞	negatif

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 "
\"".
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.

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,

Tabel 5. Contoh Hasil Setelah Pembersihan

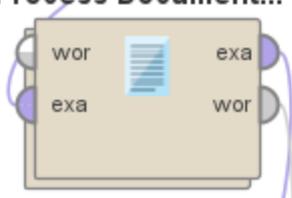
No	Sebelum	Sesudah
1	Setuju	Setuju
2	Lanjutkn	Lanjutkn
3	Bagusss. Yg penting penting pertalite jangan di hilang kan...	Bagusss Yg penting pertalite jangan di hilang kan
4	Mencekik rakyat	Mencekik rakyat
5	Jannn ribettttttt	Jannn ribettttttt
6	Ruwet semua aturran di rubah	Ruwet semua aturran di rubah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

poll..., 😊 😊 😊

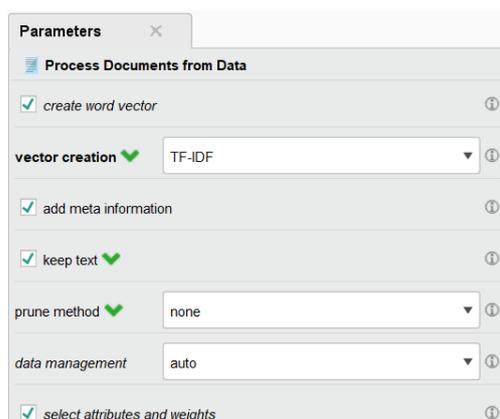
4.1.3 TRANSFORMATION

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:

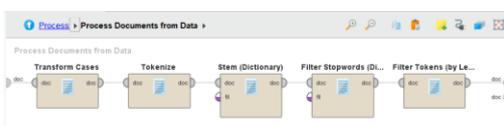
Process Document...



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 [16].

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 [11].

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 [17].

Contoh:

Sebelum: ['jann','ribetttttt']

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 [18]. 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* [19]. 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.

File No.	label	abal	absurd	anom	atf	adu	aluh	abuh	adi	abat
928	negatif	0	0	0	0	0	0	0	0	0
927	negatif	0	0	0	0.053	0	0	0	0	0
928	positif	0	0	0	0	0	0	0	0	0
929	negatif	0	0	0	0.249	0	0	0	0	0
940	negatif	0	0	0	0.439	0	0	0	0	0
941	positif	0	0	0	0	0	0	0	0	0
942	positif	0	0	0	0	0	0	0	0	0
943	negatif	0	0	0	0	0	0.418	0	0	0
944	negatif	0	0	0	0	0	0	0	0	0
945	negatif	0	0	0	0	0	0	0	0	0
946	negatif	0	0	0	0	0	0	0	0	0
947	negatif	0	0	0	0.488	0	0	0	0	0
948	positif	0	0	0	0	0	0	0	0	0
949	negatif	0	0	0	0	0	0	0	0	0
950	positif	0	0	0	0	0	0	0	0	0
951	negatif	0	0	0	0.377	0	0	0	0	0

Gambar 9. Proses Pembobotan TDF-IDF

Tabel 6. Contoh Hasil Word List

Word	Total Occurrences	Document Occurrences	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.

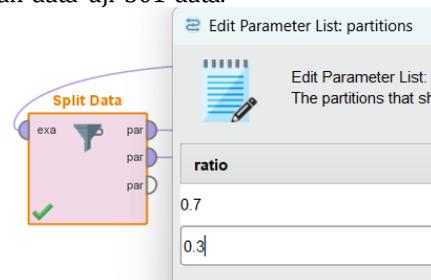
- Word menunjukkan kata yang dianalisis
- Total occurrences menunjukkan jumlah total kata dalam seluruh data
- Document occurrences mengindikasikan jumlah komentar yang mengandung kata tersebut
- Kolom negatif dan positif memperlihatkan distribusi kemunculan kata dalam masing-masing kategori sentimen.

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:

- 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

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

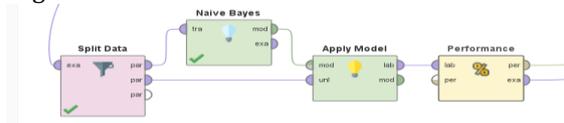


Gambar 11. Algoritma Naive Bayes

4.1.5 Evaluation

Setelah model Naive Bayes dilatih dan diuji, evaluasi performa model dilakukan dengan menggunakan operator Apply Model dan Performance 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 Performance digunakan untuk mengevaluasi kemampuan algoritma Naive Bayes untuk secara akurat mengkategorikan

sentimen komentar menjadi kategori positif atau negatif.



Gambar 12. Proses Evaluasi Naive Bayes

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

Class	Actual	Predicted	Count
negatif	negatif	negatif	212
negatif	positif	negatif	41
positif	negatif	positif	83
positif	positif	positif	227

Gambar 13. Hasil Evaluasi Naive Bayes

Berdasarkan confusion matrix pada Gambar 4.13 Hasil Evaluasi Model Sentimen Naive 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.

Model Naive Bayes menunjukkan accuracy 79,40%, dengan kinerja terbaik dalam mengidentifikasi komentar negatif (precision 90,99%, recall 83,79%). Namun, model kesulitan mengidentifikasi komentar positif (precision 39,71%, recall 56,25%), disebabkan oleh ketidakseimbangan data (869 data negatif dan

135 data positif). Meskipun efektif dalam mengklasifikasikan komentar negatif, kinerja untuk komentar positif masih perlu perbaikan. Penelitian ini belum mengoptimasi model, sehingga masih ada potensi untuk meningkatkan prediksi 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 Occurrence	Document occurrences	Negatif	Positif
mobil	320	210	210	254
subsidi	279	204	204	246
rakyat	230	169	169	221
aturan	193	163	163	184
pertalite	129	109	109	102
pemerintah	107	89	89	100
pertamax	88	78	78	70
susah	65	57	57	65
ribet	65	60	60	61
masyarakat	49	43	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 Occurrence	Document occurrences	Negatif	Positif
mobil	320	210	210	254
oke	43	26	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 sentimen yang dilakukan menunjukkan bahwa sentimen negatif lebih dominan dibandingkan dengan sentimen positif dalam komentar masyarakat terkait kebijakan ini. Sentimen negatif ini terutama berfokus pada kesulitan dalam proses pendaftaran, seperti masalah teknis dalam pengunggahan foto dan aksesibilitas aplikasi yang mempersulit pengguna. Komentar-komentar seperti "Sudah coba daftar..susah..upload foto ga bisa2..sengaja dibuat susah apa gmn?" menunjukkan betapa frustrasinya masyarakat dengan proses yang rumit ini.

Namun, meskipun sentimen negatif lebih dominan, sentimen positif juga hadir meskipun dalam jumlah yang lebih kecil. Beberapa komentar yang menyatakan dukungan terhadap kebijakan tersebut, seperti "Oke, setuju, mantap" menunjukkan bahwa meskipun ada beberapa kendala, sebagian masyarakat tetap melihat potensi positif dari penerapan kebijakan subsidi BBM ini.

Hasil analisis yang dilakukan menggunakan algoritma Naïve Bayes menunjukkan accuracy 79,40%, dengan tingkat precision yang lebih tinggi pada kategori negatif dibandingkan kategori positif. Hal ini mencerminkan bahwa sistem lebih mampu mengidentifikasi komentar negatif dengan lebih tepat, meskipun sentimen positif tetap ada, meskipun lebih sedikit.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa peningkatan transparansi, penyederhanaan prosedur pendaftaran, dan komunikasi yang lebih efektif sangat penting untuk mengurangi persepsi negatif yang ada. Terlebih lagi, jika aspek-aspek teknis dalam pendaftaran dapat diperbaiki dan disederhanakan, maka masyarakat akan lebih mudah menerima kebijakan ini. Oleh karena itu, disarankan agar pihak terkait memperbaiki sistem pendaftaran agar lebih mudah diakses dan dipahami oleh masyarakat, serta meningkatkan komunikasi untuk memberikan penjelasan yang lebih jelas mengenai proses 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, penelitian ini menekankan pentingnya transparansi kebijakan, penyederhanaan proses pendaftaran, 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 *accuracy* 79,40%, namun perlu peningkatan. Penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain seperti *Random Forest* dan *SVM*, serta mempertimbangkan penyeimbangan data guna meningkatkan *accuracy*, *precision*, dan *recall*, terutama dalam mengatasi ketidakseimbangan data sentimen positif dan negatif.
3. Integrasi Analisis Sentimen dengan Kebijakan sentimen
Penelitian ini memberikan wawasan penting yang dapat digunakan oleh pihak terkait dalam merumuskan atau menerapkan kebijakan subsidi BBM yang lebih baik. Di masa depan, penelitian dapat berfokus pada integrasi analisis sentimen dengan data wawancara atau survei, yang memberikan konteks lebih mendalam tentang pendapat masyarakat. Ini akan membantu dalam merumuskan kebijakan yang lebih sesuai dengan harapan dan kebutuhan masyarakat.
4. Peningkatan Strategi Komunikasi Pemerintah
Berdasarkan temuan bahwa sentimen negatif lebih dominan, pihak pemerintah disarankan untuk meningkatkan komunikasi dengan masyarakat mengenai kebijakan subsidi BBM. Penyederhanaan proses pendaftaran atau prosedur akses

subsidi dapat menjadi langkah awal untuk meredakan ketidakpuasan masyarakat.

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.

DAFTAR PUSTAKA:

- [1] M. Taufik Sugandi, Martanto, and U. Hayati, "Analisis Sentimen Komentar Pengguna Youtube terhadap Kebijakan Baru Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 218–227, 2024.
- [2] R. Situmorang, U. M. Husni Tamyis, and L. S. Andar Muni, "Analisis Sentimen Destinasi Wisata Di Jawabarot Pada Twitter Menggunakan Algoritma Naive Bayes Classifier," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 339–342, 2023, doi: 10.51876/simtek.v8i2.287.
- [3] Ardiansyah and Nur'aini, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga Bbm Dengan Metode Naive Bayes," *J. Inf. Syst. Informatics Eng.*, vol. 8, no. 1, pp. 1–9, 2024, [Online]. Available: <https://doi.org/10.35145/joisie.v8i1.3838>
- [4] N. Basedt, E. Supriyadi, and A. S. Nugroho, "Perbandingan Algoritma Klasifikasi dalam Analisis Sentimen Opini Masyarakat tentang Kenaikan Harga Bbm," *Joined J. (Journal Informatics Educ.*, vol. 6, no. 2, p. 219, 2024, doi: 10.31331/joined.v6i2.2893.
- [5] R. Rahmadani, A. Rahim, and R. 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,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [6] J. Muliawan and E. Dazki, “Analisis Sentimen Pemindahan Ibu Kota Negara Indonesia Menggunakan Tiga Algoritma: Naïve Bayes, Knn, Dan Random Forest,” *J. Tek. Inform.*, vol. 4, no. 5, pp. 1227–1236, 2023, doi: 10.52436/1.jutif.2023.4.5.1436.
- [7] Widia, Z. Y. Aqsalia, S. Sari, N. U. Khoirunisa, and F. Kurniawan, “Optimasi Algoritma Naive Bayes Untuk Menganalisis Sentimen Pada Konten Pemindahan Ibu Kota di Youtube,” *J. Comput. Inf. Syst. Ampera*, vol. 5, no. 2, pp. 68–83, 2024.
- [8] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- [9] M. T. Nitamia and H. Februariyanti, “Analisis Sentimen Ulasan Ekpedisi J&T Ekspres Menggunakan Algoritma Naive Bayes,” *J. Manaj. Inform. Sist. Inf.*, vol. 5, no. 1, pp. 20–29, 2022.
- [10] F. Meila, A. Sofyan, N. Sulistiyowati, and A. Voutama, “ANALISIS SENTIMEN TERHADAP RESPON PERUBAHAN NAMA TWITTER MENJADI ‘ X ’ MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” vol. 8, no. 5, pp. 10987–10994, 2024.
- [11] A. Rifa’I, R. Ardhani, D. Pratama, and F. Fatihanursari, “Analisis Sentimen Terhadap Layanan Aplikasi Grab Indonesia Menggunakan Metode Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 303–309, 2024, doi: 10.36040/jati.v8i1.8425.
- [12] H. Hajaroh, T. Suprapti, and R. Narasati, “Implementasi Algoritma Naive Bayes Untuk Analisis Sentimen Ulasan Produk Makanan Dan Minuman Di Tokopedia,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 111–118, 2024, doi: 10.36040/jati.v8i1.8237.
- [13] R. Fansuri, E. Tohidi, E. Wahyudin, K. Kaslani, and I. Iin, “Analisis Pola Transaksi Pembelian Pada Bisnis Food and Beverage Menggunakan Algoritma Fp-Growth,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 203–208, 2024, doi: 10.36040/jati.v8i1.8293.
- [14] Wartumi, R. Kurniawan, and A. Y. Wijaya, “Analisis Data Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store dengan Klasifikasi Algoritma Naïve Bayes,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 164–170, 2024.
- [15] M. Melia, B. Irawan, and O. Nurdiawan, “Analisis Sentimen Terhadap Pengguna Gojek Dan Grab Pada Media Sosial Twitter Menggunakan Random Forest,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3614–3618, 2024, doi: 10.36040/jati.v7i5.7694.
- [16] M. K. Insan, U. Hayati, and O. Nurdiawan, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di,” *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [17] A. Erfina and N. Resti Wardani, “Analisis Sentimen Perguruan Tinggi Termewah Di Indonesia Menurut Ulasan Google Maps Menggunakan Support Vector Machine (Svm),” *J. Manaj. Inform. Sist. Inf.*, vol. 5, pp. 77–85, 2022.
- [18] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [19] K. A. Padhana and M. Sadikin, “Analisis Sentimen Masyarakat Terhadap Kondisi Perekonomian di Indonesia Pada Masa Pandemi 2020,” *J. Ilmu Tek dan Komput.*, vol. 5, no. 2, pp. 268–277, 2021.

REFERENSI:

- [1] “CNN Indonesia,” Agustus 2024. [Online]. Available: <https://www.cnnindonesia.com/ekonomi/20240827164219-85-1138155/bahlil-dpr-sepakat-patok-kuota-bbm-subsidi-1941-juta-kl-di-2025>.