



ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN SUBSIDI TEPAT MYPERTAMINA DI APLIKASI X MENGGUNAKAN ALGORITMA *NAÏVE* *BAYES*

Nasya Rahmadina¹, Abdi Muhaimin², Yulisman³

^{1,2,3}Universitas Hang Tuah Pekanbaru

^{1,2,3}Jl. Mustafa Sari No.5, Tangkerang Selatan, Kecamatan Bukit Raya
28281 Kota Pekanbaru Riau^{1,2,3}

E-mail : nasyadina26@gmail.com¹

ABSTRAK

Penerapan kebijakan pendaftaran subsidi tepat dan penggunaan QR Code untuk pembelian BBM bersubsidi di seluruh SPBU sejak tanggal 1 Oktober 2024 telah memicu berbagai respon dari masyarakat. Kebijakan ini bertujuan untuk memastikan subsidi BBM diterima oleh masyarakat yang berhak. Namun, banyak tantangan dan kendala dalam pelaksanaannya. Berdasarkan fenomena tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan subsidi tepat dengan mengambil data dari postingan pengguna di media sosial X. Penelitian ini menggunakan algoritma Naïve Bayes yang diawali dengan pengumpulan data melalui teknik crawling, kemudian tahap cosine similarity hingga pre-processing data untuk menghasilkan data yang siap dianalisis. Model diuji dengan perbandingan 80:20 terhadap 900 data dengan total data latih sebesar 720 data dan data uji 180 data. Hasil penelitian menunjukkan mayoritas sentimen masyarakat bersifat netral sebesar 44.2% disusul dengan sentimen negatif yang juga cukup signifikan sebesar 40.6% dan positif sebesar 15.2%. Dengan tingkat akurasi model sebesar 68%, penelitian ini memberikan hasil mendalam mengenai persepsi publik terhadap kebijakan subsidi tepat dan penggunaan aplikasi MyPertamina

Kata kunci : Analisis sentimen, Kebijakan Subsidi Tepat, MyPertamina, Naïve Bayes, QR Code BBM

ABSTRACT

The implementation of the targeted subsidy registration policy and the use of QR Codes for purchasing subsidized fuel at all gas stations on October 1, 2024, has sparked various pro and contra responses among the public. This policy aims to ensure that fuel subsidies are received by the rightful beneficiaries. However, its implementation has encountered numerous challenges and obstacles. Based on this phenomenon, this study aims to analyze public sentiment toward the targeted subsidy policy by collecting data from user posts on the social media platform X. This research employs the Naïve Bayes algorithm, beginning with data collection through crawling techniques, followed by the cosine similarity method to identify buzzer posts. Subsequently, data pre-processing is conducted to generate data ready for analysis. The model is tested using an 80:20 ratio for training and test data. The results indicate that the majority of public sentiment is neutral, followed by a significant proportion of negative sentiment. With a model accuracy of 68%, this study provides in-depth insights into public perceptions of the targeted subsidy policy and the use of the MyPertamina application.

Keywords: MyPertamina, Naïve Bayes, QR Code BBM, Sentiment Analysis, Targeted Fuel Subsidy Policy



1. PENDAHULUAN

Perkembangan teknologi dan internet telah membawa perubahan besar di berbagai bidang kehidupan masyarakat, termasuk pada bidang industri. Transformasi ini disebut sebagai revolusi industri 4.0, yakni pemanfaatan teknologi digital seperti *Artificial Intelligence* (AI), *Internet of Things* (IoT), dan sistem otomatisasi lainnya di bidang industri yang menyebabkan perubahan di berbagai sektor (Herman et al., 2023)(Purba et al., n.d.). Pertamina sebagai salah satu perusahaan sektor industri energi terbesar di Indonesia selalu berinovasi dengan memanfaatkan teknologi untuk meningkatkan efisiensi operasional layanan mereka kepada masyarakat. Salah satu implementasinya ialah aplikasi MyPertamina yang berguna untuk memudahkan proses transaksi konsumen dan mencatat pembelian BBM bersubsidi agar lebih tepat sasaran (G. Darmawan et al., 2023)(Alvita Wagiswari D et al., 2023).

Program subsidi tepat merupakan kebijakan pemerintah yang bertujuan memastikan penyaluran subsidi BBM secara tepat sasaran. Kebijakan ini diatur dalam Peraturan Presiden Nomor 191 Tahun 2014 tentang Penyediaan, Pendistribusian dan Harga Jual Eceran BBM. Jenis Bahan Bakar Minyak (BBM) yang termasuk bersubsidi adalah Biosolar dan Pertalite. Pemerintah menetapkan bahwasanya Bahan Bakar Minyak (BBM) subsidi memiliki jumlah yang terbatas sesuai kuota yang ditetapkan pemerintah dan hanya diperuntukkan bagi konsumen tertentu. Konsumen tersebut antara lain (Kurniawan & Ferdian, 2023):

1. Kendaraan pribadi roda empat, dengan batas pembelian 60 liter/hari.
2. Kendaraan umum angkutan orang atau barang roda empat, dengan batas pembelian 80 liter/hari.
3. Kendaraan angkutan barang (kecuali untuk pengangkut hasil pertambangan dan perkebunan dengan roda > 6), dengan batas pembelian 200 liter/hari
4. Mobil layanan umum seperti ambulance, mobil jenazah, truk sampah dan pemadam kebakaran.

Pembelian Bahan Bakar Minyak (BBM) bersubsidi menggunakan aplikasi MyPertamina dan *Qr Code* mulai diterapkan secara bertahap pada tanggal 1 Juli 2022 (I. Maulana et al., 2023), dan baru diberlakukan secara efektif di seluruh Stasiun Pengisian Bahan Bakar Umum (SPBU) di Indonesia pada tanggal 1 Oktober 2024. Melalui kebijakan ini, masyarakat pengguna BBM jenis pertalite dan solar diwajibkan untuk mendaftarkan kendaraannya di aplikasi MyPertamina (Wardhani et al., 2023). Dengan demikian aplikasi MyPertamina berperan penting dalam menjalankan kebijakan tersebut. Namun, kebijakan ini memicu berbagai respon dari masyarakat.

Beberapa masyarakat setuju dengan kebijakan ini, namun tidak sedikit pula yang mengkritisnya. Data yang diperoleh dari survei oleh Litbang Kompas menunjukkan bahwa sekitar 81,1% responden tidak yakin penyaluran BBM dapat tepat sasaran meskipun pembelian dilakukan melalui aplikasi MyPertamina (Mutia Annur, 2022). Selain itu, aplikasi ini juga kerap mendapat ulasan negatif di Google Play Store dengan keluhan seperti masalah saat proses pendaftaran, kesulitan dalam menerima kode OTP, dan kesulitan akses aplikasi (R. Maulana et al., 2023). Ketidaksetujuan terhadap kebijakan ini juga banyak ditemui di media sosial X. Aplikasi X atau yang dulunya dikenal sebagai Twitter merupakan media sosial terpopuler di Indonesia dengan jumlah pengguna sebesar 24 juta orang berdasarkan data yang dirilis oleh World Of Statistics pada Agustus 2023 (Fahusni, 2023). Aplikasi ini telah menjadi media sosial andalan bagi masyarakat Indonesia untuk mengungkapkan opini berdiskusi mengenai isu-isu sosial, serta mengkritisi kebijakan pemerintah. (Harun et al., 2023)(Abduh & Cangara, 2022).

Penelitian ini mengambil data yang relevan dengan kebijakan subsidi tepat dari media sosial X, seperti postingan, *repost*, maupun komentar dari pengguna X untuk dilakukan analisis sentimen. Analisis sentimen merupakan teknik analisis data tekstual dengan menggunakan *Natural Language Processing* (NLP) dan *Machine Learning* (ML) yang bertujuan untuk mengelompokkan teks secara otomatis berdasarkan perasaan penulis, baik perasaan positif, negatif, maupun netral (Purnamasari et al., 2023). Pada penelitian ini algoritma yang digunakan adalah *Naive Bayes Classifier* yang merupakan algoritma pengklasifikasian dan pemodelan data (Harun et al., 2023). Berdasarkan penelitian yang dilakukan oleh (G. Darmawan et al., 2023) terhadap analisis sentimen pengguna MyPertamina menggunakan algoritma *Naive Bayes* didapatkan hasil bahwasanya algoritma ini memiliki keunggulan dari segi akurasi dan efisiensi waktu dibandingkan dengan algoritma lainnya yang sejenis. Hasil *confusion matrix* pada penelitian tersebut menunjukkan tingkat akurasi sebesar 91%. Penelitian lainnya oleh (Harun et al., 2023) juga menyatakan bahwa algoritma tersebut merupakan metode yang tepat untuk mengklasifikasikan teks.

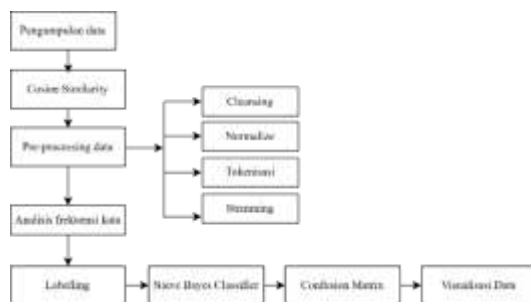
Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan subsidi tepat di aplikasi X menggunakan metode *Naive Bayes* serta mengidentifikasi faktor-faktor yang mempengaruhi opini masyarakat terhadap kebijakan tersebut. Hasil dari analisis ini berupa website sederhana yang menampilkan visualisasi data menggunakan *Streamlit*. *Streamlit* merupakan sebuah *framework open-source* yang digunakan untuk menampilkan hasil visualisasi data, model prediksi, dan analisis data lainnya dalam bentuk web interaktif menggunakan bahasa pemrograman *python* (Setiawan et al., 2024). Menurut penelitian oleh (D. Darmawan,



2024), penggunaan streamlit efektif sebagai media visualisasi data yang interaktif dengan akses yang mudah untuk masyarakat umum maupun pihak berwenang. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai persepsi masyarakat terhadap kebijakan subsidi tepat BBM.

2. METODE PENELITIAN

Dalam pelaksanaan penelitian ini dilakukan beberapa tahapan yang dapat dilihat pada gambar berikut.



Gambar 1. Metode Penelitian

Gambar di atas menampilkan tahapan dalam penelitian ini yang dimulai dari pengumpulan data, cosine similarity, *pre-processing data*, analisis frekuensi kata, labelling, uji model, dan visualisasi data. Tahapan-tahapan tersebut akan dijelaskan pada sub bab berikut.

2.1. Pengumpulan Data

Dataset untuk penelitian ini diambil dari media sosial X menggunakan metode *crawling data* dengan kata kunci “Subsidi tepat BBM”, “barcode BBM”, “subsidi pertalite”, dan “daftar MyPertamina”. Data yang digunakan merupakan data dari bulan September 2024 hingga November 2024.

2.2. Cosine Similarity

Pada tahap ini akan diidentifikasi data yang memiliki kemiripan kata menggunakan teknik cosine similarity. Teknik ini menghitung kemiripan antara dua objek berdasarkan bobotnya (Supiyanto & Sriyono, n.d.). Hal ini bertujuan untuk menemukan postingan yang diduga dibuat oleh buzzer. Data set kemudian akan dihitung *score* atau bobotnya untuk mengetahui tingkat kesamaannya. Data yang memiliki *score* >0.8 akan dideteksi sebagai postingan buzzer dan akan dihapus dari dataset.

2.3. Pre-processing Data

Berikutnya data akan diproses melalui beberapa tahapan, antara lain sebagai berikut:

a. Cleaning

Pada tahap ini data akan dibersihkan terlebih dahulu. Caranya dengan membersihkan data dari elemen-elemen yang tidak relevan. Pada tahap ini peneliti akan menghapus URL,

mention username, hashtag, emoji atau simbol khusus lainnya yang tidak dibutuhkan.

b. Normalisasi

Selanjutnya peneliti akan mengubah kalimat tidak baku atau slang menjadi kata baku. Contohnya kata “gak” diubah menjadi “tidak”.

c. Tokenisasi

Setelah proses normalisasi dilakukan, berikutnya peneliti akan memecah kalimat menjadi unit-unit kata. Contohnya, kalimat “Kebijakan subsidi tepat membantu masyarakat” akan dipecah menjadi [Kebijakan, subsidi, tepat, membantu, masyarakat].

d. Stemming

Berikutnya peneliti akan mengubah kata ke bentuk dasar untuk menyederhanakan kata. Contohnya seperti kata “membantu” diubah menjadi “bantu”.

2.4. Analisis Frekuensi Kata

Tahap berikutnya akan dilakukan analisis frekuensi kata. Pada tahap ini digunakan metode *N-Grams*. *N-Grams* merupakan metode dalam *Natural Language Processing* (NLP) yang digunakan untuk untuk menemukan kata-kata yang sering muncul pada dataset. Metode ini berguna untuk memperkirakan kemungkinan kemunculan kata-kata berikutnya berdasarkan konteks *N-Grams* sebelumnya (Jain, 2024). Hasil analisis terbagi menjadi 3, yakni *Unigram* (satu kata), *Bigram* (dua kata), dan *Trigram* (tiga kata).

2.5. Labelling

Selanjutnya, data akan diberi label positif, negatif dan netral menggunakan metode *Lexicon*. Lalu akan dilakukan pengecekan secara manual untuk memastikan data telah dilabeli dengan tepat. Jika terdapat label yang tidak sesuai, maka akan dilakukan perbaikan secara manual.

2.6. Naive Bayes Classifier

Berikutnya data diolah menggunakan algoritma *Naive Bayes*. Metode ini merupakan salah satu metode pada klasifikasi data mining. Dengan menggunakan algoritma ini data dapat diklasifikasikan dengan menghitung probabilitas pada data dan mengasumsikan semua atribut independen atau tidak terhubung satu sama lain (Mustika et al., 2021).

2.7. Confusion Matrix

Selanjutnya untuk mengukur performa model dilakukan evaluasi menggunakan *confusion matrix*. Metode ini akan menghitung akurasi, presisi, *recall* dan *F-1 score* dari model yang digunakan. *Confusion matrix* digunakan untuk mengetahui apakah model yang digunakan benar atau salah dalam melakukan klasifikasi (Simamora et al., 2024).

2.8. Visualisasi Data

Setelah didapat hasil analisis data maka akan dibuat visualisasinya menggunakan *framework* Streamlit berupa *website* yang berisi grafik data.



3. HASIL DAN PEMBAHASAN

3.1. Hasil

a. Pengumpulan Data

Pada penelitian ini data set diperoleh dari media sosial X meliputi opini publik yang relevan dengan topik kebijakan subsidi tepat, postingan, *repost*, dan komentar pengguna media sosial X. Pengambilan data dilakukan menggunakan teknik *crawling data* dengan menggunakan kata kunci “Subsidi tepat BBM”, “barcode BBM”, “subsidi pertalite”, dan “daftar MyPertamina”. Proses *crawling data* dilakukan secara bertahap karena adanya perubahan kebijakan pada media sosial X berupa limit pengambilan data per hari sebesar 600 data. Kemudian, dalam proses penelitian ini ditemukan bahwasanya penggunaan kata kunci secara bersamaan menghasilkan data yang terbatas, maka untuk mengatasinya kata kunci dimasukkan satu per satu atau secara terpisah guna memaksimalkan jumlah data yang diperoleh. Total data yang didapat dari hasil *crawling* adalah sebanyak 1405 data. Data yang telah didapat kemudian digabungkan ke dalam satu file csv dan diurut berdasarkan tanggal postingan.

Tabel 1. Hasil Crawling Data

No	Mention
1	@neVerAl0nely @msaid_didu @prabowo Mafia barcode bbm bersubsidi byk juga yg terlibat
2	@SelebtwitMobil brarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bbm jg bisa tepat sasaran
3	@SelebtwitMobil duh padahal kan adanya kebijakan qr code itu biar bbm subsidi tepat sasaran buat yang berhak inimah emg orangnya aja sih serakah.
4	@SelebtwitMobil daftar di mypertamina langsung rijek duluan
5	Ribet banget layanan subsidi tepat aplikasi mypertamina udah berhasil daftar sampe dapet QR Code masih aja harus daftar ulang lagi

b. Cosine Similarity

Tahapan berikutnya adalah *cosine similarity*. Tahapan ini dilakukan untuk mengidentifikasi postingan buzzer pada data set yang didapat dari media sosial X.

```

Teks 102: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Teks 873: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Cosine Similarity: 0.91

Teks 102: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Teks 875: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Cosine Similarity: 0.91

Teks 102: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Teks 881: Ombudsman RI bareng Pertamina turun tangan ke SPBU!
Cosine Similarity: 1.00

```

Gambar 2. Hasil Cosine Similarity

Gambar di atas menampilkan data yang memiliki kemiripan dari segi kalimat. Pada tahap ini ditemukan sebanyak 503 data yang memiliki kemiripan. Setelah dilakukan pengecekan secara manual, data tersebut kebanyakan berasal dari akun yang menyebarkan berita atau artikel secara berulang. Data tersebut kemudian dihapus salah satunya dari data set agar tidak mempengaruhi hasil dari analisis sentimen. Sehingga data set berkurang dari 1405 data menjadi 900 data.

c. Pre-Processing Data

Setelah mendapatkan data mentah dari hasil *crawling data* berikutnya dilakukan *pre-processing data* yang dilakukan menggunakan Google Colab dan bahasa pemrograman Python. Pada tahap ini dataset akan diproses terlebih dahulu melalui beberapa tahapan untuk menghasilkan data yang siap untuk dianalisis. Berikut tahapan dari *pre-processing data*:

1. Cleaning

Proses yang pertama dilakukan adalah *cleaning*. Pada tahap ini data set dibersihkan dari elemen-elemen yang tidak relevan. Elemen yang dihapus berupa link, hashtag, emoji, tanda baca, mention, dan angka.

Tabel 2. Hasil Cleaning Data

No	Sebelum	Sesudah
1	@neVerAl0nely @msaid_didu @prabowo Mafia barcode bbm bersubsidi byk juga yg terlibat	Mafia barcode bbm bersubsidi byk juga yg terlibat
2	@SelebtwitMobil brarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bbm jg bisa tepat sasaran	brarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bbm jg bisa tepat sasaran
3	@SelebtwitMobil duh padahal kan adanya kebijakan qr code itu biar bbm subsidi tepat sasaran buat yang berhak inimah emg orangnya aja sih serakah.	duh padahal kan adanya kebijakan qr code itu biar bbm subsidi tepat sasaran buat yang berhak inimah emg orangnya aja sih serakah
4	@SelebtwitMobil daftar di mypertamina langsung rijek duluan	daftar di mypertamina langsung rijek duluan
5	Ribet banget layanan subsidi tepat aplikasi mypertamina udah berhasil daftar sampe dapet QR Code masih aja harus daftar ulang lagi	Ribet banget layanan subsidi tepat aplikasi mypertamina udah berhasil daftar sampe dapet QR Code masih aja harus daftar ulang lagi

2. Normalisasi

Berikutnya dilakukan proses normalisasi untuk mengubah kata tidak baku, singkatan, dan kata slang menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Pada tahapan ini digunakan data set Kamus Kata Baku yang diperoleh dari kaggle.com. Sebelum dinormalisasi, teks pada data set diubah terlebih dahulu menjadi huruf kecil dengan teknik *case folding* agar proses normalisasi dapat berjalan dengan akurat dan efisien.



Tabel 3. Hasil Normalisasi Data

No	Sebelum	Sesudah
1	Mafia barcode bbm bersubsidi byk juga yg terlibat	mafia barcode bahan bakar minyak bersubsidi banyak juga yang terlibat
2	brarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bbm jg bisa tepat sasaran	berarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bahan bakar minyak juga bisa tepat sasaran
3	aduh padahal kan adanya kebijakan qr code itu biar bbm subsidi tepat sasaran buat yang berhak inimah emg orangnya aja sih serakah	aduh padahal kan adanya kebijakan qr code itu biar bahan bakar minyak subsidi tepat sasaran buat yang berhak inimah memang orangnya saja sih serakah
4	daftar di mypertamina langsung rjiek duluan	daftar di mypertamina langsung rjiek lebih dahulu
5	Ribet banget layanan subsidi tepat aplikasi mypertamina udah berhasil daftar sampe dapet QR Code masih aja harus daftar ulang lagi	ribet banget layanan subsidi tepat aplikasi mypertamina sudah berhasil daftar sampai dapat qr code masih saja harus daftar ulang lagi

3. Tokenisasi

Setelah data selesai dinormalisasi, berikutnya dilakukan proses tokenisasi. Pada tahap ini data akan dibagi menjadi unit-unit yang lebih kecil yang disebut token.

Tabel 4. Hasil Tokenisasi Data

No	Sebelum	Sesudah
1	mafia barcode bahan bakar minyak bersubsidi banyak juga yang terlibat	['mafia', 'barcode', 'bahan', 'bakar', 'minyak', 'bersubsidi', 'banyak', 'juga', 'yang', 'terlibat']
2	berarti tujuan adanya pembayaran menggunakan qr kode ini membantu kan supaya subsidi bahan bakar minyak juga bisa tepat sasaran	['berarti', 'tujuan', 'adanya', 'pembayaran', 'menggunakan', 'qr', 'kode', 'ini', 'membantu', 'kan', 'supaya', 'subsidi', 'bahan', 'bakar', 'minyak', 'juga', 'bisa', 'tepat', 'sasaran']
3	aduh padahal kan adanya kebijakan qr code itu biar bahan bakar minyak subsidi tepat sasaran buat yang berhak inimah memang orangnya saja sih serakah	['aduh', 'padahal', 'kan', 'adanya', 'kebijakan', 'qr', 'code', 'itu', 'biar', 'bahan', 'bakar', 'minyak', 'subsidi', 'tepat', 'sasaran', 'buat', 'yang', 'berhak', 'inimah', 'memang', 'orangnya', 'saja', 'sih', 'serakah']
4	daftar di mypertamina langsung rjiek lebih dahulu	['daftar', 'di', 'mypertamina', 'langsung', 'rjiek', 'lebih', 'dahulu']
5	ribet banget layanan subsidi tepat aplikasi mypertamina sudah berhasil daftar sampai dapat qr code masih aja harus daftar ulang lagi	['ribet', 'banget', 'layanan', 'subsidi', 'tepat', 'aplikasi', 'mypertamina', 'sudah', 'berhasil', 'daftar', 'sampai', 'dapat', 'qr', 'code', 'masih', 'aja', 'harus', 'daftar', 'ulang', 'lagi']

4. Stemming

Proses yang terakhir adalah *stemming*. Pada tahapan ini, tiap kata dalam data set disederhanakan dengan mengubah kata ke dalam bentuk dasarnya.

Tabel 5. Hasil Stemming Data

No	Sebelum	Sesudah
1	['mafia', 'barcode', 'bahan', 'bakar', 'minyak', 'bersubsidi', 'banyak', 'juga', 'yang', 'terlibat']	mafia barcode bahan bakar minyak subsidi banyak juga yang libat
2	['berarti', 'tujuan', 'adanya', 'pembayaran', 'menggunakan', 'qr', 'kode', 'ini', 'membantu', 'kan', 'supaya', 'subsidi', 'bahan', 'bakar', 'minyak', 'juga', 'bisa', 'tepat', 'sasaran']	arti tuju ada bayar guna qr kode ini bantu kan supaya subsidi bahan bakar minyak juga bisa tepat sasaran
3	['aduh', 'padahal', 'kan', 'adanya', 'kebijakan', 'qr', 'code', 'itu', 'biar', 'bahan', 'bakar', 'minyak', 'subsidi', 'tepat', 'sasaran', 'buat', 'yang', 'berhak', 'inimah', 'memang', 'orangnya', 'saja', 'sih', 'serakah']	aduh padahal kan ada bijak qr code itu biar bahan bakar minyak subsidi tepat sasaran buat yang hak inimah memang orang saja sih serakah
4	['daftar', 'di', 'mypertamina', 'langsung', 'rjiek', 'lebih', 'dahulu']	daftar di mypertamina langsung rjiek lebih dahulu
5	['ribet', 'banget', 'layanan', 'subsidi', 'tepat', 'aplikasi', 'mypertamina', 'sudah', 'berhasil', 'daftar', 'sampai', 'dapat', 'qr', 'code', 'masih', 'aja', 'harus', 'daftar', 'ulang', 'lagi']	ribet banget layan subsidi tepat aplikasi mypertamina sudah berhasil daftar sampai dapat qr code masih saja harus daftar ulang lagi

5. Analisis Frekuensi Kata

Tahap berikutnya dilakukan analisis frekuensi kata menggunakan N-Grams. Hal ini dilakukan untuk melihat frekuensi kata yang muncul pada data set guna mengetahui pemahaman masyarakat terhadap kebijakan subsidi tepat. Berikut hasil yang diperoleh dari analisis N-Grams :

Tabel di atas menampilkan 10 kata yang sering muncul pada data set. Pada kolom unigram terdiri atas satu kata, bigram terdiri atas dua kata, dan trigram terdiri atas tiga kata. Berdasarkan hasil analisis frekuensi kata ini barulah kemudian dilakukan analisis secara manual untuk mengidentifikasi bagaimana tingkat pemahaman masyarakat terhadap kebijakan subsidi tepat dan penggunaan aplikasi MyPertamina.

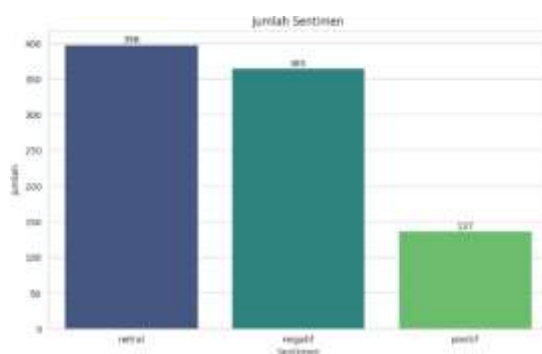
Berikut ini adalah tabel yang menampilkan hasil dari N-Grams

Tabel 6. Hasil N-Grams

No	Unigram	Bigram	Trigram
1	bakar	bahan bakar	bahan bakar minyak
2	bahan	bakar minyak	bakar minyak subsidi
3	minyak	tepat sasaran	subsidi bahan bakar
4	subsidi	minyak subsidi	tidak tepat sasaran
5	tepat	subsidi bahan	beli bahan bakar
6	sasar	subsidi tepat	subsidi tepat sasaran
7	yang	qr code	isi bahan bakar
8	dan	tidak tepat	bantu langsung tunai
9	tidak	beli bahan	minyak subsidi tepat
10	daftar	isi bahan	lebih tepat sasaran

6. Labelling Data

Setelah *pre-processing data* selesai dilakukan, diperoleh sebanyak 900 data yang siap untuk diolah. Berikutnya data dilabeli dengan kategori positif, negatif, dan netral menggunakan metode *Lexicon*. Kemudian, dilakukan pengecekan secara manual untuk memastikan bahwa seluruh data telah mendapat label yang sesuai. Berikut tampilan grafik hasil pelabelan data sentimen masyarakat terhadap kebijakan subsidi tepat. Dari hasil pelabelan menggunakan *lexicon* ditemukan bahwasanya banyak label yang tidak sesuai dikarenakan metode ini tidak mampu mendeteksi teks yang mengandung sarkasme. Oleh karena itu dataset dilabeli kembali secara manual. Hasil dari labelling secara manual dapat dilihat pada gambar berikut.



Gambar 2 Grafik Labelling Dataset

7. Uji Naive Bayes

Setelah serangkaian proses pengolahan data telah selesai, berikutnya dilakukan pengujian data menggunakan metode *Naive Bayes*. Data yang telah dilabeli akan diuji kembali menggunakan algoritma

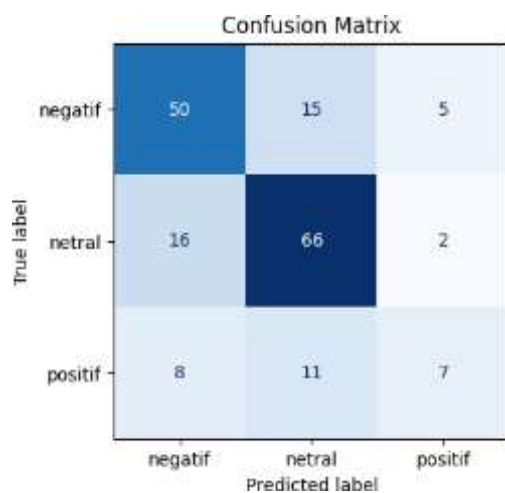


Naïve Bayes dengan perbandingan data latih dan data uji yakni 80:20. Data latih yang digunakan sebesar 720 data dan data uji sebesar 180 data. Berikutnya dilakukan perhitungan manual untuk

menunjukkan cara kerja algoritma *Naïve Bayes*. Sebagai contoh salah satu data yang akan diklasifikasikan adalah "**Daftar MyPertamina ribet**". Teks tersebut akan dikategorikan ke dalam sentimen negatif, netral, atau positif.

8. Confusion Matrix

Berdasarkan hasil evaluasi confusion matrix menggunakan google colab, diperoleh hasil sebagai berikut



Gambar 3 Hasil *Confusion Matrix*

Berdasarkan gambar tersebut, dapat diketahui bahwa model memiliki kinerja yang baik dalam memprediksi data netral karena sebagian besar data netral berhasil diklasifikasikan dengan benar sebesar 66 data. Namun, model masih kesulitan ketika membedakan data positif dan negatif, terutama data positif dengan akurasi prediksi yang rendah hanya sebesar 7 data. Berikut ditampilkan juga tabel hasil evaluasi *confusion matrix* pada tabel berikut

Tabel 6. Hasil *Confusion Matrix*

True Label	Predicted Negatif	Predicted Netral	Predicted Positif
Negatif	50	15	5
Netral	16	66	2
Positif	8	11	7

Berikut perhitungan *confusion matrix* secara manual:

Precision

$$Negatif = \frac{50}{50 + 16 + 8} = \frac{50}{74} = 0.6757 = 67\%$$

$$Netral = \frac{66}{15 + 66 + 11} = \frac{66}{92} = 0.7174 = 71\%$$

$$Positif = \frac{7}{5 + 2 + 7} = \frac{7}{14} = 0.5 = 50\%$$

Recall

$$Negatif = \frac{50}{50 + 15 + 5} = \frac{50}{70} = 0.7143 = 71\%$$

$$Netral = \frac{66}{16 + 66 + 2} = \frac{66}{84} = 0.7857 = 78\%$$

$$Positif = \frac{7}{8 + 11 + 7} = \frac{7}{26} = 0.2692 = 26\%$$

F-1 Score

$$Negatif = 2 \times \frac{0.6757 \times 0.7143}{0.6757 + 0.7143} = 0.6945 = 69\%$$

$$Netral = 2 \times \frac{0.7147 \times 0.7857}{0.7147 + 0.7857} = 0.7501 = 75\%$$

$$Positif = 2 \times \frac{0.5 \times 0.2692}{0.5 + 0.2692} = 0.3529 = 35\%$$

Accuracy

$$Total Sample = \frac{50 + 66 + 7}{50 + 15 + 5 + 16 + 66 + 2 + 8 + 11 + 7} = 0.6833 = 68\%$$

Tabel 7. Hasil Uji *Confusion Matrix*

Sentiment	Precision	Reccal	F1-Score
Negatif	0.68	0.71	0.69
Netral	0.72	0.79	0.75
Positif	0.50	0.27	0.35
Accuracy	0.68		

Berdasarkan tabel di atas dapat diketahui bahwasanya algoritma ini memiliki tingkat akurasi yang cukup baik, yakni sebesar 68%. Namun performa model ini masih kurang optimal dalam mengenali sentimen positif dilihat dari nilai *recall* dan F1-score yang rendah pada sentimen positif.

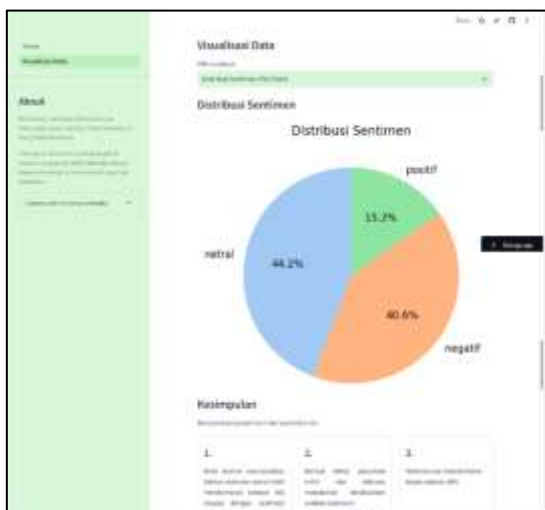
9. Visualisasi Data

Hasil visualisasi ini berupa *website* interaktif yang berisi dua page. Berikut tampilan dari *website* yang telah dibuat.



Gambar 4 Halaman Utama

Pada page pertama berisi konten mengenai kebijakan subsidi tepat seperti penjelasan singkat mengenai kebijakan tersebut, langkah-langkah untuk mendaftar subsidi tepat, serta dokumen yang perlu dipersiapkan untuk mendaftar subsidi tepat. Page ini bertujuan untuk menambah pemahaman pembaca mengenai kebijakan subsidi tepat. Berikutnya user dapat melihat hasil visualisasi data pada page kedua yang dapat diklik pada menu *sidebar navigation*.



Gambar 5 Halaman Visualisasi Data

Pada laman kedua terdapat penjelasan mengenai hasil analisis sentimen dan tampilan visualisasi data. User dapat memilih tampilan visualisasi data berupa grafik, *pie chart*, dan wordcloud melalui fitur select box. Web ini juga dapat diakses secara publik oleh pembaca.

3.2. PEMBAHASAN

Setelah melalui berbagai tahapan analisis, diperoleh hasil klasifikasi sentimen masyarakat terhadap kebijakan Subsidi Tepat dan penggunaan Aplikasi MyPertamina ke dalam tiga kategori, yaitu netral sebesar 44.2%, negatif 40.6%, dan positif 15.2%, dengan tingkat akurasi model sebesar 68%. Beberapa penelitian terdahulu yang relevan dan menjadi rujukan dalam penelitian ini antara lain adalah studi oleh (Dwi Anggita et al., 2021) yang menganalisis sentimen masyarakat terhadap

kebijakan Pembatasan Sosial Berskala Besar (PSBB) menggunakan algoritma Naive Bayes dengan teknik seleksi fitur Improved Gini Index. Hasil akurasi yang diperoleh pada penelitian tersebut sebesar 64% tanpa seleksi fitur, dan meningkat menjadi 68% setelah seleksi fitur. Dibandingkan dengan penelitian tersebut, penelitian ini memiliki hasil akurasi sebesar 68% yang dicapai tanpa penerapan seleksi fitur tambahan, yang menunjukkan bahwa model dapat bekerja secara optimal meskipun dengan pendekatan yang lebih sederhana.

Penelitian lainnya oleh (Sergio et al., 2023) menganalisis sentimen masyarakat terhadap kebijakan penghapusan subsidi BBM menggunakan Bernoulli Naive Bayes dan teknik ekstraksi fitur N-Gram TF-IDF. Akurasi tertinggi diperoleh pada Unigram sebesar 69%, sementara Bigram dan Trigram mengalami penurunan akurasi hingga 57%. Jika dibandingkan dengan kedua penelitian tersebut, penelitian ini memiliki akurasi yang lebih unggul sebesar 68%. Penelitian ini juga memiliki keunggulan dalam hal kualitas data, yaitu dengan menerapkan cosine similarity untuk menyaring opini dari akun *buzzer* atau bot, sehingga data yang dianalisis lebih representatif terhadap opini masyarakat yang sesungguhnya. Proses pelabelan sentimen juga dilakukan secara lebih teliti, yakni dengan mengombinasikan pendekatan *lexicon* dan pelabelan manual, yang meningkatkan akurasi label dan mengurangi risiko bias klasifikasi.

Meski demikian, penelitian ini tetap memiliki beberapa keterbatasan seperti sumber data yang hanya berasal dari satu platform dan tanpa melakukan survei atau wawancara langsung ke masyarakat, sehingga belum sepenuhnya mewakili opini publik secara luas. Selain itu, asumsi independensi antar fitur dalam algoritma Naive Bayes menjadi tantangan dalam menangkap makna utuh suatu teks. Model klasifikasi yang digunakan juga masih bersifat biner (positif dan negatif), tanpa mempertimbangkan keberadaan opini sarkasme atau ambigu.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut:

1. Hasil analisis sentimen menggunakan algoritma *naive bayes* menunjukkan bahwa sentimen netral lebih mendominasi sebesar 398 data disusul dengan sentimen negatif sebesar 354 data dan sentimen positif sebesar 137 data. Berdasarkan hasil analisis sentimen tersebut dan analisis frekuensi kata dapat diketahui bahwasanya mayoritas masyarakat cenderung memiliki pemahaman yang terbatas mengenai kebijakan subsidi tepat dan penggunaan aplikasi MyPertamina.
2. Faktor-faktor yang menyebabkan kritik dan keluhan dari masyarakat antara lain sebagai berikut:



- a) Ketidaktepatan sasaran subsidi BBM.
 - b) Wacana penghapusan subsidi BBM dan pengalihan ke Bantuan Langsung Tunai (BLT).
 - c) Lamanya proses verifikasi pendaftaran subsidi tepat.
 - d) Kegagalan mendapatkan kode OTP.
 - e) Nomor polisi kendaraan yang tidak terdaftar pada aplikasi MyPertamina.
 - f) Kesalahan deteksi jenis kendaraan (kendaraan roda 4 terdeteksi sebagai roda 2).
 - g) Masalah mafia barcode yang melakukan pemalsuan barcode untuk membeli BBM subsidi dalam jumlah besar.
3. Algoritma *Naive Bayes* mampu mengklasifikasikan sentimen masyarakat dengan tingkat akurasi sebesar 68%. Algoritma ini dapat memberikan gambaran awal terkait tingkat pemahaman masyarakat terhadap kebijakan subsidi tepat. Namun, algoritma ini masih kurang optimal dalam mengenali sentimen positif dikarenakan dataset yang digunakan banyak mengandung sarkasme dan ambiguitas.
- ## 5. REFERENSI
- Abduh, I. M., & Cangara, H. (2022). *Kritik Sosial Kebijakan Pemerintah Dalam Platform Media Sosial Dengan Pendekatan Komunikasi* (Vol. 8, Issue 1).
- Alvita Wagiswari D, P. R., Susilawati, I., Witanti, A., Studi Informatika, P., Teknologi Informasi, F., & Mercu Buana Yogyakarta, U. (2023). Analisis Sentimen pada Komentar Aplikasi MyPertamina dengan Metode Multinomial Naïve Bayes. In *Jurnal ForAI* (Vol. 1, Issue 1).
- Darmawan, D. (2024). *Metodologi Penelitian Dalam Pengembangan Aplikasi Streamlit: Studi Kasus Visualisasi Data Gempa Bumi Di Indonesia*. <https://doi.org/10.13140/RG.2.2.26320.32008>
- Darmawan, G., Alam, S., Imam Sulisty, M., Studi Teknik Informatika, P., Tinggi Teknologi Wastukencana Purwakarta, S., & Artikel, R. (2023). *Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes* Info Artikel Abstrak. 2(3), 100–108. <https://doi.org/10.55123>
- Dwi Anggita, K., Sari, Y. A., & Adinugroho, S. (2021). *Analisis Sentimen terhadap Opini Masyarakat mengenai Kebijakan PSBB menggunakan Metode Naïve Bayes dengan Seleksi Fitur Improved Gini Index* (Vol. 5, Issue 12). <http://j-ptiik.ub.ac.id>
- Fahusni, F. (2023, August 25). *Pengguna X Terbanyak di Dunia Menurut Akun World of Statistics*. Selular.ID. <https://selular.id/2023/08/pengguna-x-terbanyak-di-dunia-menurut-akun-world-of-statistics/>
- Harun, R., Ishak, R., & Panna, S. S. (2023). Analisis Sentimen Opini Publik Pengguna Twitter Terhadap Kenaikan Harga BBM Menggunakan Algoritma Naïve Bayes. *Copyright @BALOK*, 2(1).
- Herman, S., Nugraha, Y., Budiman, T., Rahayu, D., & Mutia, A. S. (2023). *Pemanfaatan Perkembangan Sistem Informasi dan Teknologi Informasi Dalam Menghadapi Tantangan Revolusi Industri 4.0 Pada Siswa-Siswi SMK Insan Prima Mandiri Garut*. 2(1), 74–83. <https://doi.org/10.56910/sewagati.v2i2>
- Jain, A. (2024, February 5). *N-grams in NLP*. <https://medium.com/@abhishekjainindore/24/n-grams-in-nlp-a7c05c1aff12>
- Kurniawan, R., & Ferdian, A. (2023, February 5). *Pembatasan BBM Solar Subsidi, Mobil Pribadi Dapat Jatah 60 Liter*. Kompas.Com. <https://otomotif.kompas.com/read/2023/02/05/162100015/pembatasan-bbm-solar-subsidi-mobil-pribadi-dapat-jatah-60-liter>
- Maulana, I., Apriandari, W., & Pambudi, A. (2023). Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Mypertamina Menggunakan Support Vector Machine. In *Idealis: Indonesia Journal Information System* (Vol. 6, Issue 2). <http://jom.fti.budiluhur.ac.id/index.php/I-DEALIS/index>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi Mypertamina Pada Google Play Store Menggunakan Algoritma Nbc. *Jurnal Teknologi Terpadu*.
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Agnes Manuhutu, M., Ridwan, M., Khrisna Wardhani, A., Alim, S., Romli, I., Religia, Y., Tri Octafian, D., Utan Sufandi, U., & Ernawati, I. (2021). *Data Mining dan Aplikasinya*. www.penerbitwidina.com
- Mutia Annur, C. (2022, July 8). *Survei: Efektivitas MyPertamina Banyak Diragukan* Warga. https://databoks.katadata.co.id/energi/statistik/7fd06110eb03ab8/survei-efektivitas-mypertamina-banyak-diragukan-warga?utm_source=chatgpt.com



- Purba, N., Yahya, M., & Nurbaiti. (n.d.). Revolusi Industri 4.0 : Peran Teknologi Dalam Eksistensi Penguasaan Bisnis Dan Implementasinya. In *JPSB* (Vol. 9, Issue 2).
- Purnamasari, D., Bayu, A., Desy, A., Fanka, W. A. P., Reza, A., Safrila, M., Yanda, O. N., & Hidayati, U. (2023). *Pengantar Metode Analisis Sentimen*.
- Sergio, Y., Putranta, V., Rahayudi, B., & Purnomo, W. (2023). *Analisis Sentimen Masyarakat terhadap Kebijakan Penghapusan Subsidi BBM pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram TF-IDF* (Vol. 7, Issue 3). <http://j-ptiik.ub.ac.id>
- Setiawan, A., Febrio Waleska, R., Adji Purnama, M., & Efrizoni, L. (2024). Komparasi Algoritma K-Nearest Neighbor (K-Nn), Support Vector Machine (Svm), Dan Decision Tree Dalam Klasifikasi Penyakit Stroke. In *Jurnal Informatika & Rekayasa Elektronik* (Vol. 7, Issue 1). <http://e-journal.stmiklombok.ac.id/index.php/jir> eISSN.2620-6900
- Simamora, S. D., Irwiensyah, F., & Hasan, F. N. (2024). Analisis Sentimen Terkait Konflik Palestina-Israel Pada Media Sosial X Menggunakan Algoritma Naïve Bayes Classifier. *Technology and Science (BITS)*, 6(1). <https://doi.org/10.47065/bits.v6i1.5344>
- Supiyanto, & Sriyono. (n.d.). *Metode Cosine Similarity Untuk Mendeteksi Kemiripan Pada Dokumen Teks*. <https://ejournal.uncen.ac.id/index.php/SAINS>
- Wardhani, P. A., Luqman, Y., & Noor, W. (2023). *Sikap Media detik.com Terhadap Kebijakan Subsidi Tepat Sasaran Mypertamina*. <https://www.fisip.undip.ac.id>