

SENTIMENT ANALISIS UNTUK IDENTIFIKASI KEPUASAN MASYARAKAT TERHADAP KENAIKAN BBM MENGGUNAKAN ALGORITMA NAÏVE BAYES

Ahmad Nur Badri, Noviandi², Farisa Anastya³,
dan Muhammad Roland⁴

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul Jakarta, Indonesia

Email: badrijkt@student.esaunggul.ac.id noviandi@esaunggul.ac.id²,
farisaanastia10@student.esaunggul.ac.id³, rolandmuhammad48@student.esaunggul.ac.id⁴

Abstrak

Kenaikan harga BBM tidak hanya menimbulkan beban yang lebih besar bagi masyarakat umum, tetapi juga berdampak pada dunia usaha. Dampak dari naiknya harga bahan bakar minyak (BBM) menyebabkan pro dan kontra dimasyarakat. Salah satu media sosial yang sering digunakan masyarakat untuk menyampaikan pendapat adalah Twitter. Twitter merupakan salah satu platform media sosial yang sering digunakan oleh masyarakat untuk menyampaikan pendapat. Dari beberapa media sosial yang ada, Twitter memiliki jumlah pengguna aktif terbanyak yaitu sekitar 211 juta orang. Dengan mengacu pada masalah tersebut, peneliti sebelumnya telah melakukan penelitian menggunakan analisis sentimen terhadap kenaikan harga bahan bakar min-yak (BBM) dengan menerapkan metode Naïve Bayes. Percobaan dilakukan menggunakan 1000 dataset dengan melakukan penambahan penerapan optimasi seleksi fitur Generate N-gram pada tahap preprocessing dan pelabelan memakai lexicon dan split data 80:20. Langkah awal dalam membagi dataset adalah dengan menerapkan Teknik Stratified Random Sampling, di mana dataset dibagi menjadi data train dan data test. Data train kemudian digunakan untuk membuat model Naïve Bayes, sedangkan data test digunakan untuk mengevaluasi kinerja model melalui penggunaan confusion matrix setelah melalui proses pelatihan. Berdasarkan hasil evaluasi dari penelitian yang telah dilakukan menghasilkan nilai accuracy 81%, precision 88%, recall 85% dan f1-score 87%.

Kata Kunci: BBM, Naïve Bayes, Analisis Sentimen, N-Gram, Twitter.

Abstract

The increase in fuel prices not only creates a greater burden for the general public, but also affects the business world. The impact of rising fuel prices has caused pros and cons in society. One of the social media that is often used by people to express their opinions is Twitter. Twitter is one of the social media platforms that is often used by people to express their opinions. Of the several social media available, Twitter has the highest number of active users, around 211 million people. With reference to this issue, previous researchers have conducted re-search using sentiment analysis on the increase in fuel prices by applying the Naïve Bayes method. Experiments were conducted using 1000 datasets by adding feature selection optimization, generating N-gram in the pre-processing stage, and labeling using lexicons, as well as splitting data 80:20. The initial step in dividing the dataset is by applying the Stratified Random Sampling technique, where the dataset is divided into train and test data. The train data is then used to create a Naïve Bayes model, while the test data is used to evaluate the performance of the model through the use of a confusion matrix after the training process. Based on the evaluation results of the research that has been conducted, it produces an accuracy value of 81%, precision of 88%, recall of 85%, and an F1-score of 87%.

KeyWords: BBM, Naïve Bayes, Analisis Sentimen, N-Gram, Twitter.

I. PENDAHULUAN

Bahan bakar minyak (BBM) adalah bahan baku yang mempunyai peranan penting dalam segala aspek kegiatan ekonomi. Kenaikan harga BBM tidak hanya menimbulkan beban yang lebih besar bagi masyarakat umum, tetapi juga berdampak pada dunia usaha [1]. Menurut laporan Kementerian ESDM, kenaikan ICP disebabkan oleh berbagai faktor geopolitik yang terjadi di banyak negara, seperti konflik Ukraina-Rusia. Faktor lain yang turut berperan adalah masalah logistik di Kazakhstan dan masalah politik di Uni Emirat Arab (UEA) [2]. Harga minyak mentah dan ICP belum turun sejak mengajukan tambahan subsidi dan kompensasi bahan bakar dan listrik ke DPR. Namun, tren harga minyak menunjukkan kenaikan yang signifikan. Menurut proyeksi harga minyak dari EIA hingga akhir tahun, harga minyak diperkirakan akan mencapai US\$104,8 per barel dan bahkan ada consensus yang memperkirakan harga minyak bisa mencapai US\$105 [3]. Dampak dari naiknya harga bahan bakar minyak (BBM) menyebabkan pro dan kontra di masyarakat. Twitter merupakan salah satu platform media sosial yang sering digunakan oleh masyarakat untuk menyampaikan pendapat.

Karena mudah diakses, memiliki jumlah pengikut yang tak terbatas, dan batasan karakter yang singkat hanya 280 karakter, Twitter dianggap sebagai tempat terbuka bagi pengguna untuk mengekspresikan pikiran dan pendapat mereka [4]. Dari beberapa media sosial yang ada, Twitter memiliki jumlah pengguna aktif terbanyak yaitu sekitar 556 juta pengguna di seluruh dunia [5]. Banyaknya pengguna Twitter bisa digunakan sebagai sumber informasi untuk memahami pendapat masyarakat tentang kenaikan harga BBM.

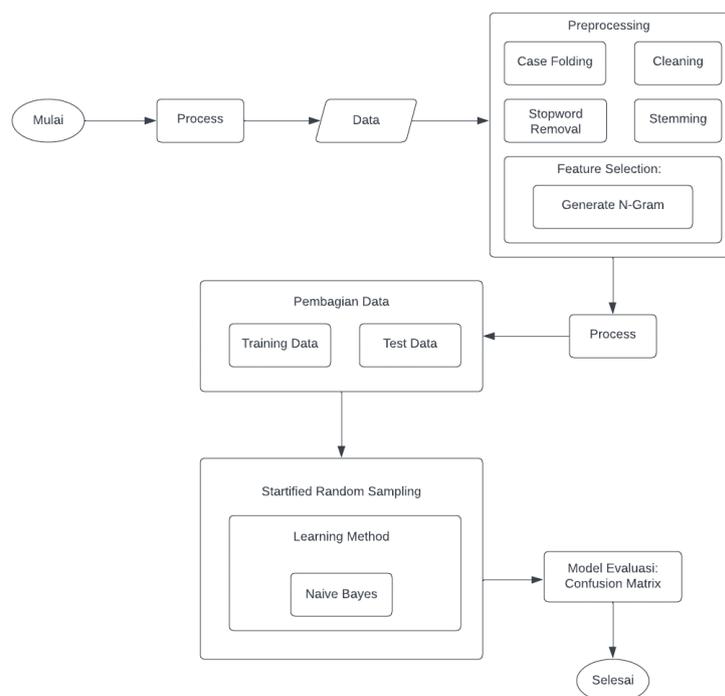
Menentukan apakah sebuah tweet memiliki sentimen negatif, positif, atau netral bisa dilakukan secara manual, namun hal

tersebut memerlukan waktu dan usaha yang signifikan dalam menganalisis semua tweet. Maka diperlukan klasifikasi teks salah satunya yakni metode Naïve Bayes [6]. Sejumlah penelitian sebelumnya telah menggunakan metode klasifikasi Naïve Bayes untuk menganalisis sentimen, contohnya adalah penelitian yang dilakukan oleh Ratnawati dkk, yakni menganalisa sentimen opini film melalui Twitter dengan implementasinya memperoleh tingkat akurasi 90%, *precision* 92%, *recall* 90% dan *f-measure* 90% [7]. Penelitian lain oleh Junaedi dkk, yakni analisis sentimen maskapai citilink pada twitter menghasilkan tingkat akurasi 0.77 dan peneliti yang melakukan analisis sentimen jasa transportasi online pada twitter oleh Pintoko dkk, mendapatkan hasil tingkat akurasi Naïve Bayes 86.80% [8]. Naïve Bayes adalah sebuah teknik klasifikasi yang menggunakan Teorema Bayes dengan menggunakan probabilitas dan statistik untuk memprediksi probabilitas suatu kejadian berdasarkan pengalaman masa lalu. Metode ini bersifat naif karena asumsi yang sangat kuat mengenai ketergantungan setiap kondisi. Naïve bayes mempunyai beberapa kelebihan seperti sederhana, cepat, akurasi yang tinggi dan banyak peneliti menggunakan Naïve Bayes untuk klasifikasi sentimen [9]. Pada ketiga paper tersebut, terdapat kelemahan yang dapat ditemukan pada paper pertama, di mana hanya terdapat 400 data yang digunakan sebagai data pelatihan. Akibatnya, ketika data uji dimasukkan, beberapa data tidak dapat dikenali karena tidak tercakup dalam data pelatihan, yang mengakibatkan hasil prediksi dan label yang tidak sesuai. Pada penelitian yang dilakukan Junaedi dkk, tidak adanya tahap seleksi fitur sehingga nilai akurasi masih belum maksimal [10]. Peneliti lainnya yang dilakukan oleh Pintoko dkk memanfaatkan pelabelan sentimen data secara manual menyebabkan terjadinya kesalahan prediksi pada sistem, yang pada akhirnya mengakibatkan hasil akurasi yang kurang optimal. Berdasarkan hasil penelitian sebelumnya, penulis tertarik untuk melakukan penelitian dengan menggunakan metode klasifikasi Naïve Bayes untuk menganalisis sentimen masyarakat terkait kenaikan harga bahan bakar minyak (BBM) yang diungkapkan melalui Twitter dengan melakukan penerapan optimasi seleksi fitur Generate N-gram pada tahap *preprocessing* untuk meningkatkan hasil akurasi yang tinggi dan melakukan pelabelan sentimen data secara otomatis menggunakan *library lexicon* untuk meminimalisir kesalahan dalam prediksi yang menyebabkan hasil akurasi tidak maksimal. Harapannya, penelitian ini dapat mengatasi kelemahan yang ditemukan pada penelitian-penelitian sebelumnya.

II. METODE PENELITIAN

A. Tahapan Penelitian

Dalam penelitian yang berjudul "Sentiment Analisis Untuk Identifikasi Kepuasan Masyarakat Terhadap Kenaikan BBM Menggunakan Algoritma Naïve Bayes", terdapat beberapa tahapan yang harus dilakukan, yaitu dimulai dengan pengumpulan data (*crawling data*), *Cleaning data (preprocessing)*, seleksi fitur N-gram, pelabelan data (*labelling*) menggunakan *library lexicon*, klasifikasi Naïve bayes, dan evaluasi menggunakan *confusion matrix*. Diagram tahapan-tahapan penelitian dapat ditunjukkan pada Gambar 1.



Gambar 1: Tahap-tahap Penelitian

B. Crawling Data

Tahap pertama yaitu pengambilan data (*crawling data*) dengan library Tweepy. *Crawling data* adalah proses memperoleh dataset berupa tweet, untuk menghasilkan input pada langkah preprocessing [11]. Pengumpulan data dilakukan dengan mengambil sejumlah 1000 tweet pengguna Twitter berdasarkan parameter kata kunci yaitu 'bbm naik'. Tweet tersebut akan digunakan sebagai data untuk menganalisis sikap masyarakat terhadap kenaikan harga BBM.

C. Preprocessing

Langkah selanjutnya adalah *preprocessing*, yang berperan dalam memproses data mentah sebelum dilakukan pemrosesan berikutnya. *Preprocessing* melibatkan penghapusan data yang tidak diperlukan dan pengubahan format data menjadi yang lebih mudah diproses oleh sistem komputasi. Tahap ini bertujuan untuk menghasilkan data tweet yang bersih agar *classifier* Naïve Bayes dapat diuji secara optimal. *Preprocessing* data melibatkan beberapa tahapan, yaitu:

1) Case Folding

Langkah awal adalah *case folding*, yaitu mengubah semua huruf kapital dalam dokumen menjadi huruf kecil. Proses *case folding* dapat dilihat pada Tabel I.

Tabel I: *Case Folding*

Sebelum <i>case folding</i>	Sesudah <i>case folding</i>
Ketika blm berkuasa sprti ini de- mo kenaikan harga bbm,tp saat ini berkuasa bbm naik sante saja,rindu tangisan mereka dulu waktu harga bbm naik. https://t.co/iY1e8oirpE	ketika blm berkuasa sprti ini de- mo kenaikan harga bbm,tp saat ini berkuasa bbm naik sante saja,rindu tangisan mereka dulu waktu harga bbm naik. https://t.co/iY1e8oirpE

2) Cleaning

Cleaning merupakan tahap membersihkan data untuk menghilangkan angka, emoticon, hashtag (#), URL, username (), maupun simbol karena tidak dibutuhkan [12]. Proses *cleaning* dapat dilihat pada Tabel II.

Tabel II: Proses *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
ketika blm berkuasa sprti ini de- mo kenaikan harga bbm,tp saat ini berkuasa bbm naik sante saja,rindu tangisan mereka dulu waktu harga bbm naik. https://t.co/iY1e8oirpE	ketika blm berkuasa sprti ini de- mo kenaikan harga bbm,tp saat ini berkuasa bbm naik sante saja,rindu tangisan mereka dulu waktu harga bbm naik.

3) Stemming

Stemming adalah sebuah metode yang digunakan untuk menggabungkan kata-kata yang berbeda dalam bentuk yang sama [13]. Proses *stemming* ditunjukkan pada Tabel III.

Tabel III: Proses *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
ketika blm berkuasa sprti ini de- mo kenaikan harga bbm,tp saat ini berkuasa bbm naik sante saja,rindu tangisan mereka dulu waktu harga bbm naik.	ketika blm kuasa sprti ini demo na- ik harga bbm tp saat ini kuasa bbm naik sante saja rindu tangis mereka dulu waktu harga bbm naik.

4) Stop Word Removal

Stop word removal merupakan langkah dalam menghapus kata-kata dari daftar *stop word* Bahasa Indonesia. Tujuan dari proses *stopword removal* adalah untuk mengurangi kata-kata umum yang kurang penting dalam kamus *stopword*. Proses *stop word removal* dilihat pada Tabel IV.

Tabel IV: *Stop Word Removal*

Sebelum <i>Stop Word Removal</i>	Sesudah <i>Stop Word Removal</i>
ketika belum berkuasa seperti ini demo kenaikan harga bbm,tapi sa- at ini berkuasa bbm naik santai saja,rindu tangisan mereka dulu waktu harga bbm naik.	ketika belum seperti ini, tapi saat ini santai saja, rindu mereka dulu waktu naik.

5) *Generate N-Gram*

Generate N-gram adalah rangkaian n-kata yang muncul secara berurutan dalam sebuah kalimat atau kumpulan teks. N-gram bukanlah representasi teks, tetapi N-gram bisa berfungsi sebagai fitur yang mewakili teks itu sendiri [12]. Pada penelitian ini akan menggunakan sampai dengan jenis Uni-gram. Contoh dari proses pemisahan kalimat menggunakan N-gram ditunjukkan pada Tabel II-C5.

Tabel V: Generate N-Gram

<i>Uni- Gram</i>	[‘ketika’, ‘belum’, ‘berkuasa’, ‘se-perti’, ‘ini’, ‘demo’, ‘kenaikan’, ‘harga’, ‘bbm’, ‘tapi’, ‘saat’, ‘ini’, ‘berkuasa’, ‘BBM’, ‘naik’, ‘san-tai’, ‘saja’, ‘rindu’, ‘tangisan’, ‘me-reka’, ‘dulu’, ‘waktu’, ‘harga’, ‘BBM’, ‘naik’]
----------------------	---

D. *Labelling*

Tahap *Labelling* adalah proses pemberian kategori pada dokumen atau kalimat berdasarkan ciri-ciri atau karakteristik yang dimilikinya. Proses ini melibatkan memberikan label pada semua data berdasarkan karakteristik dari kalimat tersebut. Penelitian ini menggunakan pelabelan otomatis dengan *Library Lexicon*. *Library Lexicon* adalah perpustakaan Python (versi 2 dan 3) yang berfungsi untuk memproses data teks. Library ini menyediakan API yang dapat digunakan untuk pemrosesan bahasa alami (NLP), seperti ekstraksi frasa kata benda, analisis sentimen, klasifikasi, dan penerjemahan [14].

E. *Klasifikasi Naïve Bayes*

Tahap klasifikasi menggunakan metode Naïve bayes, yaitu metode klasifikasi yang menggunakan probabilitas dan statistik. Bertujuan untuk menduga kemungkinan masa depan berdasarkan pengalaman masa lalu [15]. Pada metode Naïve bayes, klasifikasi memiliki dua langkah yaitu tahap training untuk proses pembelajaran bagi dokumen yang kelasnya telah diketahui dan tahap test untuk proses pengujian dokumen yang kelasnya tidak diketahui [16]. Pada tahap ini digunakan Teknik Stratified Random Sampling yang tujuannya adalah agar distribusi training data dan test data proporsional dan tidak terdistorsi berdasarkan kumpulan data. Sistem ini menggunakan data split rasio 80 (Training data):20 (Test data). Penerapan metode klasifikasi Naïve bayes dapat dirumuskan sebagai berikut 1:

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

Keterangan:

X : Ini mengacu pada data yang tidak memiliki kelas yang diketahui.

H : Hipotesis dari data X adalah bahwa itu termasuk dalam kelas tertentu.

$P(H|X)$: Probabilitas posterior hipotesis H, yang didasarkan pada kondisi X.

$P(H)$: Probabilitas prior dari hipotesis H.

$P(X)$: probabilitas dari X

Input: Training Data Set T

$F = (f_1, f_2, f_3, \dots, f_n)$ Nilai variabel prediktor dalam pengujian pengumpulan data

Output: Kelas dataset pengujian

Tahapan:

- 1) Baca dataset pelatihan T
- 2) Hitung mean dan standar deviasi dari variabel prediktor di setiap kelas
- 3) Ulang Hitung probabilitas menggunakan gauss density equation di setiap kelas; Sampai probabilitas semua variabel prediktor F telah dihitung.
- 4) Hitung kemungkinan untuk setiap kelas
- 5) Dapatkan kemungkinan terbesar dengan pseudocode sebagai berikut 1.

Algorithm 1 Pseudocode Klasifikasi Naïve Bayes

```

FUNCTION train_naive_bayes(x_train, y_train):
    clf = MultinomialNB().fit(x_train, y_train)
    return clf

FUNCTION evaluate_naive_bayes(clf, x_test, y_test):
    predicted = clf.predict(x_test)
    accuracy = accuracy_score(y_test, predicted)
    precision = precision_score(y_test, predicted, average="micro")
    recall = recall_score(y_test, predicted, average="micro")
    f1 = f1_score(y_test, predicted, average="micro")
    cm = confusion_matrix(y_test, predicted)
    report = classification_report(y_test, predicted)
    RETURN accuracy, precision, recall, f1, cm, report
    
```

F. Evaluasi

Diperlukan evaluasi guna mengukur nilai akurasi, presisi, dan recall dari model Naïve Bayes yang diuji. Evaluasi tersebut menggunakan Confusion Matrix sebagai metode untuk menganalisis dan mengevaluasi kinerja sistem yang telah dirancang. Performa diukur dengan presisi, akurasi, dan recall. Rumus menghitung nilai akurasi, presisi dan recall terdapat pada persamaan 2, 3, dan 4.

$$[H]Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$[H]Presisi = \frac{TP}{TP + FP} \quad (3)$$

$$[H]Recall = \frac{TP}{TP + FN} \quad (4)$$

Keterangan:

TP : Istilah yang digunakan ketika sistem mengidentifikasi data positif sebenarnya sebagai data positif.

TN : Istilah yang digunakan ketika sistem mengidentifikasi data negatif sebenarnya sebagai data negatif.

FP : Istilah yang digunakan ketika sistem memprediksi data negatif sebenarnya sebagai data positif.

FN : Istilah yang digunakan ketika sistem memprediksi data positif sebenarnya sebagai data negatif.

III. HASIL DAN PEMBAHASAN

A. Crawling Data

Langkah pertama dalam penelitian ini adalah crawling data atau pengumpulan data sebagai bahan penelitian untuk menganalisis opini masyarakat terhadap kenaikan harga BBM. Dataset penelitian berasal dari media sosial Twitter berupa data tweet. Dataset tersebut diproses melalui crawling data dengan memakai library tweepy. Proses crawling data dapat dilihat pada Tabel VI.

Tabel VI: Stop Word Removal

<i>Index</i>	<i>Username</i>	<i>Text</i>
1	Mr NgibuliN	RT bachrum_achmadi: Harga Terbaru BBM Pertamina, Pertamina Turbo Naik Jadi Rp15.200. Ahok memang hebat!
2	HunterKiller-rzxx	RT bachrum_achmadi: Harga Terbaru BBM Pertamina, Pertamina Turbo Naik Jadi Rp15.200. Ahok memang hebat!
3	RahmanJabir Jabir	RT keuangannnews_id: Mulai Hari Ini, Harga BBM Pertamina Naik! https://t.co/qdfPIEDjOC
4	Zakirman Kari	RamliRizal Benar benar telah menjadikan rakyat sebagai sapi perahan bayangkan harga tiket kereta api cepat dgn ker https://t.co/deV04FZybT
—	—	—
999	Babe galak banget	Yang penting BBM subsidi ngak naik dan tetap satu harga untuk seluruh Indonesia https://t.co/6bGDKWR60j https://t.co/ZfIE2MIYOv
1000	Christian Halim	RT adhams: Yak. Shell sudah naik juga. Walau di web-nya blom update. Pokoknya semua Shell Web https://t.co/vKo34neJ4g udah updated.

Berdasarkan Tabel VI, terdapat kolom berupa username dan text. Data tweet dari 1000 pengguna Twitter berdasarkan parameter kata kunci yaitu ‘bbm naik’ pada isu kenaikan harga BBM di Indonesia. Dataset yang diekstraksi selama proses crawling data yaitu: username, dan text (tweet). Hasilnya kemudian disimpan ke dalam file CSV (.csv).

B. Preprocessing

Tahap preprocessing dilakukan dalam beberapa tahap yaitu case folding untuk mengubah semua kata-kata yang terdapat huruf besar menjadi huruf kecil. Selanjutnya cleaning tujuan dari proses ini adalah untuk membersihkan data dengan cara menghilangkan angka, emoticon, hashtag (#), url, username (), tanda baca, whitespace dan karakter selain huruf a-z diubah menjadi spasi. Kemudian Stemming adalah cara menggabungkan bentuk-bentuk kata yang berbeda menjadi satu bentuk yang sama. Proses Selanjutnya proses stopword removal bertujuan untuk mengurangi kata umum yang tidak mempunyai arti dalam kamus stopword. Proses terakhir, yaitu generate N-gram bertujuan untuk memberikan kemungkinan kata-kata berikutnya yang dapat digunakan untuk membuat kemungkinan kombinasi dari seluruh kalimat. Proses preprocessing dapat dilihat pada Tabel VII.

Tabel VII: Preprocessing Data

<i>Preprocessing</i>	<i>Text</i>
Data mentah	RT bachrum_achmadi: Harga Terbaru BBM Pertamina, Pertamina Turbo Naik Jadi Rp15.200. Ahok memang hebat!
Case folding	rt bachrum_achmadi: harga terbaru bbm pertamina, pertamax turbo naik jadi rp15.200. ahok memang hebat!
Cleaning	rt bachrum_achmadi harga terbaru bbm pertamina, pertamax turbo naik jadi rp ahok memang hebat
Stemming	rt achmadi harga baru bbm pertamina pertamax turbo naik jadi rp ahok memang hebat
Stopword removal	achmadi harga baru bbm pertamina pertamax turbo rp ahok hebat
Generate N-gram	‘achmadi’, ‘harga’, ‘baru’, ‘bbm’, ‘pertamina’, ‘pertamax’, ‘turbo’, ‘rp’, ‘ahok’, ‘hebat’

Proses tahap cleaning dan preprocessing pada Tabel VII dibutuhkan untuk memproses data yang lebih bersih, sebelum masuk ke tahap labelling. Proses ini juga data akan diubah kedalam bentuk yang lebih dipahami sistem dan menghilangkan beberapa masalah pada data yang akan mengganggu saat pemrosesan data.

C. Labelling

Proses selanjutnya adalah memberi label data atau menetapkan polaritas ke dataset yang telah dibersihkan (preprocessing). Proses pengelompokan polaritas ini secara otomatis dilakukan menggunakan lexicon yang terdiri dari 3 polaritas yaitu positif (diberi nilai 1), netral (diberi nilai 0), dan negatif (diberi nilai -1). Dalam proses memberi label data terdapat perubahan dataset dari bahasa Indonesia menjadi bahasa Inggris menggunakan library googletrans agar dapat dilakukan labelling. Proses translate ditunjukkan pada Tabel VIII.

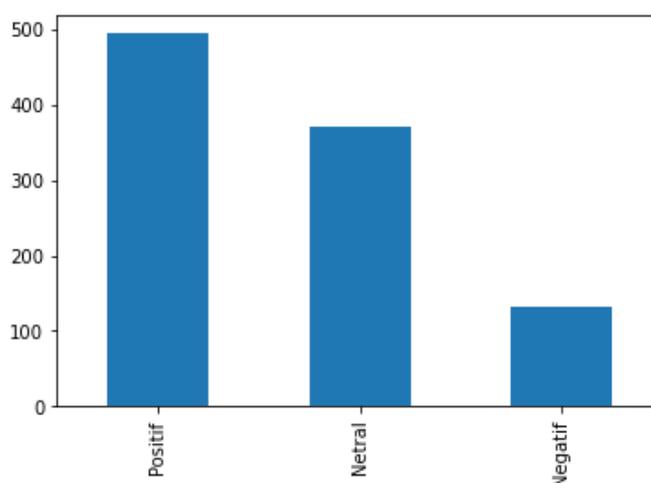
Tabel VIII: Translate Data

Sebelum	Sesudah
['jadi', 'rakyat', 'sapi', 'perah', 'bayang', 'harga', 'tiket', 'kereta', 'api', 'cepat']	['so', 'people', 'cow', 'milk', 'shadow', 'price', 'ticket', 'train', 'fire', 'fast']
['etis', 'rawan', 'izin', 'presiden', 'unt', 'tempur', 'rakyat', 'rakyat', 'susah']	['ethical', 'prone', 'permit', 'preside', 'unt', 'combat', 'the people', 'the people', 'difficult']
['achmadi', 'tolol', 'hutang', 'mikir', 'bebanin', 'rakyat']	['achmadi', 'stupid', 'debt', 'think', 'burden', 'the people']
['rambut', 'presiden', 'putih', 'mukaknya', 'kerut', 'bong']	['hair', 'president', 'putin', 'his face', 'wrinkles', 'bong']

Tabel IX: Labelling Data

Data	Polarity Score	Polarity
['so', 'people', 'cow', 'milk', 'shadow', 'price', 'ticket', 'train', 'fire', 'fast']	-0.34	Negatif
['ethical', 'prone', 'permit', 'preside', 'unt', 'combat', 'the people', 'the people', 'difficult']	-0.1531	Negatif
['achmadi', 'stupid', 'debt', 'think', 'burden', 'the people']	-0.8316	Negatif
['hair', 'president', 'putin', 'his face', 'wrinkles', 'bong']	0	Netral

Berdasarkan Tabel III-C, Untuk melakukan pelabelan, nilai pada angka akan dijumlahkan dengan mempertimbangkan bobot yang diberikan oleh kamus pada lexicon. Penentuan polarity ditentukan berdasarkan polarity score jika >0 maka diklarifikasikan negatif, jika $=0$ maka diklarifikasikan netral, jika <0 maka diklarifikasikan menjadi positif. Hasil labelling data dapat dilihat pada Gambar 2.



Gambar 2: Hasil Labelling Dataset

Berdasarkan pada Gambar 2 proses labelling menghasilkan sentimen terbanyak positif yang menyatakan setuju kenaikan BBM. Menurut pemerintah agar dapat mengurangi subsidi BBM, yaitu melalui kenaikan harga BBM karena penerapan subsidi

BBM terlalu besar, dari Rp 170an triliun naik menjadi Rp 520 triliun [17], jika pemerintah melakukan penyesuaian harga BBM, maka Indonesia akan terhindar dari dampak negative seperti krisis dan kebangkrutan negara.

D. Pembagian Data

Tahap ini melibatkan pembagian dataset ke dalam data train dan data test dengan menggunakan Teknik Stratified Random Sampling. Tujuannya adalah untuk memastikan proporsi dataset seimbang dan menghindari bias pada data. Total data tweet yaitu 1000. Pada pembagian data tersebut dilakukan secara random oleh sistem dengan data rasio 80:20. Dari data tersebut, terdapat 800 data yang digunakan sebagai data train dan 200 data yang digunakan sebagai data test.

E. Klasifikasi Naïve Bayes

Tahap klasifikasi Naïve Bayes melibatkan perhitungan probabilitas untuk setiap kelas yang terdapat pada dataset, yaitu negatif, netral, dan positif. Selanjutnya dilakukan perhitungan setiap kata pada kelas. Setelah tahap labeling dan pembagian data, dilakukan tahap klasifikasi menggunakan metode Naïve Bayes. Pada tahap ini, data training digunakan untuk mengembangkan model Naïve Bayes, sementara data testing digunakan untuk mengevaluasi kinerja model yang telah dilatih. Hasil klasifikasi yang telah dilakukan terdapat didalam Tabel X.

Tabel X: Hasil Klasifikasi

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
81%	88%	85%	87%

Berdasarkan Tabel X, didapatkan hasil accuracy sebesar 0.81, precision sebesar 0.81, recall 0.81 dan f1-score sebesar 0.81.

F. Evaluasi

Evaluasi adalah tahap terakhir dalam proses klarifikasi yang sudah dilakukan, bertujuan untuk menguji seberapa akurat sistem yang telah dibangun. Data yang digunakan berjumlah 200 data yang terdiri dari 32 negatif, 69 netral dan 99 positif. Setelah evaluasi data, hasilnya akan ditampilkan dalam bentuk confusion matrix yang mencakup perhitungan accuracy, precision, recall dan f1-score. Hasil evaluasi model terdapat pada Tabel XI.

Tabel XI: Hasil Evaluasi Menggunakan Confusion Matrix

	Precision	Recall	F1-score	Support
Negatif	0.95	0.59	0.73	32
Netral	0.69	0.86	0.77	69
Positif	0.88	0.85	0.87	99
Accuracy			0.81	200
Macro avg	0.84	0.77	0.79	200
Weighted avg	0.83	0.81	0.81	200

Accuracy adalah keakuratan kumpulan data yang diprediksi. *Precision* adalah Akurasi kasus positif yang diprediksi dengan benar. *Recall* adalah Akurasi kasus positif yang teridentifikasi dengan benar [18]. *F1-score* adalah Ukuran kinerja yang mengambil presisi p dan mengembalikan r untuk menghitung skor [19]. Berdasarkan Tabel X, hasil evaluasi dari penelitian yang telah dilakukan menghasilkan nilai accuracy 81%, precision 88%, recall 85% dan f1-score 87%.

IV. KESIMPULAN

Setelah dilakukan pengujian pada sentimen kenaikan harga BBM menggunakan algoritma Naïve Bayes dan fitur seleksi N-Gram, diperoleh hasil akurasi sebesar 81%, presisi sebesar 88%, recall sebesar 85%, dan f1-score sebesar 87%. Hasil ini menunjukkan bahwa model klasifikasi memiliki kemampuan yang cukup baik dalam mengenali sentimen negatif, netral, dan positif terkait dengan kenaikan harga BBM di media sosial Twitter. Nilai akurasi yang diperoleh menunjukkan persentase data yang diklasifikasikan dengan benar oleh model. Nilai presisi menunjukkan seberapa banyak dari data yang diklasifikasikan se-bagai sentimen negatif, netral, atau positif benar-benar memiliki label tersebut. Nilai recall menunjukkan seberapa banyak data yang seharusnya memiliki label sentimen negatif, netral, atau positif berhasil diklasifikasikan dengan benar oleh model. Sedangkan, f1-score merupakan harmonic mean dari nilai presisi dan recall.

Dari hasil tersebut, dapat disimpulkan bahwa model klasifikasi memiliki kemampuan yang baik dalam mengenali sentimen terkait kenaikan harga BBM. Oleh karena itu, hasil analisis sentimen dapat digunakan sebagai acuan dalam mengambil

keputusan atau mengevaluasi opini masyarakat terhadap kenaikan harga BBM. Namun, untuk meningkatkan performa prediksi sentimen pada penelitian selanjutnya, penulis menyarankan untuk menambahkan algoritma optimasi seperti algoritma PSO untuk meningkatkan nilai akurasi, presisi, recall, dan f1-score dari model klasifikasi. Dengan demikian, hasil analisis sentimen dapat menjadi lebih akurat dan relevan bagi pengambil keputusan.

PUSTAKA

- [1] D. Yuliani and S. Saryono and D. Apriani and M. Maghfiroh and M. Ro, "Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok (Sembako) Di Kecamatan Tambun Selatan Da-lam Masa Pandemi," *J. Citizsh. Virtues*, vol. 2, no. 2, pp. 320-326, 2022, doi:10.37640/jcv.v2i2.1533.
- [2] M. S. Andini, "Kenaikan Harga Minyak Mentah: Dongkrak atau Pasak APBN?," *Buletin APBN*, vol. VII, ed. 5, pp. 8-11, 2022.
- [3] kemenkeu.go.id, "Harga Minyak Mentah Terus Naik, Sementara Harga Jual Eceran Masih Dijaga Pemerintah," *Kemenkeu.Go.Id*, 2002, <https://www.kemenkeu.go.id/informasi-publik/publikasi/berita-utama/Harga-Minyak-Mentah-Terus-Naik> (accessed Apr. 29, 2023).
- [4] S. N. J. Fitriyyah and N. Safriadi and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279-285, 2019, doi:10.26418/jp.v5i3.34368.
- [5] C. M. Annur, "Pengguna Twitter di Indonesia Capai 24 Juta hingga Awal 2023, Peringkat Berapa di Dunia?," *Databoks*, 2023, <https://databoks.katadata.co.id/datapublish/2023/02/27/pengguna-twitter-di-indonesia-capai-24-juta-hingga-awal-2023-peringkat-berapa-di-dunia> (accessed Apr. 29, 2023).
- [6] Mujahidin and B. Prasetyo and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform)*, vol. 10, no. 3, pp.17-24, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [7] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, pp. 50–59, 2018, doi: 10.35314/isi.v3i1.335.
- [8] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [9] R. Amelia, D. Darmansah, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, pp. 338–343, 2022, doi: 10.30865/jurikom.v9i2.3895.
- [10] M. Yasid and L. Junaedi, "Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naive Bayes," *J. Ilm. Inform.*, vol. 7, no. 2, pp. 82–86, 2019, doi: 10.36802/jnaloka.2020.v1-no2-81-86.
- [11] M. Rifki and I. Imelda, "Analisis Sentimen Wacana Kenaikan Harga Tiket Candi Borobudur Menggunakan Multinomial Naive Bayes," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 156–163, 2022, doi: 10.33387/jiko.
- [12] V. Zuliana and G. Garno, and I. Maulana, "Analisis Sentimen Program Migrasi TV Digital Menggunakan Algoritma Naive Bayes Dengan Chi Square," *J. Inf. dan Komput.*, vol. 10, no. 2, pp. 90–95, 2022.
- [13] N. Hafidz and D. Y. Liliana, "Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 213–219, 2021, doi: 10.29207/resti.v5i2.2960.
- [14] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naive Bayes," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 267–281, 2022.
- [15] H. K. C. A. Pratama, W. Suharso, and Q. A'yun, "Pengklasifikasian Kanker Payudara Dan Kanker Paru-Paru Dengan Metode Gaussian Naive Bayes, Multinomial Naive Bayes, Dan Bernoulli Naive Bayes," *J. Smart Teknol.*, vol. 3, no. 4, pp. 350–355, 2022, [Online]. Avail-able: <http://jurnal.unmuhjember.ac.id/index.php/JST>.
- [16] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022, doi: 10.35870/jtik.v6i4.530.
- [17] D. Dano, "Analisis Dampak Konflik Rusia–Ukraina Terhadap Harga Bahan Bakar Minyak Indonesia," *CENDEKIA J. Ilmu Pengetah.*, vol. 2, no. 3, pp. 261–269, 2022, doi: 10.51878/cendekia.v2i3.1494.
- [18] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
- [19] W. Sugiarto and Y. Kristian and E. R. Setyaningsih, "Estimasi Arah Tatapan Mata Menggunakan Ensemble Convolutional Neural Network," *Teknika*, vol. 7, no. 2, pp. 94–101, 2018, doi: 10.34148/teknika.v7i2.126.