Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

Komparasi Naive Bayes dan Gradient Boosting Machine untuk Klasifikasi Sentimen Publik terkait Kenaikan Harga Pertalite di Media Sosial Twitter

Arnila Sandi[1] *, Yahya[2]
[1]Program Studi Teknik Komputer, Universitas Hamzanwadi
[2]Program Studi Sistem Informasi, Universitas Hamzanwadi
arnilasandi1878@gmail.com

Abstrak

Kenaikan harga bahan bakar pertalite di Indonesia telah menimbulkan berbagai reaksi dari masyarakat, yang banyak diungkapkan melalui media sosial seperti twitter. Bahan bakar minyak (BBM) merupakan salah satu kebutuhan dasar yang sangat penting bagi masyarakat Indonesia karena berperan besar dalam mendukung berbagai aktivitas sehari-hari. BBM digunakan tidak hanya sebagai bahan bakar untuk kendaraan bermotor, tetapi juga sebagai sumber energi bagi berbagai peralatan industri. Penelitian ini bertujuan untuk mengklasifikasi sentimen masyarakat terhadap isu tersebut dengan menggunakan algoritma Naive Bayes dan Gradient Boosting Machine (GBM) sebagai metode klasifikasi. Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yang berisi kumpulan tweet terkait isu harga bahan bakar khususnya pertalite. Proses analisis diawali dengan prapemrosesan teks, seperti pembersihan data, tokenisasi, stopword removal, dan stemming. Data tersebut sudah memiliki label sentimen (positif, negatif, netral) dan dibagi untuk pelatihan dan pengujian model. Hasil evaluasi menunjukkan bahwa algoritma GBM mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 60% sedangkan algoritma naive bayes memiliki tingkat akurasi sebesar 90%. hasil ini membuktikan bahwa antara naive bayes memiliki tingkat akurasi lebih tinggi dari algoritma GBM, sehingga dapat digunakan dalam mengolah data teks dari media sosial untuk memahami opini publik terhadap kebijakan pemerintah, khususnya terkait kenaikan harga bahan bakar.

Kata kunci : Analisis sentimen, GBM, klasifiaksi, naive bayes, pertalite, twitter.

Abstract

The increase in the price of pertalite fuel in Indonesia has caused various reactions from the public, which are widely expressed through social media such as twitter. Fuel oil (BBM) is one of the basic needs that is very important for the Indonesian people because it plays a major role in supporting various daily activities. BBM is used not only as fuel for motor vehicles, but also as a source of energy for various industrial equipment. This study aims to classify public sentiment towards the issue using the naive bayes algorithm and Gradient Boosting Machine (GBM) as a classification method. The data used in this study were obtained from the Kaggle platform, which contains a collection of tweets related to the issue of fuel prices, especially pertalite. The analysis process begins with text preprocessing, such as data cleaning, tokenization, stopword removal, and stemming. The data already has a sentiment label (positive, negative, neutral) and is divided for model training and testing. The evaluation results show that the GBM algorithm is able to classify sentiment with an accuracy rate of 60% while the Naive Bayes algorithm has an accuracy rate of 90%. These results prove that naive bayes has a higher level of accuracy than the GBM algorithm, so it can be used in processing text data from social media to understand public opinion on government policies, especially regarding fuel price increases.

Keywords : *Sentiment analysis, GBM, classification, naive bayes, pertalite, twitter.*

1. Pendahuluan

Kenaikan harga bahan bakar minyak di Indonesia menjadi isu yang viral di media sosial. Pada awal September 2022, pemerintah Indonesia mengumumkan penyesuaian harga terhadap tiga jenis bahan bakar bersubsidi, yaitu pertalite,

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

pertamax, dan solar. Kenaikan ini dipengaruhi oleh beberapa faktor utama, seperti tingginya harga minyak mentah global, meningkatnya beban subsidi, serta depresiasi nilai tukar Rupiah terhadap Dolar AS. Kebijakan pemerintah ini memicu reaksi dari masyarakat, yang disampaikan baik secara langsung maupun melalui media sosial[1].

Salah satu media sosial yang digunakan untuk menyampaikan reaksinya adalah twitter yang merupakan salah satu media komunikasi yang populer di kalangan masyarakat Indonesia, peningkatan jumlah pengguna serta cepatnya menyebabkan penyebaran informasi dan berita yang sedang tren menjadi lebih cepat.

Di Indonesia, twitter termasuk platform media sosial yang banyak digunakan secara aktif. Indonesia sendiri menempati posisi kelima sebagai negara dengan jumlah pengguna twitter terbanyak di dunia. Berdasarkan data dari *We Are Social*, jumlah pengguna twitter di Indonesia pada tahun 2022 mengalami peningkatan sebesar 31,3%, dari 14,05 juta pada tahun sebelumnya menjadi 18,45 juta. Angka ini mewakili sekitar 4,23% dari total 436 juta pengguna Twitter secara global [2].

Analisis sentimen yang dilakukan pada data twitter, dapat diketahui apakah mayoritas masyarakat Indonesia mendukung, menolak, atau bersikap netral terhadap kebijakan pemerintah tersebut. Analisis sentimen bisa disebut juga *opinion mining*, yang merupakan salah satu cabang dari klasifikasi teks dalam bidang pemrosesan bahasa alami, serta dapat digunakan untuk mengumpulkan dan mengelola berbagai respons publik yang berasal dari media sosial. Teknik ini bertujuan untuk memahami serta mengekstrak data opini secara otomatis dengan menentukan sentimen yang terkandung dalam suatu opini. Dengan penerapan metode *machine learning*, hasil analisis sentimen biasanya dikategorikan ke dalam tiga kelas utama, yaitu sentimen positif, netral, dan negatif [3].

Pada kasus kenaikan harga Pertalite ini, metode yang digunakan untuk menganalisis seluruh opini di media sosial twitter, adalah algoritma *Gradient Boosting Machine (GBM) dan* algoritma Naive Bayes. Naive Bayes adalah metode klasifikasi sederhana yang menghitung semua probabilitas berdasarkan teorema Bayes yang dikombinasikan dengan kombinasi nilai frekuensi basis data. Dalam tugas klasifikasi, tujuannya adalah untuk memprediksi label kelas dari sampel tertentu berdasarkan serangkaian fitur atau karakteristik. Metode ini digunakan karena sesuai dengan data yang ada yaitu data dalam bentuk teks [4][5].

Dua algoritma digunakan untuk mengetahui performa algoritma tersebut dalam menganalisis data teks, dengan mencari nilai akurasi, presisi, recall dan F1-Score. Oleh karena itu dari uraian tersebut maka penelitian ini diberi judul

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

"Komparasi Naive Bayes dan Gradient Boosting Machine untuk Klasifikasi Sentimen Publik terkait Kenaikan Harga Pertalite di Media Sosial Twitter"

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Adapun penelitian sebelumnya yang menggunakan algoritma Naïve Bayes antara lain:

- Laura Sari dan kawan-kawan yang berjudul *"Implementation of LightGBM and Random Forest in Potential Customer Classification".* Penelitian ini menggunakan metode LGBM dan Random Forest untuk mengklasifikasikan calon konsumen. Hasil penelitian ini menunjukkan bahwa pada data ketidakseimbangan metode LightGBM memiliki akurasi yang lebih baik dibandingkan dengan Random Forest yaitu sebesar 85,49%, ketika Random Forest tidak mampu menghasilkan model. Metode SMOTE yang digunakan dalam penelitian ini mempengaruhi akurasi random forest tetapi tidak mempengaruhi akurasi LightGBM. ecara keseluruhan nilai *Accuracy, Recall*, *Specificity,* dan *Precision*, Random Forest menghasilkan nilai yang baik dibandingkan dengan LightGBM pada data seimbang. Sementara itu LightGBM mampu menangani data tidak seimbang [6].
- Nenny Anggraeni dan kawan-kawan dalam penelitiannya yang berjudul *"A Comparative Analysis of Random Forest, XGboost, and LightGBM Algorithms for Emotion Classification in Reddit Comments"*. Penelitian ini memanfaatkan dataset GoEmotions Fine-Grained yang difilter hingga 7.325 komentar Reddit dengan 5 label emosi dasar yang berbeda. Dalam penelitian ini, dilakukan langkah-langkah praproses data, ekstraksi fitur menggunakan CountVectorizer dan TF-IDF, serta penyetelan hiperparameter menggunakan GridSearchCV untuk setiap algoritma. Selanjutnya, evaluasi model dilakukan menggunakan Cross-Validation dan matriks konfusi. Hasil penelitian menunjukkan bahwa Random Forest mengungguli algoritma XGBoost dan LightGBM dengan akurasi 75,38% dibandingkan dengan XGBoost dengan akurasi 69,05% dan LightGBM dengan akurasi 66,63% [7].
- Penelitian yang dilakukan oleh Ibnu Zahy dan rekan-rekan dengan judul "Implementasi Metode Klasifikasi LightGBM dan Analisis Survival dalam Memprediksi Pelanggan Churn" yang berisi tentang upaya mempertahankan pangsa pasar dengan melakukan pemodelan prediksi dan proyeksi pelanggan churn sebagai upaya retensi pelanggan, pada penelitian ini algoritma yang digunakan adalah kalasifikasi LightGBM, algoritma ini digunakan untuk prediksi churn

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

pelanggan dan memanfaatkan analisis survival untuk proyeksi masa mendatang. Hasil penelitian juga dimanfaatkan untuk mencegah churn pelanggan pada perusahaan, khususnya PT Kasir Pintar Internasional. Performa klasifikasi LightGBM yang diukur dengan evaluasi model mencapai nilai Accuracy, Precision, Recall, dan F1-score memiliki nilai masing-masing 0,964, 0,971, 0,990, dan 0,980. Model klasifikasi LightGBM memberikan informasi tentang lima fitur penting yang memengaruhi churn pelanggan. Perusahaan dapat menggunakan fitur tersebut sebagai bahan untuk merancang strategi retensi pelanggan. Selain itu, model survival Cox Proportional Hazard memiliki nilai evaluasi C-index sebesar 0,83 yang memiliki arti cukup mampu untuk memproyeksikan survival pelanggan. Model survival ini juga menunjukkan bahwa saat ini pelanggan yang tidak churn memiliki harapan survival rata-rata selama 15 bulan [8].

- Paulus Samotana Zalukhu dan kawan-kawan melakukan penelitian dengan judul Analisis Sentimen Terhadap Kenaikan Harga BBM di Indonesia pada Media Sosial Twitter menggunakan metode Naive Bayes". Hasil penelitian dijadikan acuan untuk melakukan penilaian terhadap kebijakan pemerintah dimata masyarakat. Penelitian dilakukan dengan 5000 data komentar twitter dan akan diolah menggunakan algoritma Naive Bayes. Nilai akurasi yang diperoleh adalah 75% dengan sentimen negatif yang mendominasi hasil pengolahan data, namun untuk keseluruhan rancangan dikatan baik[9].

2.2. Landasan Teori

1. Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) adalah teknik ensemble yang membangun model prediktif sebagai rangkaian model yang lebih lemah, biasanya pohon keputusan. Model-model ini dibangun secara bertahap, dengan setiap model baru mengoreksi kesalahan model sebelumnya[10].

2. Naive Bayes

Metode ini awalnya diperkenalkan oleh seorang ilmuwan asal Inggris, Thomas Bayes, yang menggunakannya untuk memperkirakan kemungkinan suatu kejadian di masa depan berdasarkan data atau pengalaman di masa lalu. Oleh karena itu, metode ini dikenal dengan nama Teorema Bayes. Naive Bayes sendiri merupakan salah satu metode klasifikasi yang sangat terkenal dan termasuk dalam daftar sepuluh algoritma terbaik dalam bidang data mining[11].

Naïve Bayes merupakan salah satu teknik yang berguna dalam mengelompokkan data ke dalam kategori yang sesuai. Dan sebagai metode statistik yang membantu memprediksi seberapa besar kemungkinan suatu data akan masuk ke

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

dalam kelas tertentu[12][13]. Naive Bayes adalah salah satu Algoritma sederhana yang menghitung probabilitas bersyarat untuk setiap kelas Keputusan dengan asumsi bahwa atribut dari sebuah objek saling independen. Algoritma ini juga memperkirakan kemungkinan-kemungkinan suatu objek ke dalam kelas tertentu berdasarkan nilai-nilai atribut yang dimiliki. Probabilitas dihitung dengan mengacu pada keadaan kelas yang sedang di evaluasi adalah benar. Frekuaensi kemunculan data dalam tabel Keputusan menjadi faktor yang utama untuk menentukan besarnya probabilitas akhir[14]

3. Klasifikasi

Klasifikasi merupakan proses pengelompokan data berdasarkan kriteria tertentu dan kesamaan dalam variabel-variabelnya, dengan tujuan membangun sebuah model yang mampu memprediksi kelas dari data baru yang belum diketahui sebelumnya. Proses ini bisa dilakukan secara manual maupun dengan memanfaatkan teknologi komputasi[15][16].

4. Text Mining

Text mining adalah proses mengekstraksi informasi dari sekumpulan dokumen, di mana pengguna berinteraksi dengan dokumen-dokumen tersebut menggunakan alat analisis yang merupakan bagian dari data mining, seperti klasifikasi teks dan pengelompokan teks. Data yang digunakan dalam text mining berasal dari pola-pola bahasa alami yang memungkinkan untuk menganalisis data dalam bentuk teks yang semi-terstruktur maupun tidak terstruktur[17].

3. Metode Penelitian

Penelitian ini bersifat penelitian kuantitatif komperatif karena menggunakan data numerik untuk hasil evaluasinya berupa akurasi, presisi dan F1-Score dan komperatif karena membandingkan dua kinerja algoritam yaitu Naive Bayes dan Gradient Boosting Machine (GBM) dalam mengklasifikasikan sentimen, serta eksperimen karena melatih kedua model dengan data yang sama dan uji kinerja keduanya. Sumber data berasal dari Kaggle dengan jenis data berbentuk teks. Metode penelitian ini terdiri dari beberapa tahapan, yaitu:

1. Teknik Pengumpulan Data

Data yang digunakan dari platform Kaggle dengan isu tentang kenaikan harga bahan bakar Pertalite dalam bentuk file .csv. Data Tweet yang dikumpulkan disaring berdasarkan bahasa Indonesia yang sesuai dengan topik/isu.

2. Pra-pemrosesan Data

Tahapan ini dilakukan untuk membersihkan dan mempersiapkan data teks agar dapat dianalisis lebih lanjut. Proses yang dilakukan meliputi:

- Case folding: Mengubah seluruh huruf menjadi huruf kecil.
- Tokenisasi: Memisahkan teks menjadi kata-kata tunggal.

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

- Stopword removal: Menghapus kata-kata umum yang tidak memiliki makna penting (misal: "yang", "dan", "di").
- Stemming: Mengubah kata ke bentuk dasar menggunakan algoritma stemmer bahasa Indonesia.
- Filtering: Menghapus karakter khusus, tautan, emoji, dan mention.

3. Labeling Sentimen

Data kemudian diberi label sentimen secara manual atau semi-otomatis dengan tiga kategori Positif, Negatif dan Netral.

4. Ekstraksi Fitur

Fitur teks dikonversi ke dalam bentuk numerik menggunakan metode TF-IDF(Term Frequency-Inverse Document Frequency) untuk merepresentasikan kata-kata yang penting dalam dokumen.

5. Penerapan Algoritma Klasifikasi

Dalam penelitian ini menggunakan dua algoritma pembelajaran mesin yaitu :

1. Naive Bayes (Multinomial Naive Bayes): Algoritma probabilistik sederhana yang mengasumsikan independensi antar fitur.
2. Gradient Boosting Machine (GBM): Model ensembel berbasis pohon keputusan yang membangun model prediksi secara bertahap.

Keduanya dilatih dengan data latih yang sama dan diuji pada data uji yang sama.

6. Evaluasi Kinerja

Model dievaluasi menggunakan metrik evaluasi klasifikasi akurasi, Presisi, Recall, dan F1-Score

7. Analisis dan Interpretasi

Hasil dari kedua model kemudian dibandingkan untuk mengetahui algoritma mana yang mampu memberikan performa terbaik dalam mengklasifikasikan sentimen publik terkait kenaikan harga Pertalite. Selain itu, dilakukan juga analisis tambahan untuk memahami jenis-jenis kesalahan klasifikasi yang sering terjadi serta melihat bagaimana persebaran sentimen publik secara keseluruhan, apakah lebih condong ke arah positif, negatif, atau netral.

4. Hasil dan Pembahasan

Algoritma Naïve Bayes dan Gradient Boosting Machine diterapkan untuk mengklasifikasikan sentimen publik terhadap kenaikan harga pertalite. Pengolahan data dilakukan menggunakan Google Colab dengan metode klasifikasi dengan algoritma Naïve Bayes dan GBM. Total data yang terkumpul adalah 511 komentar atau tweet.

4.1. Hasil Pra-Pemrosesan Data

Untuk penelitian ini dimulai dengan prapemrosesan data (preprocessing data) yaitu tahap pembersihan data dan data dinormalisasi untuk dihilangkan beberapa hal seperti: ulasan ganda (redundant data) dan ulasan kosong. Selain itu, proses pembersihan juga dilakukan

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

untuk setiap ulasan seperti penghilangan tanda baca, penghapusan kata yang tidak relevan (*stopword removal),* pengubahan kata dalam bentuk *lowercase*, dan mengubah kata menjadi kata dasar. Prapemrosesan data dimaksud untuk mempermudah penghitungan pada model algoritma yang di gunakan. Gambar 1 adalah tampilan dataset dalam bentuk CSV.

Gambar 1 Tampilan dataset dalam CSV

Tampilan komentar yang akan diproses terdiri dari 510 data tweet. Sampel data dapat dilihat pada gambar 2.

```
            DATE        PENGGUNA  \
0  6/16/2022 4:57  usahadarirumah
1  6/16/2022 4:53         p___a___u
2  6/16/2022 4:45       infokapuas
3  6/16/2022 4:25       MadechanS
4  6/16/2022 4:23    XinwenMetroTV
5  6/16/2022 4:16         detikcom
6  6/16/2022 4:16     detikfinance
7  6/16/2022 4:07       AingGunar
8  6/16/2022 4:06  abrakadabra678
9  6/16/2022 4:05       InforiauCo

                                          TWEET   LABEL
0  b'Harga Bahan Bakar Minyak Naik, Bensin Premiu...  POSITIF
1          b'@muaIes motor i juga pake pertalite weh'      NaN
2  b'Bupati Kapuas Kalteng Larang Kendaraan Dinas...  POSITIF
3  b'@jokowi Sak karepmu pak.....rakyat hanya bis...  NEGATIF
4  b'Untuk memastikan kecukupan pasokan dan distr...   NETRAL
5  b'Ini daftar harga Pertalite dan Pertamax hari...   NETRAL
6  b'Ini daftar harga Pertalite dan Pertamax hari...   NETRAL
7  b'@DraftAnakUnpad Aku sih kasih ucapan selamat...  POSITIF
```

Gambar 2 Tampilan data yang akan di proses

Beberapa tahapan utama dalam *preprocessing* meliputi:

1. Case Folding: Mengubah semua huruf menjadi huruf kecil agar seragam dan menghilangkan tanda baca seperti tanda petik (‘, “), koma (,), dan At (@). Misalnya, kata "b’Harga Bahan Bakar Minyak…." diubah menjadi "bharga bahan bakar minyak…". Contoh dari hasil dari proses Case Folding dapat dilihat pada gambar 3

```
                                              TWEET  \
0  b'Harga Bahan Bakar Minyak Naik, Bensin Premiu...
2  b'Bupati Kapuas Kalteng Larang Kendaraan Dinas...
3  b'@jokowi Sak karepmu pak.....rakyat hanya bis...
4  b'Untuk memastikan kecukupan pasokan dan distr...
5  b'Ini daftar harga Pertalite dan Pertamax hari...

                                      cleaned_tweet
0  bharga bahan bakar minyak naik bensin premium ...
2  bbupati kapuas kalteng larang kendaraan dinas ...
3  bjokowi sak karepmu pakrakyat hanya bisa berbu...
4  buntuk memastikan kecukupan pasokan dan distri...
5  bini daftar harga pertalite dan pertamax hari ...
```

Gambar 3 Hasil Case Folding

2. Tokenization: Memecah kalimat menjadi kata-kata atau token. Salah satu contoh proses tokenization pada kalimat sebagai berikut: “b'@jokowi Sak karepmu pak.....rakyat hanya bisa... “, maka hasil dari tokenization adalah ['b', "'", '@', 'jokowi', 'Sak', 'karepmu', 'pak', '.....', 'rakyat', 'hanya', 'bisa', '...'], proses hasil dapat dilihat pada gambar 4.

```
Token kata: ['b', "'", '@', 'jokowi', 'Sak', 'karepmu', 'pak', '.....', 'rakyat', 'hanya', 'bis', '...']
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
```

Gambar 4 Hasil Tokenization

3. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti contoh pada potongan kalimat berikut “b'@mazzini_gsp Kasus gini sangat memungkinkan, di tambah nih yee buat gue yg orang kampung ke pom gede ampir 2jam tuh jadi serba susah….”, maka hasil setelah melalui proses stopword removal menjadi seperti ini “b'@mazzini_gsp gini memungkinkan, nih yee gue yg orang

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

kampung pom gede ampir 2jam tuh serba susah....". berikut tampilan hasil proses stopword removal menggunakan code pada gambar 5.

```
Hasil setelah stopword removal:
b'@mazzini_gsp gini memungkinkan, nih yee gue yg orang kampung pom gede ampir 2jam tuh serba susah.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Gambar 5 Hasil *Stopword removal*

4. Normalization: Menyelaraskan kata yang memiliki bentuk berbeda tetapi makna yang sama menjadi bentuk baku. Contohnya kata "Yg", "tdk", "sdg" dan seterusnya maka akan dirubah menjadi bentuk baku seperti "yang", "tidak", dan "sedang". Diambil contoh dari kalimat berikut sebelum proses Normalization "Yg tidak punya android, tdk punya kuota, itu rerata orang yg membutuhkan pertalite'. Setelah mengalami Normalization akan menjadi "yang tidak punya android, tdk punya kuota, itu rata-rata orang yang membutuhkan pertalite'". Gambar 6 adalah hasil dari proses Normalization menggunakan python.

```
Teks asli:
Yg tidak punya android, tdk punya kuota, itu rerata orang yg membutuhkan pertalite'

Hasil setelah normalisasi:
yang tidak punya android, tdk punya kuota, itu rata-rata orang yang membutuhkan pertalite'
```

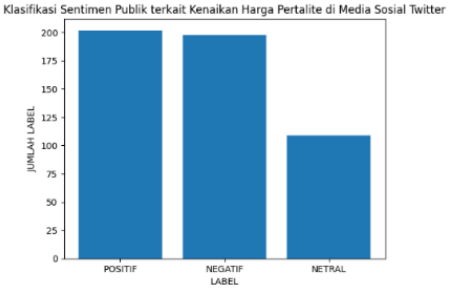
Gambar 6 Hasil *Normalization*

5. Stemming: Mengubah kata ke bentuk dasarnya. Contohnya, kata "membutuhkan", akan diubah menjadi "butuh". Agar terlihat jelas maka pada gambar 7 akan terlihat perubahannya.

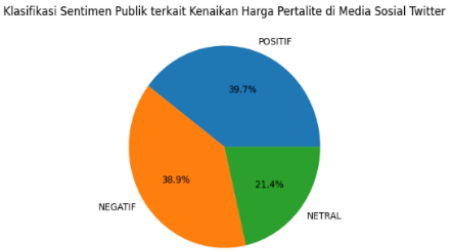
```
Teks asli:
Yg tidak punya android, tdk punya kuota, itu rerata orang yg membutuhkan pertalite.

Hasil setelah stemming:
yg tidak punya android tdk punya kuota itu rerata orang yg butuh pertalite
```

Gambar 7 Hasil *stemming*

Setelah pemrosesan data, maka hasil dari proses dapat dilihat dari tampilan visual berupa grafik batang dan lingkaran pada gambar 8 dan 9 secara berturut-turut.



Gambar 8 Tampilan Grafik Batang



Gambar 9 Tampilan Grafik Lingkaran

4.2. Pembagian Data

Setelah melalui tahap prapemrosesan, data kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji. Sebanyak 20% dari total dataset dialokasikan sebagai data uji, sedangkan sisanya digunakan untuk pelatihan model. Dengan demikian, terdapat 407 data untuk pelatihan dan 102 data untuk pengujian. Berikut gambar 10 untuk hasil pembagian datanya.

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

```
Jumlah data pelatihan: 407
Jumlah data pengujian: 102
Proporsi label di data pelatihan:
 LABEL
POSITIF    0.398034
NEGATIF    0.388206
NETRAL     0.213759
Name: proportion, dtype: float64
Proporsi label di data pengujian:
 LABEL
POSITIF    0.392157
NEGATIF    0.392157
NETRAL     0.215686
Name: proportion, dtype: float64
```

Gambar 10 Hasil pembagian data

Contoh hasil dari seluruh pemrosesan dapat dilihat pada gambar 11 berikut:

	DATE	PENGGUNA	TWEET	LABEL
0	6/16/2022 4:57	usahadarirumah	b'Harga Bahan Bakar Minyak Naik, Bensin Premiu...	POSITIF
2	6/16/2022 4:45	infokapuas	b'Bupati Kapuas Kalteng Larang Kendaraan Dinas...	POSITIF
3	6/16/2022 4:25	MadechanS	b'@jokowi Sak karepmu pak.....rakyat hanya bis...	NEGATIF

clean_TWEET	case_folding
bharga bahan bakar minyak bensin premium ga pe...	b'harga bahan bakar minyak naik, bensin premiu...
bbupati kapua kalteng larang kendaraan dina di...	b'bupati kapuas kalteng larang kendaraan dinas...
bjokowi sak karepmu pakrakyat berbuat apabiarl...	b'@jokowi sak karepmu pak.....rakyat hanya bis...

Gambar 11. Hasil pemrosesan data

4.3. Evaluasi

Untuk mengetahui performa dari dua algoritma yang digunakan maka kita melakukan evaluasi dari data yang ada, yaitu mencari nilai akurasi, presisi, recall dan F1-Score. Model yang pertama adalah dengan algoritma Gradient Boosting Machine dan hasil yang diperoleh dapat kita lihat pada gambar 12 adalah sebagai berikut :

```
Classification Report:
              precision    recall  f1-score   support

     NEGATIF       0.56      0.76      0.64        33
      NETRAL       0.67      0.32      0.43        25
     POSITIF       0.62      0.64      0.63        44

    accuracy                           0.60       102
   macro avg       0.61      0.57      0.57       102
weighted avg       0.61      0.60      0.58       102
```

Gambar 12. Gambar hasil evaluasi menggunakan Algoritma GBM

Dan untuk gambar 13 adalah hasil visualisasi menggunakan algoritma GBM.

Gambar 13. Tampilan visual dari hasil pemodelan GBM

Model selanjutnya menggunakan algoritma Naive Bayes dengan hasil 90% untuk nilai akurasi dan 90% utnuk Presisi dan 89% utnuk recall, hasil proses dapat dilihat pada gambar 14 berikut.

```
Hasil Prediksi Sentimen pada Dataset:
              precision    recall  f1-score   support

     NEGATIF       0.93      0.94      0.93       198
      NETRAL       0.86      0.83      0.84       109
     POSITIF       0.91      0.91      0.91       202

    accuracy                           0.90       509
   macro avg       0.90      0.89      0.89       509
weighted avg       0.90      0.90      0.90       509
```

Gambar 14. Gambar hasil evaluasi menggunakan Algoritma Naive Bayes

Sedangkan untuk tampilan visual dari model Naive Bayes dapat dilihat pada gambar 15 berikut:

Gambar 15. Tampilan visual dari hasil pemodelan Naive Bayes

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

Adapun adalah tabel perbandingan dari hasil Evaluasi dua algoritma tersebut adalah sebagai berikut:

Tabel 1. Perbandingan hasil evaluasi algoritma Gradient Boosting Machine dan Naive Bayes

Algoritma	Akurasi	Presisi	Recall	F1-Score
GBM	60 %	61%	57%	57%
Naive Bayes	90%	90%	89%	89%

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan, metode klasifikasi untuk sentimen publik terkait kenaikan harga pertalite di media sosial Twitter, untuk algoritma Naive Bayes mendapatkan nilai akurasi sebesar 90%, presisi 90%, recall 89% dan 89% untuk F1-Score. Sedangkan untuk algoritma GBM mendapatkan nilai Akurasi sebesar 60%, presisi 61%, recall 57% dan F1-Score 57%. Dari nilai akurasi yang diperoleh dapat disimpulkan bahwa masyarakat performance atau kinerja untuk algoritma Naïve Bayes sangat akurat digunakan untuk menganalisis sentimen publik terkait isu kenaikan harga Pertalite di media sosial Twitter.

6. Daftar Pustaka

[1] N. J. Hutagalung, Tony, "Analisis Sentimen Pada Opini Kenaikan Harga Bahan Bakar Minyak Pada Media Sosial Twitter," *J. Sist. Inf. DAN Tek. Komput.*, vol. 8, no. 2, pp. 280–284, 2023.

[2] Tiara Danirmala and Y. S. Nugroho, "Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1258–1268, 2023, doi: 10.33022/ijcs.v12i3.3199.

[3] B. Darmawan, A. Dwi Laksito, M. Resa, A. Yudianto, and A. Sidauruk, "Krea-TIF: Jurnal Teknik Informatika Analisis Perbandingan Ekstraksi Fitur Teks pada Sentimen Analisis Kenaikan Harga BBM," *J. Mhs. Inform.*, vol. 11, no. 1, pp. 53–63, 2023, doi: 10.32832/krea-tif.v11i1.13819.

[4] R. Harun, R. Ishak, and S. Panna, "Analisis Sentimen Opini Publik Pengguna Twitter Terhadap Kenaikan Harga BBM Menggunakan Algoritma Naïve Bayes," *J. Ilm. Ilmu Komput. Banthayo Lo Komput.*, vol. 2, no. 1, pp. 26–33, 2023, doi: 10.37195/balok.v2i1.414.

[5] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[6] L. Sari, A. Romadloni, R. Lityaningrum, and H. D. Hastuti, "Implementation of LightGBM and Random Forest in Potential Customer Classification," *TIERS Inf. Technol. J.*, vol. 4, no. 1, pp. 43–55, 2023, doi: 10.38043/tiers.v4i1.4355.

[7] N. Anggraini, S. J. Putra, L. K. Wardhani, F. D. U. Arif, N. Hakiem, and I. M. Shofi, "A Comparative Analysis of Random Forest, XGBoost, and LightGBM Algorithms for Emotion Classification in Reddit Comments," *J. Tek. Inform.*, vol. 17, no. 1, pp. 88–97, 2024, doi: 10.15408/jti.v17i1.38651.

[8] I. Zahy, A. Illah, W. Syaifullah, J. Sapu, and A. T. Damaliana, "Implementasi Metode Klasifikasi LightGBM dan Analisis Survival dalam Memprediksi Pelanggan Churn," vol. 8, no. 1, pp. 43–53, 2024.

[9] P. S. Zalukhu, T. Handhayani, and M. Sitorus, "Analisis Sentimen Terhadap Kenaikan Bbm Di Indonesia Pada Media

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.447-457

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31102 Link : https://dx.doi.org/10.29408/jit.v8i2.31102

Sosial Twitter Menggunakan Metode Naïve Bayes," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 1, pp. 65–69, 2023, doi: 10.51876/simtek.v8i1.177.

[10] J. Chaidir and P. Herwanto, "Sentiment Analysis Of Youtube on The Vina Cirebon Case," vol. 1, no. 2, pp. 98–106, 2024.

[11] M. Wasil, "Infotek : Jurnal Informatika dan Teknologi Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naive Bayes Pelayanan publik adalah kegiatan atau rangkaian kegiatan dalam ran," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 140–146, 2020.

[12] M. Ridho Handoko, "Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[13] M. Saiful1*, Syamsuddin2, and Moh. Farid Wajdi3, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 96–104, 2021.

[14] J. Sihombing, "Klasifikasi Data Antroprometri Individu Menggunakan Algoritma Naïve Bayes Classifier," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 1–10, 2021, doi: 10.37148/bios.v2i1.15.

[15] A. M. Nur, N. Nurhidayati, and I. Fathurrahman, "Penerapan Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP).," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 1, pp. 93–102, 2024, doi: 10.29408/jit.v7i1.23995.

[16] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, 2021, doi: 10.29408/jit.v4i1.2981.

[17] Rina Noviana and Isram Rasal, "Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter," *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.