

Analisis Sentimen Masyarakat Tentang Naiknya Harga BBM Dengan Metode Fasttext dan Naïve Bayes

Diterima:

10 Mei 2023

Revisi:

10 Juli 2023

Terbit:

1 Agustus 2023

^{1*}Dedy Surya Wijaya, ²Ardi Sanjaya, ³Wahyu Cahyo Utomo
¹⁻³Prodi Teknik Informatika, Fakultas Teknik, Universitas Nusantara
PGRI Kediri

Abstrak—Pada 3 September 2022, pemerintah menginformasikan melalui media sosial bahwa penyesuaian harga BBM dilakukan karena kenaikan harga minyak dunia dan subsidi BBM yang tidak tepat sasaran. Tidak sedikit masyarakat memberikan komentar tentang naiknya harga BBM dengan opini yang berbeda-beda sehingga sulit diklasifikasi. Berdasarkan uraian tersebut peneliti tertarik melakukan analisis sentimen masyarakat tentang naiknya harga BBM dengan metode vektorisasi Fasttext dan Gaussian Naïve Bayes yang bertujuan untuk mengklasifikasi komentar. Data komentar berasal dari komentar Youtube CNN Indonesia sebanyak 3000 komentar. Proses pemberian label dilakukan oleh sistem yang berdasar pada perhitungan nilai polarity menggunakan Vader Sentiment. Validasi k-fold dimana k=10 mendapat hasil terbaik dengan akurasi sebesar 73,67% dan data negatif lebih mendominasi model.

Kata Kunci—Analisis Sentimen;Fasttext;Naïve Bayes

Abstract— On September 3, 2022, the government informed through social media that fuel price adjustments were made due to rising world oil prices and fuel subsidies that were not well-targeted. Not a few people commented on the increase in fuel prices with different opinions so that it was difficult to classify. Based on this description, researchers are interested in analyzing public sentiment about rising fuel prices with the Fasttext vectorization method and Gaussian Naïve Bayes which aims to classify comments. Comment data comes from CNN Indonesia Youtube comments totaling 3000 comments. The labeling process is done by the system based on the calculation of the polarity value using Vader Sentiment. The k-fold validation where k = 10 gets the best results with an accuracy of 73.67% and negative data dominates the model.

Keywords—Sentiment Analisis;Fasttext;Naïve Bayes

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Dedy Surya Wijaya ,
Fakultas Teknik,
Universitas Nusantara PGRI Kediri,
Email: dedysurya2000@gmail.com

I. PENDAHULUAN

Pada 3 September 2022, Presiden Joko Widodo menginformasikan melalui media sosial bahwa penyesuaian harga BBM dilakukan karena kenaikan harga minyak dunia dan subsidi BBM yang tidak tepat sasaran. Dengan 204,7 pengguna internet di Indonesia dan 68,9% pengguna media sosial membuat informasi tersebut menyebar ke seluruh kalangan masyarakat Indonesia [1]. Tidak sedikit masyarakat juga menyampaikan pendapat dan opininya berupa komentar dengan akun media sosial yang dimiliki. Setiap komentar memiliki makna yang berbeda – beda membuat pandangan umum tentang naiknya harga BBM sulit diketahui. Berdasarkan permasalahan tersebut penulis tertarik untuk melakukan analisis sentimen masyarakat terhadap keputusan pemerintah untuk menaikkan harga BBM.

Analisis sentimen juga dikenal sebagai *opinion mining* adalah studi komputasi tentang pendapat, penilaian, dan perasaan orang lain yang terkandung dalam entitas, fakta, dan atribut yang dimiliki [2]. Proses analisis sentimen perlu dilakukan *text mining* dengan menggunakan NLP (*Natural Language Processing*) untuk menganalisa dan mendapat makna pada setiap kalimat komentar [3].

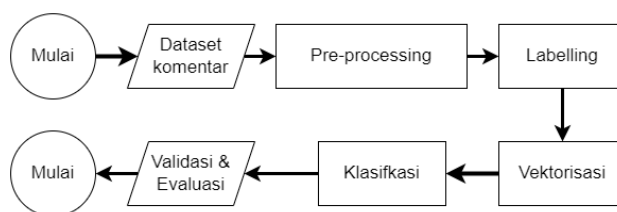
Ada banyak metode yang dapat digunakan untuk klasifikasi guna melakukan analisis sentimen, salah satunya adalah metode Naïve Bayes. Penelitian – penelitian analisis sentimen dengan metode Naïve Bayes dengan studi kasus tentang kebijakan pemerintah mendapatkan metode tersebut dapat melakukan klasifikasi dengan baik dan mendapatkan akurasi yang baik [3][4][5][6].

Dalam proses analisis sentimen perlu dilakukan vektorisasi atau mengubah sebuah kalimat kedalam bentuk vektor agar dapat dilakukan klasifikasi. Beberapa metode yang dapat digunakan adalah Word2vec, Glove, Stanford CoreNLP dan Fasttext. Beberapa penelitian dilakukan untuk membandingkan kinerja metode vektorisasi kalimat dan diantaranya menghasilkan metode Fasttext mendapatkan akurasi lebih baik dengan metode klasifikasi yang diajukan [7][8][9]. Dalam penelitian lain penggunaan Fasttext untuk vektorisasi kalimat review aplikasi MyPertamina menghasilkan Fasttext dapat digunakan untuk analisis sentimen [10].

Berdasarkan uraian pendahuluan dan penelitian sebelumnya diatas penulis mencoba untuk melakukan analisis sentimen masyarakat tentang naiknya harga BBM dengan algoritma Naïve Bayes dan Fasttext. Algoritma Naïve Bayes akan digunakan sebagai klasifikasi data menjadi sentimen positif atau negatif dan metode Fasttext akan digunakan untuk mengubah kalimat menjadi vektor.

II. METODE

Teknik penelitian yang digunakan dalam penelitian ini adalah teknik deskriptif kuantitatif. Penelitian kuantitatif adalah suatu proses mengumpulkan, mengolah, menganalisis, dan menyajikan data menggunakan perhitungan statistik, dan perhitungan matematika atau komputasi cerdas [11][12]. Prosedur penelitian yang digunakan adalah metode *waterfall* dimana memiliki kelebihan pengembangan dilakukan fase per fase sehingga mengurangi terjadinya kesalahan [13]. Metode pengumpulan data dilakukan dengan menggunakan metode kualitatif dengan teknik dokumen dimana dokumen dapat berupa tulisan, video, film dan gambar yang berisi informasi tentang proses penelitian dan dataset yang digunakan untuk inputan sistem [14]. Gambaran proses sistem dapat dilihat pada gambar 1.



Gambar 1. Diagram Alir Analisis Sentimen

2.1 Dataset

Dataset berupa komentar akan digunakan dalam penelitian. Dataset didapatkan dari proses *scraping* komentar video youtube CNN Indonesia yang memberitakan tentang pemerintah yang menaikkan harga BBM pada tanggal 3 September 2022.

2.2 Pre-processing

Teks *pre-processing* adalah proses membersihkan dan mengubah data teks yang tidak terstruktur menjadi format yang dapat dibaca mesin, membuat dokumen lebih konsisten dan memfasilitasi representasi teks dan penemuan pengetahuan [15]. Tahapan – tahapan yang dilakukan diantaranya.

1. *Translate emoticon* : tahap mengubah sebuah *emoticon* ke dalam bentuk kalimat.
2. *Cleaning text* : tahap menghapus *punctuation* dan mengubah huruf ke *lowercase*.
3. Fungsi terbilang : proses mengubah sebuah angka kedalam bentuk kalimat [16].
4. *Tokenizing* : proses memotong sebuah kalimat menjadi *array* kata.
5. Normalisasi : proses validasi kata yang tidak sesuai dengan kaidah bahasa Indonesia.
6. *Stemming* : proses mengubah sebuah kata yang diberi imbuhan menjadi kata dasar.

2.3 Labelling

Tahap *labelling* atau pemberian label pada komentar adalah tahap pengelompokan *polarity* yang dijadikan nilai pengetahuan untuk pemberian label. Penghitungan *polarity* menggunakan

metode *Vender Sentiment Lexicon Based*. Tujuan *labelling* adalah untuk menjadi nilai target dalam proses klasifikasi.

2.4 Fasttext

FastText adalah *library* untuk membuat representasi dan kata-kata yang efisien untuk mendukung klasifikasi teks *out-of-the-box* [17]. FastText menghasilkan pembagian parameter antara kata yang berbeda dengan n-gram membagi menjadi suku - suku kata untuk membuat penyematan untuk kata itu sendiri, ketika beberapa kata hilang dari kosa kata pelatihan atau jarang ada, masih bisa mewakili kata-kata itu jika n-gramnya ada sebagai bagian dari kata lain [18]. Tahapan proses yang harus dilakukan untuk vektorisasi yaitu dengan melatih corpus bahasa Indonesia selanjutnya dilakukan vektorisasi pada dataset.

2.5 Naïve Bayes

Naïve Bayes adalah sebuah algoritma yang dapat digunakan untuk pengklasifikasian untuk data *binary* dan *multiclass*. Klasifikasi akan menggunakan tipe Gaussian Naïve Bayes. Gaussian Naive Bayes dapat digunakan untuk memperkirakan distribusi data, tetapi Gaussian paling mudah digunakan karena hanya perlu memperkirakan rata – rata dan simpangan baku dari data pelatihan[19]. Perhitungan probabilitas dalam Gaussian Naïve Bayes dapat dirumuskan pada rumus (1).

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (1)$$

Dari rumus (1) dapat dijelaskan sebagai berikut:

$P(x_i|y)$ = Probability nilai kelas i terhadap kelas y

x_i = Nilai kelas i

σ = Variant

μ = Mean kelas y

π = 3.14

2.6 Validasi dan Evaluasi

Validasi dilakukan untuk mengetahui peforma dari sebuah model yang diajukan dimana tahap ini diuji validasi dengan *k- fold validation*. Tujuan dilakukanya *k-fold validation* adalah untuk mengeksplorasi data yang dilakukan beberapa kali agar mendapatkan estimasi akurasi yang dapat diandalkan [20]. Eksperimen menggunakan nilai k = 10 sehingga data kan dibagi menjadi 10 dengan perulangan klasifikasi sebanyak 10 kali. Untuk proses evaluasi akan dilakukan dengan menggunakan *confusion matrix*, akurasi, presisi dan *recall*.

III. HASIL DAN PEMBAHASAN

Scraping dilakukan menggunakan bahasa pemrograman Python dan menghasilkan 3000 komentar teratas. Data komentar selanjutnya disimpan dalam bentuk JSON dan akan dimasukkan dalam proses *Pre-processing*.

3.1 Pre-processing

Data komentar diubah kedalam bentuk *dataframe* lalu di masukan kedalam sistem *pre-processing*. Output dari proses ini menghasilkan data yang bersih dari *punctuation* tapi tidak mengurangi makna dari kalimat itu sendiri. Contoh hasil *pre-processing* dapat ditunjukkan pada tabel 1.

Tabel 1. Hasil *Pre-processing*

<i>Index</i>	<i>Text</i>	<i>pre-processing</i>
0	Turunkan tunjangan untuk DPR, alihkan untuk subsidi rakyat....	[turun,tunjang,untuk,dpr,alih, untuk,subsidi,rakyat,...]
1	Terimakasih pak de Saya seorang tukang ojek kampung...	[terimakasih,bapak,de,saya,orang,tukang, ojek,kampung...]
2	Saya sebagai rakyat kecil mendengar BBM naik sungguh sangat sedih pa...	[saya,bagai,rakyat,kecil,dengar, BBM,naik,sungguh,sangat,sedih, ...]
...
2997	saran saya daripada ngasih BLT yg tidak merata...	[saran,saya,daripada,ngasih,blt,yang, tidak,rata, ...]
2998	Presiden blusukan mantapp 😊	[presiden,blusukan,mantapp,wajah,dengan, air,mata,sukacita]
2999	Dari dulu hingga skarang sama harga bukn turun...	[dari,dulu,hingga,skarang,sama,harga ,bukn,turun...]

3.2 Labelling

Proses pelabelan dilakukan dengan menghitung *polarity* pada setiap komentar. Jika *polarity* kurang dari 0 maka diberi label negatif, jika polaritas lebih dari sama dengan 0 maka diberi label positif. Sebaran data hasil *labelling* adalah 2041 komentar berlabel negatif dan 959 komentar berlabel positif. Sebagai contoh pemberian label ditunjukkan pada tabel 2.

Tabel 2. Hasil *Labelling*

<i>Pre-processing</i>	<i>Polarity</i>	<i>Labels</i>
[turun, tunjang, untuk, dpr, alih, untuk, subsidi, rakyat, ...]	-0.0024	Negatif
[terimakasih, pak, de, saya, orang, tukang, ojek, kampung, ...]	-0.0011	Negatif
[saya, bagai, rakyat, kecil, dengar, BBM, naik, sungguh, sangat, sedih, ...]	0	Positif
...
[saran, saya, daripada, ngasih, blt, yg, tidak, merata, ...]	-0.0106	Negatif
[presiden, blusukan, mantapp, wajah,dengan, air,mata,sukacita]	0.1782	Positif
[dari, dulu, hingga, skarang, sama, harga, bukn, turun, ...]	0.0042	Positif

3.3 Vektorisasi Fasttext

Tahap vektorisasi dilakukan dengan dua tahap yaitu tahap *training* model untuk mendapatkan model Fasttext dan selanjutnya model akan digunakan untuk membuat vektor pada kalimat. Data yang digunakan untuk membuat model Fasttext adalah komentar yang sudah dilakukan *cleaning*. Vektorisasi akan dilakukan pada komentar yang sudah dilakukan seluruh proses *pre-processing*.

3.4 Klasifikasi Gaussian Naïve Bayes

Label akan dilakukan *encoding* menjadi bentuk numerik agar dapat mempermudah proses klasifikasi. Model klasifikasi dibuat dengan melatih data komentar yang sudah berbentuk vektor dan testing dilakukan untuk mengetahui performa dari model yang sudah dibuat.

3.5 Validasi dan Evaluasi

Proses pembagian data menggunakan *k-fold* dengan k=10 menghasilkan data terbagi menjadi 10 dimana pengujian dilakukan sebanyak 10 kali dan rasio jumlah data 9:1. Akurasi dari 10 kali pengujian dapat dilihat pada tabel 3.

Tabel 3. Validasi dengan *k-fold*

<i>K-fold</i>	Akurasi
1	70%
2	66%
3	65,3%
4	72,67%
5	70,67%
6	61,67%
7	66%
8	70,67%
9	64%
10	73,67%

Dari tabel 3 dapat ditetapkan nilai *k-fold* yang terbaik adalah k = 10 dimana menghasilkan nilai akurasi tertinggi sebesar 73,6%. Dari hasil sebelumnya dapat ditunjukkan evaluasi pada nilai k=10.

```

akurasi = 0.7366666666666667
      precision  recall  f1-score  support
0         0.76     0.92     0.83      212
1         0.60     0.30     0.40       88

accuracy                0.74      300
macro avg             0.68     0.61     0.61      300
weighted avg          0.71     0.74     0.70      300
```

Gambar 2. *Screenshot* hasil prediksi

Berdasarkan gambar 2, data negatif lebih mendominasi dengan nilai *precision* 0.76, *recall* 0.92 dan *f1-score* 0.83. Data label positif mendapatkan nilai *precision* 0.60 dan hanya mendapatkan nilai *recall* dan *f1-score* sebesar 0.30 dan 0.40 saja.

IV. KESIMPULAN

Dari hasil penelitian yang sudah tentang analisis sentimen tentang naiknya harga BBM dapat disimpulkan metode Fasttext dapat mengubah sebuah kalimat kedalam sebuah vektor meskipun kalimat menggunakan kata – kata diluar kaidah bahasa dan algoritma Naive Bayes dapat mengklasifikasi sebuah vektor dengan dimensi 100. Validasi dengan *k-fold* mendapatkan hasil yang bagus untuk evaluasi untuk data yang banyak.

Adapun kekurangan dalam penelitian ini yaitu komentar banyak yang menggunakan sindiran sehingga proses *labelling* belum menjangkau sebuah makna sebenarnya dalam kalimat. Hal lain yang perlu diperhatikan adalah tahap *pre-processing* karena kalimat komentar cenderung menggunakan kata-kata yang tidak sesuai kaidah Bahasa Indonesia. Agar penelitian selanjutnya mendapatkan hasil yang lebih baik disarankan menggunakan sistem pengenalan makna kalimat dan pendekatan tahap *pre-processing* lebih ditekankan.

DAFTAR PUSTAKA

- [1] S. Kemp, “Digital 2022: Indonesia — DataReportal – Global Digital Insights,” *Global Digital Insights*. pp. 1–103, 2022. Accessed: Oct. 24, 2022. [Online]. Available: <https://datareportal.com/reports/digital-2022-indonesia>
- [2] F. A. Nugraha, N. H. Harai, and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif Industri Nusantara, 2020. Accessed: Nov. 19, 2022. [Online].
- [3] S. Mujahidin, B. Prasetyo, and M. C. C. Utomo, “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes,” *J. Vocat. Tek. Elektron. dan Inform.*, vol. 10, no. 3, pp. 17–24, 2022, [Online].
- [4] M. Al Khadafi, K. P. Kartika, and F. Febrinita, “PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN LEXICON BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS,” *J. Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 725–733, 2022.
- [5] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [6] C. F. Hasri and D. Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector

- Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online].
- [7] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, “Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks,” *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.
- [8] Siti Khomsah, Rima Dias Ramadhani, and Sena Wijaya, “The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 3, pp. 352–358, 2022, doi: 10.29207/resti.v6i3.3711.
- [9] D. M. K. Oshadi and S. Thelijjagoda, “AppGuider: Feature Comparison System using Neural Network with FastText and Aspect-based Sentiment Analysis on Play Store User Reviews,” *3rd Int. Conf. Smart Electron. Commun. ICOSEC 2022 - Proc.*, pp. 1148–1155, 2022, doi: 10.1109/ICOSEC54921.2022.9952093.
- [10] E. Budianita, M. Fikry, and F. Yanto, “Klasifikasi Sentiment Review Aplikasi MyPertamina Menggunakan Word Embedding FastText dan SVM (Support Vector Machine),” *J. Sist. Komput. dan Inform.*, vol. 4, pp. 526–534, 2023, doi: 10.30865/json.v4i3.5695.
- [11] M. Ramdhan, *Metode Penelitian*. Surabaya, 2021. Accessed: Nov. 05, 2022. [Online].
- [12] D. Nicolaus, *Metodologi Penelitian Kuantitatif: Beberapa Konsep Dasar Untuk Penulisan Skripsi & Analisis Data dengan SPSS*. 2019. Accessed: Nov. 05, 2022. [Online].
- [13] T. Sanubari, P. Cahyo, and N. Riza, “Odol (one desa one product unggulan online) penerapan metode Naive Bayes pada pengembangan aplikasi e-commerce menggunakan Codeigniter,” *Kreatif*, 2020.
- [14] S. Siyoto and M. A. Sodik, “Dasar Metodologi Penelitian.” Literasi Media Publishing, Yogyakarta, pp. 68–70, 2015. Accessed: Nov. 05, 2022. [Online].
- [15] A. Beheshti, S. Ghodratnama, M. Elahi, and H. Farhood, *Social Data Analytics*. CRC press, 2022. doi: 10.1201/9781003260141.
- [16] A. Sanjaya and S. D. Sasongko, “UJI KEMIRIPAN KALIMAT MENGGUNAKAN FUNGSI TERBILANG PADA PRE-PROCESSING DAN COSINE SIMILARITY DALAM BAHASA INDONESIA,” *J. Ilm. Nero*, vol. 7, no. 2, pp. 95–104, 2022.
- [17] J. Bhattacharjee, *fastText Quick Start Guide: Get started with Facebook’s library for text representation and classification*. Packt Publishing, 2018.
- [18] A. Kedia and M. Rasu, *Hands-On - Python Natural Language Processing*. 2020.
- [19] J. Brownlee, “Master Machine Learning Algorithms: Discover how they work and

- implement them from scratch.,” *Mach. Learn. Mastery*, pp. 1–163, 2016, [Online].
- [20] T. T. Wong and P. Y. Yeh, “Reliable Accuracy Estimates from k-Fold Cross Validation,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020, doi: 10.1109/TKDE.2019.2912815.