

# ANALISIS SENTIMEN TERHADAP KENAIKAN BBM DI INDONESIA PADA MEDIA SOSIAL TWITTER MENGUNAKAN METODE NAÏVE BAYES

Paulus Samotana Zalukhu<sup>\*1</sup>, Teny Handhayani<sup>2</sup>, Manatap Sitorus<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, FTI, Universitas Tarumanagara

Jl. Letjen S Parman No 1, Jakarta 11440 Indonesia

e-mail: <sup>1</sup>*\*<sup>1</sup>paulus.535180150@stu.untar.ac.id*, <sup>2</sup>*tenyh@fti.untar.ac.id*, <sup>3</sup>*manataps@fti.untar.ac.id*

*Baru-baru ini masyarakat Indonesia sedang dibuat heboh dengan kebijakan yang akan memangkas subsidi pada harga BBM yang beredar. Hal ini mengakibatkan naiknya harga BBM dan dikhawatirkan akan berdampak pada sektor lainnya. Pada zaman sekarang, masyarakat dapat memberikan opini pada situasi yang sedang terjadi menggunakan media sosial, dimana nantinya diharapkan opini mereka dapat didengar. Untuk itu, analisis sentiment masyarakat terhadap kenaikan harga BBM di Indonesia perlu untuk dilakukan. Hasil analisis ini nantinya diharapkan dapat menjadi bahan acuan untuk penilai dari penerimaan masyarakat terhadap kebijakan pemerintah. Sentimen analisis ini mengambil 5000 data berupa komentar masyarakat terhadap isu kenaikan BBM dan akan diolah oleh Machine Learning menggunakan algoritma Naïve Bayes.*

**Kata Kunci:** Algoritma Naïve Bayes, Kenaikan BBM, Analisis Sentimen

## I. PENDAHULUAN

Terkait dengan membengkaknya beban subsidi dan kompensasi yang dibayarkan negara untuk tahun 2022 mengakibatkan negara perlu mengambil kebijakan dalam menguranginya. Untuk itu pemerintah mengeluarkan kebijakan yaitu berupa pengurangan subsidi Bahan Bakar Minyak (BBM). Hal ini mengundang pro dan kontra ditengah tengah masyarakat, dimana kondisi ekonomi masyarakat yang masih belum begitu pulih pasca dilanda Covid-19 dan juga kenaikan harga pangan, sekarang harus dihadapkan dengan naiknya harga BBM. Opini dan pendapat masyarakat yang begitu banyak terhadap kebijakan dari pemerintah terkait naiknya harga BBM tersebut dalam hal ini terkhusus untuk platform twitter, maka perlu dilakukan suatu analisis yang dapat dijadikan acuan dan tolak ukur dalam penentuan kebijakan mendatang terkait hal tersebut. Pada perancangan sistemnya sentimen masyarakat terkait kenaikan harga BBM di Indonesia pada Twitter dengan pemodelannya

akan menggunakan Naive Bayes. Naive Bayes merupakan pendekatan statistik yang fundamental dalam mengenal pola (Pattern Recognition). Pendekatan ini di dasarkan pada kualifikasi trade-off antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang di timbulkan dalam keputusan-keputusan tersebut. Algoritma naïve bayes digunakan pada perancangan ini karena Naïve Bayes Classifier memiliki kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi. Dikarenakan sifatnya yang sederhana, sehingga dalam melakukan pengkodean menggunakan Bahasa pemrograman juga sederhana. Metode Naïve Bayes Classifier untuk klasifikasi teks menggunakan atribut kata yang muncul dalam suatu dokumen sebagai dasar klasifikasinya. Penelitian Rish (2001) menunjukkan bahwa meskipun asumsi independensi antar kata dalam dokumen tidak sepenuhnya dapat dipenuhi, tetapi kinerja Naïve Bayes Classifier dalam klasifikasi relative sangat bagus, selain itu algoritma naïve bayes adalah algoritma yang sering digunakan dalam melakukan analisis sentiment sehingga dalam hal ini memudahkan penulis dalam mencari referensi. Pada perancangan ini akan dijabarkan tahapan – tahapan yang akan dilalui untuk melakukan pemrosesan analisis sentimen terhadap ulasan – ulasan yang ada pada Twitter terkait kenaikan harga BBM. Tahapan tersebut akan diawali dengan tahapan preprocessing, tahap analisis menggunakan metode Naïve Bayes Classifier sampai tahap visualisasi hasil analisa.

## II. METODE

Pada perancangan sistem, diperlukan metode yang menjadi acuan dan pedoman untuk diimplementasikan pada rancangan yang dibuat. Berikut teori yang digunakan antara lain Analisis Sentimen, Algoritma Naïve Bayes, TF-IDF, Confusion Matrix, dan pengambilan data.

### A. Analisis Sentimen

Analisis sentimen adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual dilakukan untuk melihat pendapat terhadap

sebuah masalah, atau untuk identifikasi kecenderungan hal di pasar. Pada era global saat ini, internet merupakan salah satu kebutuhan yang tidak dapat dipisahkan dari masyarakat. Setiap hari bahkan setiap detiknya terdapat informasi baru yang ditambahkan pada internet. Bersama dengan konten atau informasi yang masuk tersebut, komentar dan opini pengguna juga turut ikut kedalamannya. Analisis sentimen memungkinkan kita untuk dapat mengetahui persentase dukungan atau penolakan masyarakat terhadap suatu topik.

**B. Algoritma Naïve Bayes**

Algoritma Naive Bayes merupakan salah satu algoritma machine learning yang banyak digunakan untuk klasifikasi. Naive Bayes adalah sekumpulan algoritma yang didasarkan pada Teorema Bayes. Naive Bayes adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian. Naive Bayes Classifier merupakan salah satu algoritma yang sederhana namun memiliki kemampuan dan akurasi yang tinggi dan termasuk dalam metode machine learning. Ada dua tahap dalam NBC untuk proses klasifikasinya, yaitu training dan testing. Tahap pertama adalah pelatihan terhadap sentimen yang sudah diketahui kelasnya untuk membangun sebuah model probabilistik. Sedangkan tahap kedua adalah proses klasifikasi sentimen yang belum diketahui kebenaran dari kelasnya. Teorema Bayes menemukan probabilitas atau kemungkinan suatu peristiwa akan terjadi dengan memberikan probabilitas peristiwa lain yang telah terjadi. Teorema Bayes dinyatakan secara matematis dalam persamaan berikut :

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

Dimana  $P(B) \neq 0$

Keterangan:

- B : data dengan class yang belum diketahui
- A : hipotesis data berupa suatu class spesifik
- $P(A | B)$  : probabilitas hipotesis A berdasar kondisi B
- $P(A)$  : probabilitas hipotesis A
- $P(B | A)$  : probabilita B berdasarkan kondisi pada hipotesis A
- $P(A)$  : probabilitas A

Setelah itu hitung himpunan kelas dengan rumus sebagai:

$$V_{map} = \underset{v_j \in V}{argmax} \frac{P(a_1 \dots a_n | v_j) P(v_j)}{P(a_1 \dots a_n)}$$

Setelah diperoleh perhitungan, maka kelas yang didapat adalah kelas dengan nilai tertinggi. Nilai  $P(V_j)$  ditentukan berdasarkan persamaan berikut :

$$P(V_j) = \frac{|banyak\ kata\ dalam\ kelas|}{|banyak\ kata|}$$

$$P(W_k | V_j) = \frac{n_k + 1}{n + |kata|}$$

Keterangan:

- $P(W_k|V_j)$  : probabilitas kata  $W_k$  muncul dengan kelas  $V_j$
- $n_k$  : frekuensi kata  $W_k$  muncul dengan kelas  $V_j$
- $n$  : banyak kata pada kelas  $V_j$
- $|kata|$  : banyak kata pada data

**C. TF-IDF (Terms Frequency-Inverse Document Frequency)**

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Pada rancangan ini TF-IDF berfungsi sebagai metode dalam pembobotan kata sehingga nantinya dapat diolah oleh naïve bayes. Pada dasarnya, TF-IDF bekerja dalam menentukan frekuensi relatif suatu kata kemudian dibandingkan dengan proporsi kata tersebut pada seluruh dokumen (TF). Intuisinya bahwa kata yang muncul di banyak dokumen bukanlah pembeda yang baik, dan harus diberi bobot kurang dari satu yang terjadi dalam beberapa dokumen (IDF). Pembobotan TF-IDF adalah hasil perkalian dari pembobotan term frequency dan inverse document frequency dari suatu term. TF-IDF dihitung dengan menggunakan persamaan seperti berikut:

$$W_{i,j} = \frac{n_{i,j}}{\sum_{j=1}^p n_{j,i}} \log_2 \frac{D}{d_j}$$

Keterangan

- $W_{i,j}$  : Pembobotan tf-idf untuk term ke-j pada dokumen-i
- $n_{i,j}$  : Jumlah Kemunculan untuk term-j pada dokumen ke-i
- $p$  : Banyaknya term yang terbentuk
- $\sum_{j=1}^p n_{j,i}$  : Jumlah kemunculan seluruh term pada dokumen ke-i
- $D$  : Jumlah keseluruhan dokumen
- $d_j$  : Banyaknya Dokumen yang mengandung term ke-j

**D. Confusion Matrix**

Pada rancangan ini confusion matrix digunakan sebagai metode evaluasi dari akurasi yang dirancang. Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan tingkat akurasi pada data mining. Confusion matrix memuat informasi tentang klasifikasi yang diprediksi dengan benar oleh sebuah sistem klasifikasi. Terdapat empat istilah sebagai representasi hasil proses klasifikasi pada confusion matrix diantaranya adalah True Positif (TP), True Negatif (TN), False Positif (FP) dan False Negative (FN) seperti yang dapat dilihat pada tabel berikut:

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive)
	0 (Negative)	FN (False Negative)	TN (True Negative)

Gambar. 1 Confusion Matrix



3. Dari data yang dihasilkan, apakah anda dapat memahami hasil dari analisis?
4. Apakah hasil analisis tersebut merupakan data yang informatif?
5. Setelah melihat hasil analisis, apakah anda setuju bahwa dampak kenaikan bbm adalah negative terhadap masyarakat?
6. Apakah anda setuju jika kenaikan harga BBM akan berefek domino pada harga kebutuhan lain?
7. Berikan saran dan tanggapan anda terhadap aplikasi yang digunakan.

#### IV. KESIMPULAN DAN SARAN

Setelah proses pengujian yang dilakukan pada Bab IV terdapat kesimpulan yang dapat dirangkum menjadi beberapa poin yakni sebagai berikut:

1. Penggunaan Algoritma Naïve Bayes sebagai metode klasifikasi memiliki tingkat akurasi sebesar 75%. kurangnya akurasi yang dicapai pada rancangan ini disebabkan karena kendala ketidakseimbangan data yang didapat dimana sentiment negative lebih mendominasi, yang jika perbandingannya dibuat dalam bentuk angka dalah 2:1:1. Namun untuk keseluruhan rancangan sudah dapat dikatakan baik.
2. Pengujian metode perhitungan dengan menggunakan confusion matrix sangat dianjurkan untuk mengetahui keakuratan system.
3. Berdasarkan hasil pengujian user acceptance test terhadap 7 responden dapat dinyatakan bahwa aplikasi ini merupakan aplikasi yang baik dan dapat nyaman digunakan oleh user serta memenuhi tujuan dari pembuatan aplikasi ini sendiri.

Saran yang dapat diberikan untuk pengembangan dan perbaikan rancangan ini adalah:

1. Proses klasifikasi Naïve Bayes memang dapat digunakan pada analisis sentiment terhadap kenaikan bbm di Indonesia, tetapi dikarenakan tidak seimbangnya nilai sentiment, dimana sentimen negative lebih mendominasi maka dibutuhkan suatu cara penyeimbang untuk digunakan sebagai data latih.
2. Secara keseluruhan proses yang dijalankan memakan waktu yang cukup lama untuk dapat menampilkan hasil analisis. Hal ini disebabkan oleh proses yang berjalan cukup berat untuk pengambilan data real time dan penjalanan beberapa library yang tidak bersifat open source. Maka dari itu diharapkan untuk melakukan proses lebih lanjut untuk dapat meminimalisir waktu pemrosesan data. Hal ini juga berpengaruh pada proses hosting yang mana data yang dapat digunakan dalam web hosting hanya sejumlah 100 data agar tidak terkena timeout Hal ini menunjukkan bahwa hanya dengan menggunakan aplikasi latihan FPS yaitu Aim Lab, tidak berpengaruh secara keseluruhan terhadap performa di ingame Valorant dan dibutuhkannya latihan tambahan seperti sparing dan latihan lainnya.

#### DAFTAR PUSTAKA

- [1] Abipraya, D. "Implementasi Opinion Mining Untuk Provider Internet Menggunakan
- [2] Metode Naive Bayes Classification", Jurnal Ilmu Komputer Dan Sistem Informasi, Vol. 9 No. 2 (2021)
- [3] Cholissodin, I. Buku Ajar AI, Machine Learning & Deep Learning.
- [4] <https://www.researchgate.net/publication/348003841>. diakses pada 12 September 2022
- [5] Giovani A.P, Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi (<https://ejurnal.teknokrat.ac.id/2020>) Diakses pada 9 September 2022.
- [6] Ginting, H. S. "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes", e-Proceeding of Engineering : Vol.5, (2018).
- [7] Gumilang, Z. A. Implementasi Naïve Bayes Classifier Dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee Pada Situs Google Play (Yogyakarta: <https://dspace.uui.ac.id>) diakses pada 17 september 2022
- [8] Gunawan Billy, Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes (<https://jurnal.untan.ac.id/>, 2018) diakses pada 9 September 2022.
- [9] Harishamzah. perbandingan-perhitungan-bobot-tf-idf-secara-manual-dan-menggunakan-Python(<https://medium.com/bisa-ai/perbandingan-perhitungan-bobot-tf-idf-secara-manual-dan-menggunakan-python377392a165c6>) diakses pada 15 September 2022.
- [10] Hidayatulah, Buku Ajar Interface User Experience (Yogyakarta: Universitas Pembangunan Nasional "Veteran".2020) hlm.5
- [11] Ipmawati. "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen".
- [12] Indonesian Journal on Networking and Security - Volume 6 No 1.(2017).
- [13] Kalokasari D.H. "Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar (Studi Kasus: Diskominfo Kabupaten Tangerang)". JURNAL Teknik Informatika Vol.10 No.2,(2017).
- [14] Mansur Januar," Analisis Pendapat Publik Terhadap Public Figure Dengan Menggunakan Metode Naive Bayes". Jurnal Ilmu Komputer dan Sistem Informasi Vol. 7 No. 2 (2019).
- [15] Mushthofa. Informatika untuk SMA Kelas X. Jakarta: Pusat Perbukuan Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2021.
- [16] Natalius. Metode Naive Bayes Classifier dan Penggunaanya Pada Klasifikasi Dokumen.(<https://informatika.stei.itb.ac.id/~rinaldi.munir/Probstat/2010-2011/Makalah2010/MakalahProbstat2010-017.pdf>) diakses pada 16 September 2022.
- [17] Nurjannah Musfiroh, Penerapan Algoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining,(<http://e-ourrnals.unmul.ac.id/index.php/JIM/article/view/113>, 2013) diakses pada 15 September 2022.
- [18] Ramos, J. (2003). Using Tf-idf to Determine Word Relevance in Document Queries. <https://www.semanticscholar.org/>. Diakses pada 15 September 2022.
- [19] Robertson, S. (2004). "Understanding inverse document frequency: on theoretical arguments for IDF". Journal of Documentation, Vol. 60 No. 5,(2004).

- [20] Rosari M.A, "Analisis Sentimen Tanggapan Masyarakat Terhadap bantuan Sosial pemerintah Di Masa Pandemi Covid-19 Pada Platform Twitter". Jurnal Ilmu Komputer Dan Sistem Informasi Vol. 10 No. 1 (2022).
- [21] Sepri, D. "Penerapan Algoritma Naïve Bayes Untuk Analisis Kepuasan Penggunaan Aplikasi bank". Journal of Computer System and Informatics (JoSYC), VOL 2 NO 1, 135-139. (2020).
- [22] Sipayung E.M, Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier (<https://ejournal.unsri.ac.id/2016>) diakses pada 9 September 2022.
- [23] StatistikaKomputasi,seri-bayesian-untuk-pemula -teorema-bayes. (<https://statistikakomputasi.wordpress.com/2010/03/27/seri-bayesianuntuk-pemula-teorema-bayes/> ) 17 september 2022