

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER MENGENAI KEBIJAKAN KENAIKAN HARGA BAHAN BAKAR MINYAK MENGUNAKAN METODE NAIVE BAYES

Bambang Tri Buwono¹, Nurhafifah Matondang²
S1 Informatika / Fakultas Ilmu Komputer

Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450
bambangtb@upnvj.ac.id¹, nurhafifahmatondang@upnvj.ac.id²

Abstrak. Twitter adalah media sosial yang sering dimanfaatkan masyarakat untuk menyampaikan aspirasi, opini, atau keluhan mengenai kebijakan pemerintah dalam membuat suatu keputusan. Dalam penelitian ini dilakukan proses klasifikasi data tweet yang mengandung sentimen positif dan sentimen negatif mengenai kebijakan kenaikan harga bahan bakar minyak (BBM). Metode yang digunakan untuk proses klasifikasi pada penelitian ini dengan menggunakan metode Naïve Bayes. Data tweet diambil dengan menggunakan metode scraping. Data tweet akan diberikan label positif dan negatif untuk mempermudah data tersebut diproses. Tweet yang diambil adalah tweet yang menyertakan @KenaikanhargaBBM sebagai kata kunci. Saat data tweet sudah terkumpul data akan dibagi menjadi dua bagian yaitu, data latih 80% dan data uji 20%. Hasil evaluasi tersebut memiliki fungsi untuk menguji performa dari metode yang sudah dilakukan untuk mengklasifikasikan tweet kedalam kelas positif dan negatif menunjukkan nilai akurasi yang didapatkan sebesar 89%.

Kata Kunci: Twitter, Kenaikan Harga Bahan Bakar, Klasifikasi, Naïve Bayes

1. Pendahuluan

Twitter adalah salah satu media sosial berbasis microblogging dimana pengguna dapat mengirim pesan yang disebut *tweet*. Twitter adalah situs microblogging yang dioperasikan oleh Twitter, Inc. Blog mikro karena Situs ini memungkinkan pengguna untuk memposting dan membaca pesan seperti blog pada umumnya. Pesan yang disebut *tweets*, yaitu 140 karakter teks tertulis yang ditampilkan di halaman profil pengguna. Menggunakan twitter sering diakses oleh pengguna Indonesia mencapai 56% dan tempat ke-5 di media sering digunakan di jejaring sosial pada tahun 2020.

Bahan Bakar Minyak (BBM) memiliki peran yang sangat penting dalam kehidupan masyarakat. BBM merupakan salah satu kebutuhan pokok bagi masyarakat kota maupun desa, baik sebagai rumah tangga, pedagang, bahkan perusahaan, demikian juga BBM sangat memiliki peran penting disektor industri maupun transportasi. Kenaikan harga BBM bagi kehidupan masyarakat menimbulkan fenomena adanya pro dan kontra di kalangan masyarakat maupun berbagai media. Karena dampak dari perubahan harga BBM ini mempengaruhi harga produksi, distribusi, transportasi sehingga berpengaruh juga pada harga barang yang lainnya. Kebutuhan bahan makanan pokok pun juga terpengaruh dengan adanya kenaikan harga BBM, seperti beras, gula, dan minyak goreng.

Analisis sentimen atau penggalian opini adalah proses atau langkah memahami, memperoleh, dan memproses data teks otomatis untuk mengambil informasi sentimen yang terkandung dalam kalimat opini. Tren penelitian analisis sentimen berfokus pada opini yang menyalurkan sentimen positif atau negatif. Dalam proses pengklasifikasian kelas positif dan negatif, penulis menggunakan salah satu metode klasifikasi yaitu *Naive Bayes* (NB).

Algoritma Naive Bayes (NB) yang digunakan untuk klasifikasi sentimen pada penelitian ini diharapkan dapat memberikan model klasifikasi yang baik, sehingga informasi yang diperoleh dari hasil model klasifikasi tersebut kredibel dan dapat digunakan untuk analisis sentiment terhadap reaksi masyarakat Indonesia dalam hal ini para warganet pengguna twitter tentang Kenaikan Harga Bahan Bakar Minyak (BBM).

2. Metode Penelitian

2.1 Analisis Sentimen

Analisis sentimen atau penggalian opini adalah proses atau langkah memahami, memperoleh, dan memproses data teks otomatis untuk mengambil informasi sentimen yang terkandung dalam kalimat opini. dan dapat ditentukan sebagai sentimen positif, dan negatif. Menurut C. B. Saputra, A. Muzakir, & D. Udariansyah, (2019)[1], Analisis sentimen juga dapat dianggap sebagai penambangan opini. Analisis sentimen dapat digunakan di berbagai bidang yang memungkinkan, mulai dari produk konsumen, layanan kesehatan, layanan ekonomi, peristiwa sosial dan politik hingga pemilihan umum. Tren penelitian analisis sentimen berfokus pada opini yang menyalurkan atau menyiratkan sentimen positif atau negatif.

2.2 Text Mining

Text Mining adalah tahap proses analisis data dalam bentuk teks dari mana sumber data diekstraksi. dari sebuah dokumen, seperti word data kata, konsep text mining sering digunakan untuk mengklasifikasikan dokumen teks, di mana dokumen diklasifikasikan menurut subjek dokumen menurut B. R. Feldman (2013)[2]. Teks mining adalah penambangan yang dilakukan oleh komputer untuk mengekstrak beberapa pesan, sesuatu yang sebelumnya tidak diketahui, atau untuk menemukan informasi yang berarti diekstraksi dari informasi yang dihasilkan secara otomatis dari berbagai sumber data teks.

2.3 Praproses Data

Pra proses data merupakan tahapan awal yang dilakukan untuk mengolah data yang masih memiliki noise menjadi data yang siap untuk proses klasifikasi. Tujuan pada tahap ini untuk mengambil inti atau pokok dari sebuah teks, sehingga menjadi data yang siap dan tepat untuk proses klasifikasi.

a) Case Folding

Proses yang mengubah setiap huruf menjadi huruf kecil secara keseluruhan atau lowercase, semua huruf yang akan diklasifikasikan diubah menjadi lowercase karena mayoritas *tweet* pada umumnya lebih banyak menggunakan huruf yang campur antara huruf besar dan kecil.

b) Cleaning Data

Proses menghapus semua karakter yang tidak memiliki value kecuali huruf alfabet. Proses yang dimaksud seperti menghapus hastag, URL, username, dan emoticon

c) Stopword Removal

Proses pembuangan kata yang tidak memiliki makna atau stopword. Proses penghapusan stopword tergantung dari daftar kamus stopword yang digunakan. Contoh stopword yang sering muncul adalah “yang”, “pada”, “dari”, dan lain-lain.

d) Stemming

Proses mengubah kata yang memiliki imbuhan menjadi sebuah kata dasar. Dalam proses pengklasifikasian teks, banyak kata yang ditemukan memiliki imbuhan awalan atau akhiran dari sebuah kata. Penghapusan imbuhan pada penelitian ini dibantu dengan library sastrawi.

e) Tokenisasi

Proses pemotongan string input yang dilakukan pada setiap kata yang ada dalam penyusunannya supaya memudahkan dalam membedakan karakter

2.4 Pembobotan TF-IDF

Menurut Zakiyuddin (2020) [3] pembobotan TF-IDF merupakan metode untuk memberikan bobot hubungan sebuah kata terhadap suatu dokumen yang memiliki dua konsep, yaitu frekuensi kemunculan kata pada sebuah dokumen dan kebalikan frekuensi dokumen yang mengandung kata tersebut.

$$TF - IDF = TF_{t,d} \times IDF_t = TF_{t,d} \times \log \frac{|N|}{|DF_t|} \quad (1)$$

Keterangan:

$TF_{t,d}$: Jumlah frekuensi yang memiliki term t pada dokumen d

IDF_t : Jumlah inverse frekuensi dokumen pada term t

N : Jumlah seluruh dokumen yang digunakan

DF_t : Jumlah frekuensi dokumen yang disimpan pada term t

$$w_{i,j} = tf_{i,j} \log \left(\frac{N+1}{df_i+1} \right) + 1 \quad (2)$$

Keterangan:

w_{i,j} = bobot dokumen ke-i terhadap kata ke-j

tf_{i,j} = banyaknya kata i yang dicari pada sebuah dokumen j

N = total dokumen

df_i = banyaknya dokumen yang mengandung kata ke-i

2.5 Naïve Bayes

Menurut Retno (2019) [4] metode Naïve Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi atau mengatur peluang keanggotaan dalam kelas. Keuntungan untuk metode Naive Bayes, metode ini hanya membutuhkan data pelatihan data sedikit untuk menentukan kontroversi dibutuhkan dalam proses klasifikasi.

Metode Naïve Bayes membutuhkan dua set, yaitu Set Training dan set testing. Satu Set Training digunakan sebagai data training yang akan digunakan sebagai pengenalan objek yang akan diklasifikasikan, sedangkan set testing data yang digunakan sebagai alat ukur akurasi klasifikasi ini. Berikut pernyataan dari Naive Bayes:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (3)$$

Keterangan:

C : Hipotesis data merupakan suatu kelas spesifik

B : Data dengan kelas yang belum diketahui.

P(A|B) : Posterior, probabilitas hipotesis A tergantung kondisi B.

P(A) : Prior, probabilitas hipotesis A

P(B|A) : Likelihood, probabilitas B tergantung kondisi pada hipotesis A.

P(B) : Evidence, probabilitas B.

Rumus untuk menghitung probabilitas pada masing masing kelas, seperti persamaan dibawah:

$$P(c) = \frac{|doc\ c|}{|document|} \quad (4)$$

Keterangan :

P(c) : Peluang kemunculan dokumen yang memiliki kategori j.

doc c : Jumlah dari dokumen untuk tiap kategori j.

|document| : Jumlah dokumen dari setiap kategori.

Proses klasifikasi data testing menggunakan seperti berikut:

$$C_{MAP} = \operatorname{argmax}_{c \in V} P(A) \times P(X_i|A) \quad (5)$$

Keterangan :

P(A) : Peluang kemunculan suatu dokumen yang memiliki kelas A.

P(X_i|A) : Peluang kemunculan Xi pada kelas A.

Rumus untuk menentukan nilai probabilitas setiap kata dari dokumen dapat dideskripsikan gunakan persamaan sebagai berikut:

$$P(A) = \frac{\text{Count}(W_{i,A}) + 1}{|A| + |V|} \quad (6)$$

Keterangan :

- P(A) : Peluang W pada kelas A
- Count(W_{i,A}) : Jumlah term W yang ditemukan di kelas A
- |A| : jumlah term di seluruh di kelas A
- |V| : Jumlah keseluruhan term pada kelas V

2.6 Evaluasi

Menurut (Retno, 2019) [4] pengukuran akurasi diukur dengan confusion matrix. Dapat diartikan bahwa confusion matrix merupakan metode untuk menghitung akurasi pada konsep data mining. Hasil akhir dari metode ini akan memberikan suatu informasi dengan membandingkan hasil klasifikasi yang dijalankan oleh sistem dengan data sebenarnya. Berikut merupakan tabel yang mengilustrasikan hasil *confusion matrix* dari suatu kelas prediksi.

Tabel 1 Confusion Matrix

		True Class	
		Positif	Negatif
Predicted Class	Positif	True Positif(TP)	False Positif(FP)
	Negatif	False Negatif(FN)	True Negatif (TN)

Keterangan :

True Positive (TP) : data yang diprediksi benar dan realitanya benar

True Negatif (TN) : data yang diprediksi salah dan realitanya salah.

False Positive (FP) : data yang diprediksi benar dan realitanya salah.

False Negatif (FN) : data yang diprediksi salah dan realitanya benar.

a) Persamaan untuk menghitung nilai accuracy:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \quad (7)$$

b) Persamaan untuk menghitung nilai precision:

$$\text{precision} = \frac{TP}{TP + FP} \quad (8)$$

c) Persamaan untuk menghitung nilai recall:

$$\text{recall} = \frac{TP}{TP + FN} \quad (9)$$

d) Persamaan untuk menghitung nilai specificity:

$$specificity = \frac{TN}{TN+FP} \quad (10)$$

3. Hasil dan Pembahasan

Penelitian ini menggunakan data yang berasal dari media sosial twitter dengan menggunakan kata kunci pencarian yaitu #Kenaikan Harga BBM. Data *tweet* yang digunakan merupakan *tweet* yang diunggah pada media sosial twitter pada tanggal 21 April 2022 hingga 16 Juni 2022. Jumlah data yang diperoleh kata sebanyak 16742 kata dari 1000 dokumen *tweet*.

Data yang sudah dikumpulkan akan diberikan label secara manual dengan menggunakan 3 annotator, hasil pelabelan dapat dilihat pada table 2 sebagai berikut :

Data	Label 1	Label 2	Label 3	Hasil Akhir
Hari ini Komisi VI dijadwalkan RDP dengan Pertamina. Salah satu agenda: Pembahasan kenaikan harga BBM	Positif	Positif	Positif	Positif
Betul !!! Kenaikan harga BBM dunia tdk diikuti oleh penetapan kenaikan harga oleh Pertamina. Shg harga BBM ttt yg dipasarkan kpd rkyt Indonesia adl harga yg sdh dibantu pmrnth mll subsidi. Dan namanya subsidi dpt diartikan scra kasar sbg "menggerogoti" APBN	Positif	Negatif	Negatif	Positif

Setelah melakukan pelabelan data, data akan melalui praproses data untuk menghilangkan noise yang ada pada data. Hasil praproses data dapat dilihat pada Tabel 3.

Data sebelum praproses data	Data sesudah praproses data	Label
Hari ini Komisi VI dijadwalkan RDP dengan Pertamina. Salah satu agenda: Pembahasan kenaikan harga BBM	komisi vi jadwal rdp pertamina salah agenda bahas naik harga bbm	Positif
@energitodayID @pertamina @KemenBUMN Pertamina ikut memikul beban apbn agar harga BBM masih stabil gak ada kenaikan 🙏🙏🙏	pertamina pikul beban apbn harga bbm stabil gak naik	Positif

Selanjutnya melakukan Pembobotan TF-IDF terhadap setiap kata yang ada pad tweet dari data yang digunakan pada penelitian ini sejumlah 1000 data.

Tabel 2 Perhitungan Pembobotan TF-IDF

Term	Dokumen				DF	IDF	TF-IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
Harga	1	2	1	1	4	0	0	0	0	0
Bbm	1	1	1	1	4	0	0	0	0	0
Pertamina	1	0	1	0	2	0,301	0,301	0	0,301	0

Standar	1	0	0	0	1	0,602	0,602	0	0	0
Alami	1	0	0	0	1	0,602	0,602	0	0	0
Naik	1	1	1	0	3	0,200	0,200	0,200	0,200	0
Lambung	0	1	0	0	1	0,602	0	0,602	0	0
Minyak	0	1	0	0	1	0,602	0	0,602	0	0
Mentah	0	1	0	0	1	0,602	0	0,602	0	0
Dunia	0	1	0	0	1	0,602	0	0,602	0	0
Pengaruh	0	0	1	0	1	0,602	0	0	0,602	0
Masyarakat	0	0	1	0	1	0,602	0	0	0,602	0
Bantu	0	0	1	0	1	0,602	0	0	0,602	0
Mahal	0	0	0	1	1	0,602	0	0	0	0,602
Sebab	0	0	0	1	1	0,602	0	0	0	0,602
Tiket	0	0	0	1	1	0,602	0	0	0	0,602

Setelah sudah mendapatkan hasil nilai bobot pada term di dokumen, maka selanjutnya melakukan tahap metode klasifikasi dengan menggunakan naïve bayes. Proses klasifikasi naïve bayes dilakukan dapat dua cara langkah proses, yaitu data training dan data testing. Hal ini dimana data yang sudah dikumpulkan dari total 1000 data, maka perlu proses perbandingan data. Kemudian diperbandingkan terdahulu otomatis nanti nya akan dibagi secara acak ke dalam data training dan data testing. Data training mendapatkan persentase sebesar 80% sedangkan data testing mendapatkan persentase sebesar 20% dari 100%. Berikut ini merupakan hasil pembagian data pada data training dari 80% dan data testing dari 20% seperti dideskripsikan pada Tabel 4.10 tersebut.

Tabel 3 Hasil Pembagian Data

	Label Positif	Label Negatif	Total Jumlah
Data Training	713	87	800
Data Testing	178	22	200
Total Jumlah	891	109	1000

Dengan menggunakan multinomial naïve bayes hasil klasifikasi yang dapat dilihat pada Tabel 6 berupa confusion matrix yang akan digunakan untuk menghitung performa dari algoritma tersebut.

Tabel 4 Hasil Confusion Matrix

		True Class	
		Positif	Negatif
Predicted Class	Positif	178(TP)	22(FP)
	Negatif	0(FN)	0(TN)

Tabel 4.15 adalah hasil *confusion matrix* yang menunjukkan beberapa hasil dari model bentuk dengan menggunakan algoritma *Naïve Bayes* yang sudah dihasilkan *tersebut* :

- Data 178 yang akan diprediksi masuk ke dalam kategori kelas positif maka nilai prediksi benar(*True positive*)
- Data 22 yang akan diprediksi masuk ke dalam kategori kelas positif maka nilai prediksi benar(*True Negatif*)
- Data 0 yang akan diprediksi masuk ke dalam kategori kelas positif maka nilai prediksi benar(*False positive*)
- Data 0 yang akan diprediksi masuk ke dalam kategori kelas positif maka nilai prediksi benar(*False Negatif*)

Dari hasil pengujian model tersebut, maka dapat melakukan menghitung performa model *naïve bayes* dengan menggunakan rumus *confision matrix pada persamaan (2.7), (2.8), (2.9), dan (2.10)* pada BAB 2 sebagai berikut.

- Hitung nilai *accuracy* :

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} = \frac{178 + 0}{178 + 0 + 22 + 0} = 0,89$$

- Hitung nilai *precision* :

$$Precision = \frac{TP}{TP + FP} = \frac{178}{178 + 22} = 0,89$$

- Hitung nilai *recall* :

$$recall = \frac{TP}{TP + FN} = \frac{178}{178 + 0} = 1$$

- Hitung nilai *Specificity* :

$$Specificity = \frac{TN}{TN + FP} = \frac{0}{0 + 22} = 0$$

Berdasarkan hasil performa dengan menggunakan *naïve bayes* telah didapatkan dengan nilai akurasi sebesar 89%, nilai *precision* 89%, nilai *recall* mendapatkan sebesar 100%, dan nilai *specificity* mendapatkan nilai sebesar 0%

4. Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan pada hasil penelitian Analisis Sentimen Pada Media Sosial Twitter Mengenai Kebijakan Kenaikan Harga Bahan Bakar Minyak Menggunakan Metode Naïve Bayes, maka dapat menarik kesimpulan sebagai berikut :

- a) Berdasarkan hasil penelitian yang dilakukan dengan menggunakan metode naïve bayes dalam melakukan klasifikasi tweet positif dan negatif mengenai kebijakan kenaikan bahan bakar minyak, dapat disimpulkan bahwa masyarakat pengguna twitter mendominasi sentimen positif dibandingkan sentimen negatif, karena masyarakat masih memiliki kepercayaan kepada usaha yang sudah dilakukan pemerintah untuk masyarakat.
- b) Hasil performa menggunakan algoritma naïve bayes dalam klasifikasi mendapatkan hasil performa dengan nilai akurasi sebesar 89%, hasil precision 89%, hasil recall senilai 100%, dan hasil specificity mendapatkan nilai sebesar 0%

4.2 Saran

Saran yang diperoleh dalam penelitian ini untuk pengembangan supaya kedepannya dapat menjadi hasil yang lebih baik lagi, yaitu :

- a) Untuk penelitian selanjutnya dapat menambahkan algoritma yang lainnya untuk sebagai perbandingan performa dengan algoritma naïve bayes yang sudah dilakukan dalam penelitian ini.

Referensi

- [1] B. R. Feldman, "Techniques and Applications for Sentimen Analysis," *Commun. Acm*, vol. 58, no. 4, pp. 83–89, 2013.
- [2] Saputra, C. B., Muzakir, A., & Udariansyah, D. (2019). Analisis Sentimen Masyarakat Terhadap #2019gantipresiden Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier. In *Bina Darma Conference on Computer Science (BDCCS)* (Vol. 1, No. 2, pp. 403-413)
- [3] Sari, R. (2019). Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization. *Jurnal Informatika*, 6(1), 23-28.
- [4] Zakiyuddin, H. (2021). Penerapan Algoritma Cosine Similarity Dan Pembobotan TF-IDF Pada System Penerimaan Mahasiswa Baru Universitas Bumigoram Mataram (Doctoral dissertation, UNIVERSITAS BUMIGORA)[