

Analisis Perbandingan Ekstraksi Fitur Teks pada Sentimen Analisis Kenaikan Harga BBM

Briga Darmawan¹, Arif Dwi Laksito^{2*}, Muhammad Resa Arif Yudianto³, Acihmah Sidauruk⁴

¹Program Studi Informatika, Universitas Amikom Yogyakarta, Indonesia

^{2,4}Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia

³Fakultas Teknik, Universitas Muhammadiyah Magelang, Indonesia

E-mail koresponden: arif.laksito@amikom.ac.id

Diserahkan 31 Maret 2023; Direview 27 Juli 2023; Dipublikasikan 30 Agustus 2023

Abstrak

BBM merupakan bahan bakar yang digunakan kendaraan bermotor. Penggunaan BBM meningkat sejalan dengan pertumbuhan ekonomi di Indonesia. Kenaikan harga BBM di Indonesia menimbulkan berbagai macam pendapat di media sosial twitter melalui posting dan thread. Fokus penelitian ini melakukan analisis sentimen terhadap kenaikan BBM yang datanya didapat melalui twitter dengan jumlah 1667 data. Tujuan dari penelitian ini melakukan perbandingan metode ekstraksi fitur yang memiliki kinerja paling baik seperti TF-IDF, Bag of Word, dan FastText diuji dengan algoritma machine learning SVM. Untuk tahap penelitian yang pertama melakukan crawling data twitter, preprocessing data, ekstraksi fitur, pembuatan model dengan algoritma machine learning, dan kemudian dilakukan pengujian dan perbandingan model confusion matrix pada setiap ekstraksi fitur. Hasil dari penelitian ini menunjukkan bahwa penggunaan ekstraksi fitur BoW memiliki kinerja lebih baik dibandingkan model ekstraksi fitur yang lain.

Kata kunci: Analisis Sentimen, BoW, FastText, SVM, TF-IDF.

Abstract

BBM is the fuel used by motorized vehicles. The use of fuel is increasing in line with economic growth in Indonesia. The fuel price hike in Indonesia generated a wide range of opinions on social media twitter through posts and threads. This study focuses on carrying out a sentiment analysis of the increase in fuel prices, the data obtained via Twitter with a total of 1667 data. This study aims to compare feature extraction methods with the best performance, such as TF-IDF, Bag of Word, and FastText, evaluated with the SVM machine learning algorithm. The study starts with crawling Twitter data, pre-processing, feature extraction, developing machine learning, and then evaluating and comparing the confusion matrix model for each feature extraction. The results of this study indicate that the use of BoW feature extraction has better performance than other feature extraction models.

Keywords: BoW, FastText, Sentiment Analysis, SVM, TF-IDF.

PENDAHULUAN

BBM adalah singkatan dari Bahan Bakar Minyak. Istilah ini muncul karena bahan bakar yang sering digunakan kendaraan bermotor. BBM merupakan hasil dari perut bumi kemudian diolah menjadi berbagai macam produk salah satunya BBM [1]. Penggunaan BBM akan mengalami peningkatan sejalan dengan pertumbuhan ekonomi di Indonesia. Akibat kenaikan harga minyak di seluruh dunia membuat harga BBM di Indonesia mengalami peningkatan [2]. Hal tersebut menimbulkan opini dari seluruh masyarakat baik yang mendukung, menolak dan tidak keduanya.

Twitter merupakan media sosial yang lumayan populer bagi para penggunanya untuk memberikan komentar [3] karena penggunaan aplikasi yang mudah. Melalui *posting* dan *thread* twitter bisa mendapatkan berbagai macam tanggapan tentang masalah apapun. Postingan twitter dapat digunakan sebagai analisis pendapat dan opini terutama tanggapan tentang kenaikan BBM. Analisis sentimen yang dilakukan melalui twitter dapat mengetahui sebagian besar masyarakat di Indonesia mendukung, menolak atau tidak peduli terhadap kebijakan pemerintah tentang kenaikan BBM tersebut.

Analisis sentimen atau *opinion mining* merupakan bagian dari *text classification* yang mempunyai bidang bahasa alami yang luas [4][21]. Biasanya analisis sentimen digunakan untuk mengelola dan mengumpulkan berbagai macam tanggapan publik melalui media sosial seperti facebook, twitter, instagram. Analisis sentimen digunakan untuk mengetahui, memahami, dan mengekstrak [5] data opini yang secara otomatis menentukan sentimen yang ada pada sebuah opini. Melalui analisis sentimen dengan menggunakan metode *machine learning*, pendapat tersebut dibagi ke dalam tiga bagian yaitu pendapat positif, netral, dan negatif.

Machine learning memiliki hubungan dengan *artificial intelligence* untuk belajar dari data di masa lalu dengan mengembangkan teknik-teknik yang diprogramkan, karena *machine learning* merupakan alat yang digunakan untuk analisis dalam suatu *data mining* [6]. *Machine learning* dapat diklasifikasikan ke dalam dua kategori [7] yaitu *supervised machine learning* dan *unsupervised machine learning*. *Supervised machine learning* dalam analisis sentimen terdapat beberapa algoritma salah satunya adalah algoritma *Support Vector Machine* (SVM) yang memiliki akurasi tinggi untuk melakukan sentimen karena kemampuannya dalam menemukan *hyperplane* paling baik dan penggunaannya yang banyak dikembangkan saat ini [8]. Selain itu proses mendapatkan sentimen tidak terlepas dari tahap ekstraksi. Tahap ekstraksi fitur termasuk dalam pemrosesan *word embedding* pada suatu sentimen yang bertujuan untuk mengubah informasi teks menjadi *vector* yang nyata. Perkembangan *word embedding* dimulai dengan terbentuknya teknik *bag of word* yang merepresentasikan kata dengan frekuensi kata yang sering muncul pada kalimat. Peneliti Karen Sparck Jones tahun 1972 menemukan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) yang merupakan gabungan dari TF (*Term Frequency*) dan IDF (*inverse Document Frequency*) yang mengandalkan setiap statistik dari kata-kata yang ada pada dokumen [9][22]. Kemudian pada tahun 2017 facebook mengembangkan teknik untuk mempelajari representasi kata yang banyak digunakan untuk NLP dengan nama *FastText* yang bertujuan untuk menganalisis kata melalui struktur internal, teknik representasi kata ini sangat efektif karena mempelajari kata-kata secara mandiri [10].

Penelitian-penelitian sebelumnya terkait analisis sentimen telah melakukan pemodelan menggunakan pembelajaran mesin (*machine learning*). Penelitian analisis sentimen menggunakan satu algoritma yaitu SVM diterapkan pada *dataset* yang diambil dari yahoo tentang pasar saham untuk menentukan sentimen (positif, negatif, netral) [11]. Penelitian lain melakukan perbandingan algoritma antara SVM dengan *K-Nearest Neighbor* (K-NN) pada

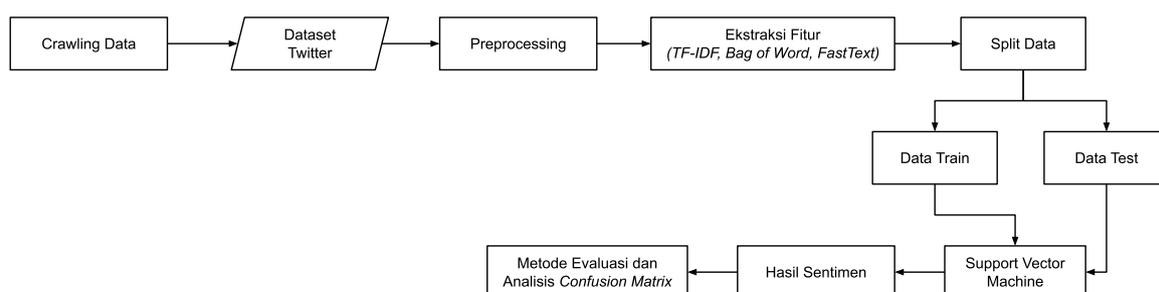
dataset yang diambil dari twitter [8]. Dari hasil kedua penelitian penggunaan algoritma SVM memiliki tingkat akurasi yang baik sehingga penelitian ini menggunakan algoritma SVM dengan tujuan mendapatkan kinerja yang lebih unggul dalam mendapatkan hasil sentimen.

Selain membandingkan algoritma *machine learning*, beberapa penelitian melakukan perbandingan metode ekstraksi fitur seperti TF-IDF, BoW, dan FastText. Penggunaan metode SVM dengan TF-IDF memiliki hasil yang paling baik dengan rata-rata F1-Score sebesar 74,6% pada dataset *University Social Media* [12]. Penelitian dengan *dataset feedback e-learning* berbahasa indonesia memiliki hasil yang sama bahwa penggunaan ekstraksi fitur TF-IDF memiliki akurasi yang lebih baik [13]. Penelitian lainnya mengenai ekstraksi fitur ialah melakukan perbandingan dengan berbagai algoritma. Perbandingan metode ekstraksi fitur BoW, TF-IDF, dan FastText yang dikombinasikan dengan algoritma klasifikasi yang populer seperti Naïve Bayes, SVM, dan XGBoost pada *dataset spam comment* instagram [14]. Hasil penelitian menunjukkan bahwa penggunaan FastText yang digabung dengan algoritma SVM memiliki nilai f1-score yang lebih baik dibandingkan dengan kombinasi algoritma yang lain. Perbandingan metode ekstraksi fitur yang populer yaitu BoW, N-Gram, TF-IDF, Word2Vec, Glove, dan FastText kemudian dilakukan kombinasi algoritma CNN, NB, SVM, ID3, dan C4.5 pada *dataset opinion mining* twitter. Hasil penelitian menunjukkan bahwa ekstraktor yang paling efisien dalam hal akurasi, presisi, recall, dan F1 adalah CNN+FastText yang memiliki hasil lebih baik dari yang lain [14][15]. Banyaknya jenis ekstraksi fitur yang bisa digunakan untuk menganalisis kenaikan harga BBM, membuat penelitian ini memerlukan referensi penggunaan ekstraksi fitur yang paling efektif. Jenis ekstraksi fitur mana yang memiliki hasil paling efisien dalam perhitungan dan perubahan kata menjadi vektor.

Penelitian ini bertujuan menganalisis tentang kenaikan harga BBM untuk melihat apakah masyarakat cenderung memberikan tanggapan publik positif, netral, atau negatif akibat dari kenaikan harga BBM. Hasil analisis didapatkan dari membandingkan ekstraksi fitur yang memiliki performa paling baik berdasarkan data twitter pada kenaikan BBM dengan menggunakan metode ekstraksi fitur antara TF-IDF, BoW, dan FastText

METODE PENELITIAN

Penelitian ini membandingkan tiga ekstraksi fitur pada analisis sentimen. Ekstraksi fitur itu terdiri dari TF-IDF, BoW, dan FastText dengan menggunakan algoritma *Support Vector Machine* (SVM). Dataset diperoleh dari proses *crawling* pada twitter mengenai kenaikan BBM yang kemudian diuji dengan tiga ekstraksi fitur untuk mendapatkan hasil akurasi dari setiap metode ekstraksi fitur yang digunakan. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Crawling Data

Crawling data tentang kenaikan BBM diambil dari tanggal 08 September 2022 sampai 26 September 2022 yang dilakukan melalui API twitter dengan menggunakan bahasa

pemrograman python memanfaatkan tools library tweepy. Data diambil dari akun twitter acak yang berjumlah 1667 data dengan keyword “bbm”, “pertamax”, dan “pertalite” yang sering dibahas di media sosial twitter. Dengan mengambil kolom *created_at*, *text*, dan *name* yang dapat dilihat di dalam Tabel 1.

Tabel 1. Dataset Twitter

<i>Created_at</i>	<i>text</i>	<i>name</i>
2022-09-08 23:59:58+00:00	@CNNIndonesia Aneh, harga BBM naik tgl 3/09 tapi kagetnya baru sekarang. Apakah kagetnya lalu kena heart attack?ðŸ˜†	@boa_cr77
2022-09-08 23:59:56+00:00	Di ujung sana, mikrofon berbunyi "jadi rakyat jangan cengeng!". Di ujung sana pula, mulut pejabat berbunyi "Postur APBN tersandra subsidi, memberatkan keuangan pemerintah, keputusan bbm naik tak bisa dihindari". Hayo siapa yg cengeng?	Sudarsono Saidi
2022-09-08 23:59:52+00:00	@CNNIndonesia Kenapa sih gak pake sistem kupon aja yg dikasih saat bayar pajak kendaraan? Kupon dikasih sejumlah brp liter jatah BBM, disobek oleh operator tiap ngisi di SPBU sebanyak liter BBM yg diisi @pertamina	Budi Permana
2022-09-08 23:59:51+00:00	Politikus Irwan (PD) dan Adian (PDIP) Bersiteru Soal Kenaikan Harga BBM https://t.co/qWBPF4OIGK	Ali Syarief

Preprocessing Data

Tahap *preprocessing data* terdiri dari beberapa sub-proses seperti *cleaning text*, *tokenization*, *stop removal* atau *stopwords*, *case folding*, dan pelabelan data[16]. Tahap *cleaning text* digunakan untuk menghilangkan tanda baca dan angka yang tidak digunakan. Selanjutnya tahap *tokenization* digunakan untuk memecah beberapa teks menjadi satuan kata yang bertujuan untuk memisahkan kata yang dapat diperlukan sebagai pemisah atau bukan. Kemudian tahap *stop removal* atau *stopwords* digunakan untuk menghilangkan kata yang dianggap tidak memiliki arti atau tidak berpengaruh. Tahap *case folding* untuk menyamaratakan huruf kapital. Tahap terakhir adalah pelabelan data yang dilakukan secara manual berdasarkan kata bersentimen tertentu, yang nanti setiap data akan diberi label dengan tulisan positif, netral, dan negatif. Proses *preprocessing* dapat dilihat di Tabel 2 dan proses labeling data dapat dilihat di Tabel 3.

Tabel 2. Proses Preprocessing

Nama Proses	Text
Data Review	@abu_waras Boros amat BBM, katanya mahal
Cleaning Text	abu waras Boros amat BBM katanya mahal
Tokenization	['', 'abu', 'waras', 'Boros', 'amat', 'BBM', 'katanya', 'mahal']
Stop Removal / Stopwords	['', 'abu', 'waras', 'Boros', 'BBM', 'mahal']
Case Folding	abu waras boros bbm mahal

Tabel 3. Labeling Data

Text	Label
ma ruf amin bbm penyesuaian ekonomiriterusmeningkat https://t.co/brtathqd7f	positif
lebanon santino blm sejarah nya demo bbm bbm gratisan	netral
abu waras boros bbm mahal	negatif

Ekstraksi Fitur

Tahapan ekstraksi fitur dilakukan dalam sentimen dimana pembelajaran mesin tidak bisa diinput dalam bentuk *string*. Metode ekstraksi menghasilkan bobot kata yang akan diubah ke dalam bentuk vektor dari setiap kata yang ada pada dokumen.

TF-IDF

TF (*Term Frequency*) yang berguna untuk menormalisasikan jumlah kata di dalam dokumen dan IDF (*Inverse Document Frequency*) yang digunakan untuk mendapatkan nilai dari beberapa jumlah kata dalam dokumen, karena IDF bekerja dengan sangat baik [17].

Bag of Word

Model *Bag of Word* merupakan representasi kata setiap kalimat dengan mengubah kalimat ke dalam bentuk vektor dengan munculnya setiap kata pada kalimat tersebut. Karena BoW merupakan cara representasi kata paling sederhana untuk mengubah teks menjadi data numerik [13]. Penelitian ini kami menggunakan library `CountVectorizer` pada `sklearn`.

FastText

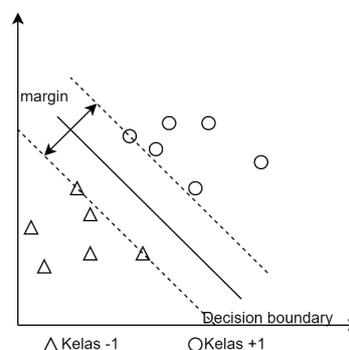
Representasi kata yang dikembangkan oleh Facebook Research memiliki kinerja yang lebih baik dan cepat dengan memasukkan karakter n -gram ke dalam model skipgram yang dimana kata direpresentasikan atas karakter n -gram [10]. Tujuan dasar melakukan representasi kata adalah untuk menganalisis struktur internal data.

Split Data

Dalam melatih data kami melakukan pemisahan antara data pelatihan dan data pengujian. Pada penelitian ini dilakukan pembagian data dengan persentase 80% yang digunakan untuk data pelatihan model dan 20% yang digunakan untuk data pengujian model menggunakan library `sklearn` dengan mengimpor `train_test_split`.

Support Vector Machine

Algoritma *Support Vector Machine* (SVM) kami implementasikan dengan library `sklearn` dengan mengimpor `SVC`. Algoritma SVM efisien dalam melakukan klasifikasi [18]. Masalah untuk menentukan sentimen dapat dijelaskan dalam usaha menemukan *hyperplane* dan *hyperplane* terbaik ditentukan dengan mengukur margin *hyperplane* untuk mendapatkan titik paling maksimal [19]. Dalam menemukan *hyperplane* dapat melihat konsep ilustrasi dalam Gambar 2.



Gambar 2. Ilustrasi SVM

Gambar 2 menjelaskan bagaimana konsep SVM bekerja yang terdapat beberapa data dengan bentuk segitiga untuk kelas -1 dan lingkaran untuk kelas +1. *Hyperplane* yang paling maksimal

ialah dengan cara menghitung setiap jarak antar margin pada masing-masing kelas. Data yang paling dekat bisa disebut sebagai *support vector* karena memiliki *hyperplane* terbaik di setiap kemungkinan. Persamaan 1 dan Persamaan 2 *support vector machine* yang memiliki *hyperplane* dan Persamaan 3 digunakan untuk menghitung margin [20].

$$w \cdot x + b = -1 \quad (1)$$

$$w \cdot x + b = +1 \quad (2)$$

$$\text{Margin} = \frac{2}{\|w\|} \quad (3)$$

Metode Evaluasi dan Analisis

Untuk metode evaluasi pada penelitian ini kami menggunakan *confusion matrix* yang digunakan untuk mengukur hasil nilai performa dari setiap percobaan yang dilakukan. *Confusion matrix* mempunyai memiliki evaluasi yang harus dianalisis yaitu *Precision* (Persamaan 4), *Recall* (Persamaan 4), *Accuracy* (Persamaan 4), dan *F1-Score* (Persamaan 4).

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

$$\text{F1 - Score} = \frac{2 \times P \times R}{P + R} \times 100\% \quad (7)$$

Keterangan:

TP	: True Positive
TN	: True Negative
FP	: False Positive
FN	: False Negative
P	: Precision
R	: Recall

Setelah melakukan tahap evaluasi menggunakan *confusion matrix* yang digunakan untuk mendapatkan nilai performa, maka dilakukan tahap perbandingan pada tiga ekstraksi fitur untuk menentukan metode mana yang memiliki kinerja dan performa paling baik saat melakukan proses analisis sentimen.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan bahasa pemrograman python 3 untuk melakukan perbandingan metode ekstraksi fitur. *Dataset* twitter yang digunakan memiliki jumlah data sebesar 1667 data dengan menggunakan API twitter yang sudah dilakukan tahap *preprocessing* pada data seluruhnya. Selanjutnya akan dilakukan *split data* dengan rasio 80:20 yaitu 80% data pelatihan dan 20% data pengujian. Data pelatihan kemudian dilakukan perhitungan dengan algoritma *Support Vector Machine* dan metode ekstraksi fitur seperti TF-IDF, BoW, dan FastText.

Langkah pertama menggunakan beberapa data yang ada pada *file* hasil *preprocessing*.

Data 1 : section pertamax pertalite sejalur

Data 2 : pertamax pertalite

Data 3 : rasain isi pertamax pertalite abis berasa ygy duitnya

Selanjutnya dari setiap kata dari tiga data tersebut dilakukan perhitungan TF-IDF yang tersaji pada Tabel 4.

Tabel 4. Perhitungan TF-IDF

	Data 1	Data 2	Data 3	TF1	TF2	TF3	IDF	TF-IDF1	TF-IDF2	TF-IDF3
section	1	0	0	0.25	0	0	0.48	0.12	0	0
pertamax	1	1	1	0.25	0.5	0.125	0	0	0	0
pertalite	1	1	1	0.25	0.5	0.125	0	0	0	0
sejalur	1	0	0	0.25	0	0	0.48	0.12	0	0
rasain	0	0	1	0	0	0.125	0.48	0	0	0.06
isi	0	0	1	0	0	0.125	0.48	0	0	0.06
abis	0	0	1	0	0	0.125	0.48	0	0	0.06
berasa	0	0	1	0	0	0.125	0.48	0	0	0.06
ygy	0	0	1	0	0	0.125	0.48	0	0	0.06
duitnya	0	0	1	0	0	0.125	0.48	0	0	0.06

Dari Tabel 4 didapatkan vektor dari masing-masing data seperti berikut:

Data 1 = [0.12, 0, 0, 0.12, 0, 0, 0, 0, 0, 0]

Data 2 = [0,0 ,0 ,0, 0, 0, 0, 0, 0, 0]

Data 3 = [0, 0, 0, 0, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06]

Sedangkan untuk Bag of Word perhitungannya dapat dilihat pada Tabel 5.

Tabel 5. Perhitungan Bag of word

Data	section	pertamax	pertalite	sejalur	rasain	isi	abis	berasa	ygy	duitnya
1	1	1	1	1	0	0	0	0	0	0
2	0	1	1	0	0	0	0	0	0	0
3	0	1	1	0	1	1	1	1	1	1

Dari Tabel 5 didapatkan vektor dari setiap data seperti berikut:

Data 1 = [1, 1, 1, 1, 0, 0, 0, 0, 0, 0,0]

Data 2 = [0, 1, 1, 0, 0, 0, 0, 0, 0, 0,0]

Data 3 = [0, 1, 1, 0, 1, 1, 1, 1, 1, 1,1]

Selanjutnya untuk FastText, nilai dari setiap kata yang mendekati kata “bors” untuk melihat kinerja dari FastText bekerja atau tidak. Contoh yang digunakan ialah kata “bors” menunjukkan kalau terdapat salah ketik dalam kalimat FastText berusaha mencari kata tersebut. Dapat dilihat pada Gambar 3.

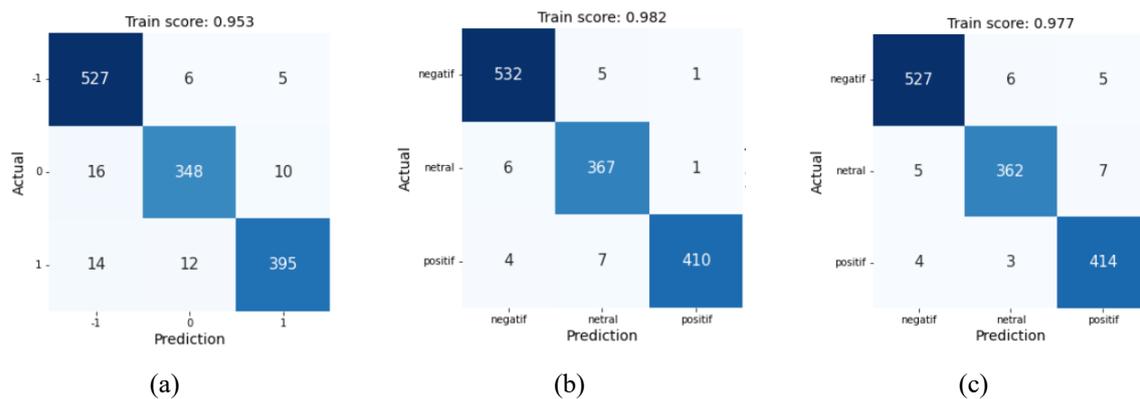
```
[('boros', 0.8142074346542358),
 ('bos', 0.47775787115097046),
 ('bocor', 0.38675904273986816),
 ('dipake', 0.38490962982177734),
 ('boss', 0.38052093982696533)]
```

Gambar 3. Mendekati kata “bors”

Dari Gambar 3 mencari kata “bors” akan mengeluarkan beberapa vektor diman FastText dapat mengidentifikasi kata setiap n -gram yang paling mendekati dan kata apa yang sering muncul bersamaan dengan kata tersebut.

Hasil dari setiap ekstraksi fitur akan dilakukan pelatihan dengan metode SVM untuk menemukan model mengidentifikasi sentimen yang dilakukan pada data. Model tersebut akan

diuji dengan data uji yang hasil prediksinya dapat dilihat pada Gambar 4 untuk masing-masing ekstraksi fitur.



Gambar 4 Hasil prediksi model SVM pada ekstraksi fitur : (a) TF-IDF, (b) BoW, dan (c) FastText

Dari Gambar 4 dapat dilihat hasil *train score* dari TF-IDF memiliki nilai sebesar 0.953 atau 95.3%, nilai *train score* BoW sebesar 0.982 atau 98.2, dan hasil *train score* FastText sebesar 0.977 atau 97.7%.

Selanjutnya model akan dilakukan perhitungan untuk menentukan akurasi berdasarkan hasil prediksi dan melakukan perbandingan hasil dari setiap model yang sudah dibangun. Jumlah data yang sudah kami gunakan memiliki label positif, netral, dan negatif yang dapat dilihat dalam pada Tabel 6.

Tabel 6. Distribusi data setiap sentimen

No.	Sentimen	Data
1.	Positif	527
2.	Netral	467
3.	Negatif	673

Langkah selanjutnya model dievaluasi menggunakan *confusion matrix* dengan menghitung *precision*, *recall*, *f1-score*, dan *accuracy*.

Hasil percobaan menggunakan *confusion matrix* dengan *machine learning* dikombinasikan dengan algoritma dan metode ekstraksi fitur yang sudah ditentukan. Hasil dari setiap ekstraksi fitur menunjukkan bahwa penggunaan metode BoW memiliki hasil akurasi yang lebih baik dibandingkan dengan metode yang lain dapat dilihat pada Tabel 7.

Tabel 7. Hasil ekstraksi fitur

	TF-IDF				BoW				FastText			
	Precision	Recall	F1-Score	Acc	Precision	Recall	F1-Score	Acc	Precision	Recall	F1-Score	Acc
Negatif	59%	81%	68%		71%	69%	70%		62%	75%	68%	
Neutral	60%	47%	53%	62.6%	55%	61%	58%	65.3%	54%	49%	52%	59.3%
Positif	73%	53%	61%		68%	64%	66%		59%	48%	53%	

Pada Tabel 7 terlihat bahwa nilai presisi dari metode TF-IDF lebih unggul dibandingkan dengan BoW pada sentiment netral dan positif. Hal tersebut terjadi karena dataset yang minimal pada kedua sentiment tersebut membuat perhitungan setiap kata yang muncul menjadi tidak optimal ketika menggunakan metode BoW. Pada penelitian ini nilai metrik yang digunakan sebagai

evaluasi model yaitu *FI-Score* disebabkan persebaran dataset yang tidak seimbang. Nilai *FI-Score* dari semua sentimen baik itu negatif, netral atau positif didominasi oleh metode BoW. Untuk memperjelas perbedaan ketiga metode ekstraksi fitur tersebut, dilakukan perhitungan rata-rata nilai *FI-Score* dari ketiga metode tersebut seperti terlihat pada Tabel 8.

Tabel 8. Perhitungan *FI-score*

No.	Ekstraksi Fitur	Rata-rata
1.	TF-IDF	61%
2.	BoW	65%
3.	FastText	58%

Dari Tabel 8 dapat dilihat bahwa penggunaan BoW memiliki hasil rata-rata paling baik dibandingkan ekstraksi fitur TF-IDF dan FastText.

KESIMPULAN

Penelitian ini melakukan perbandingan metode ekstraksi fitur TF-IDF, BoW, dan FastText yang merupakan teknik merepresentasikan data yang paling populer, kemudian dikombinasikan dengan menggunakan algoritma machine learning yaitu SVM. Hasil dari penelitian ini menunjukkan bahwa model menggunakan ekstraksi fitur BoW menggunakan algoritma SVM memiliki hasil kinerja lebih baik dibandingkan dengan model ekstraksi fitur TF-IDF dan FastText.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) dan Program Studi Informatika Universitas Amikom Yogyakarta atas dukungan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Y. Dewi, S. S, A. Dini, M. M, and R. Mauli, "Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok (Sembako) Di Kecamatan Tambun Selatan Dalam Masa Pandemi," *J. Citizsh. Virtues*, vol. 2, no. 2, pp. 320–326, 2022, doi: 10.37640/jcv.v2i2.1533.
- [2] S. Sarbaini and N. Nazaruddin, "Pengaruh Kenaikan BBM Terhadap Laju Inflasi di Indonesia," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 2, no. I, pp. 25–32, 2023, doi: 10.55826/tmit.v2i1.132.
- [3] I. Najiyah, "Analisis Sentimen Tanggapan Masyarakat Indonesia Tentang Kenaikan Bbm Menggunakan Metode Artificial Neural Network," *J. Responsif*, vol. 5, no. 1, pp. 92–100, 2023, [Online]. Available: <https://ejurnal.ars.ac.id/index.php/jti>
- [4] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Inf. Syst. Public Heal.*, vol. 4, no. 1, pp. 28–36, 2019.
- [5] S. M. Permataning Tyas, B. S. Rintyarna, and W. Suharso, "The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 1–10, 2022, doi: 10.31849/digitalzone.v13i1.9158.
- [6] M. Harahap, B. P. A. Sihombing, O. A. F. Laia, B. T. Saragih, and K. Dharma, "Analisis

- Sentimen Review Penjualan Produk Umkm Pada Kabupaten Nias Dengan Komparasi Algoritma Klasifikasi Machine Learning,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 5, no. 2, pp. 147–154, 2021, doi: 10.46880/jmika.vol5no2.pp147-154.
- [7] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [8] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [9] K. S. Jones, “A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL,” *J. Doc.*, vol. 28, no. 1, pp. 11–21, 1972, [Online]. Available: <https://doi.org/10.1108/eb026526>
- [10] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information,” *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, 2017, doi: 10.1162/tacl_a_00051.
- [11] R. Batra and S. M. Daudpota, “Integrating StockTwits with sentiment analysis for better prediction of stock price movement,” *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICOMET.2018.8346382.
- [12] F. A. Wicaksono and A. Romadhony, “Sentiment Analysis of University Social Media Using Support Vector Machine and Logistic Regression Methods,” vol. 7, no. August, pp. 15–24, 2022, doi: 10.34818/indoic.2022.7.2.638.
- [13] Dhammajoti, J. C. Young, and A. Rusli, “A comparison of supervised text classification and resampling techniques for user feedback in bahasa Indonesia,” *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, 2020, doi: 10.1109/ICIC50835.2020.9288588.
- [14] A. A. Septiandri and O. Wibisono, “Detecting spam comments on Indonesia’s Instagram posts,” *J. Phys. Conf. Ser.*, vol. 801, no. 1, 2017, doi: 10.1088/1742-6596/755/1/011001.
- [15] F. Es-Sabery, K. Es-Sabery, H. Garmani, J. Qadir, and A. Hair, “Evaluation of different extractors of features at the level of sentiment analysis,” *Infocommunications J.*, vol. 14, no. 2, pp. 85–96, 2022, doi: 10.36244/ICJ.2022.2.9.
- [16] A. J. Arifin and A. Nugroho, “Uji Akurasi Penggunaan Metode KNN dalam Analisis Sentimen Kenaikan Harga BBM pada Media Twitter,” no. September, pp. 700–708, 2022.
- [17] S. Robertson, “Understanding inverse document frequency: On theoretical arguments for IDF,” *J. Doc.*, vol. 60, no. 5, pp. 503–520, 2004, doi: 10.1108/00220410410560582.
- [18] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The impact of features extraction on the sentiment analysis,” *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.
- [19] A. A. Kasim and M. Sudarsono, “Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah,” *Semin. Nas. APTIKOM*, pp. 568–573, 2019.
- [20] I. Monika Parapat and M. Tanzil Furqon, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3163–3169, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [21] Fauzi, M. R., Pratama, R. A., Laksono, P., & Eosina, P. (2021). Penerapan Big Data Menggunakan Algoritma Multi-Label K-Nearest Neighbor dalam Analisis Sentimen

-
- Konsumen UMKM Sektor Kuliner. Krea-TIF: Jurnal Teknik Informatika, 9(1), 9–20. <https://doi.org/10.32832/kreatif.v9i1.3587>
- [22] Habyba, A. N., Djatna, T., & Anggraeni, E. (2021). Positioning E-commerce Produk UKM berdasarkan Kebutuhan Afektif Pengguna. Krea-TIF: Jurnal Teknik Informatika, 9(1), 21–28. <https://doi.org/10.32832/kreatif.v9i1.3590>