

Perbandingan Algoritma *Support Vector Machine* dan *Random Forest* untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022

Muhamad Samantri ^{1*}, Afiyati ²

^{1,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia.

article info

Article history:

Received 19 May 2023

Received in revised form

7 October 2023

Accepted 20 November 2023

Available online January 2024

DOI:

<https://doi.org/10.35870/jtik.v8i1.1202>

Keywords:

Sentiment Analysis; Fuel Prices; Support Vector Machine; Random Forest.

Kata Kunci:

Analisis Sentimen; Harga BBM; Support Vector Machine; Random Forest.

abstract

The commodity of fuel oil (BBM) is the main commodity and the driving force of business. The increase in world oil prices is a threat to countries around the world, one of which is Indonesia. With the turbulent conditions in several countries, the Indonesian government decided to cut fuel subsidies which had an impact on price increases. The policy invited all Indonesian people and criticized it on various social media. The purpose of this research is to find out which algorithm has a better accuracy rate and to provide input to the government about public opinion regarding the increase in fuel prices in Indonesia. From the test results both work well, this is evidenced by the accuracy value obtained, where the support vector machine algorithm produces an accuracy value of 77%, while the Random Forest algorithm produces an accuracy value of 76%. So it can be concluded that the support vector machine algorithm has a fairly good accuracy rate compared to the Random Forest algorithm.

abstrak

Komoditi bahan bakar minyak (BBM) merupakan komoditas utama dan motor penggerak bisnis kenaikan harga minyak dunia menjadi ancaman bagi negara-negara di seluruh dunia, salah satunya di Indonesia. Dengan adanya pergolakan kondisi pada beberapa negara, pemerintah Indonesia mengambil keputusan memangkas subsidi BBM yang berdampak pada kenaikan harga. Kebijakan tersebut mengundang seluruh masyarakat Indonesia dan mengkritiknya di berbagai media sosial. Tujuan dari penelitian ini untuk mengetahui algoritma mana yang memiliki tingkat akurasi yang lebih baik dan untuk memberikan masukan terhadap pemerintah tentang pendapat masyarakat mengenai kenaikan harga bbm di indonesia. Dari hasil pengujian keduanya bekerja dengan baik, hal ini dibuktikan dengan diperolehnya nilai akurasi, dimana algoritma support vector machine menghasilkan nilai akurasi sebesar 77%, sedangkan algoritma Random Forest menghasilkan nilai akurasi sebesar 76%. Sehingga dapat disimpulkan bahwa algoritma support vector machine memiliki tingkat akurasi yang cukup baik dibandingkan dengan algoritma Random Forest.

Corresponding Author. Email: muh.samantri07@gmail.com ^{1}.

© E-ISSN: 2580-1643.

Copyright © 2024 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISEI). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



ACM Computing Classification System (CCS)

EBSCOhost

Communication and Mass Media Complete (CMC)

1. Latar Belakang

Pada awal September 2022, pemerintah menerapkan kebijakan menaikkan harga Bahan Bakar Minyak (BBM) di Indonesia. Kebijakan ini merupakan buntut dari perang di Eropa antara Rusia dan Ukraina yang membuat harga minyak dunia meroket. Akibatnya, terjadi guncangan ekonomi yang berdampak besar pada harga minyak, termasuk di Indonesia. Embargo perdagangan minyak Rusia menyebabkan kejutan pasokan, sehingga mengakibatkan harga yang lebih tinggi di pasar dunia. Lonjakan harga minyak dunia juga membantu Indonesian Oil Price atau Indonesian Crude Oil Price (ICP) mencapai \$95,45 per barel. Sedangkan asumsi ICP dalam APBN 2022 hanya \$63 per barel. Kenaikan suku bunga dan likuiditas tentu akan mempengaruhi negara-negara berkembang, termasuk Indonesia, yang berujung pada peningkatan anggaran subsidi yang ada. BBM merupakan komoditas yang sangat penting bagi masyarakat Indonesia, khususnya bagi dunia usaha. Volatilitas harga BBM dapat mempengaruhi biaya operasional dalam kegiatan sehari-hari masyarakat. Kebijakan menaikkan harga BBM memicu kontroversi di media sosial, termasuk Twitter. Beberapa hari setelah pengumuman kenaikan harga BBM, linimasa Twitter dibanjiri *tweet* ber-tag BBM naik. Pro dan kontra yang hadir di tengah masyarakat menjadi menarik untuk dapat diteliti dengan mengetahui tanggapan masyarakat. Ada banyak cara untuk menganalisis tanggapan masyarakat, terutama di media sosial. Opini publik dapat dianalisis dengan menggunakan banyak teknik analisis, salah satunya adalah analisis sentimen. Metode analisis sentimen mencari polaritas dalam data khususnya di internet dengan melakukan analisis [1].

Dalam penelitian ini, penulis menerapkan analisis opini publik terhadap kebijakan pemerintah terkait kenaikan harga BBM di Indonesia. Analisis sentimen adalah studi terkomputerisasi tentang opini, perasaan, dan emosi yang diekspresikan dalam bentuk teks. Tujuan analisis sentimen adalah untuk menentukan peringkat polaritas teks atau opini dalam teks media sosial (*tweet*), terlepas dari itu apakah *tweet* tersebut bersifat positif atau negatif. Untuk proses analisis dan kategorisasi, penelitian ini menggunakan data teks dari hasil *crawling* Twitter, data yang dikumpulkan mengenai kenaikan harga BBM tahun 2022 dengan kata kunci “bbm naik” dan data yang

diperoleh dari hasil *crawling* berjumlah 4256 dataset. Kajian ini bertujuan untuk menggali opini publik terhadap kenaikan harga BBM dan bentuk usulan atau kontribusi kepada pemerintah Indonesia. Hasil penelitian ini menggunakan pendekatan kuantitatif untuk menunjukkan proporsi dan perbedaan pendapat masyarakat terhadap kenaikan harga BBM, dan membandingkannya dengan menggunakan metode *Support vector machine* (SVM) dan *Random Forest Classification*.

2. Metode Penelitian

Pada tahap ini menjelaskan langkah-langkah dalam proses penerapan metode support vektor machine yang akan dilakukan perbandingan dengan metode *Random Forest*. Dimana tujuan dari penelitian ini adalah untuk mengetahui hasil kinerja pada metode *support vector machine* dan *Random Forest* dalam analisis sentimen yang dapat dilihat pada Gambar 1 [2].



Gambar 1. Flow Penelitian

Berdasarkan gambar 1 diatas dapat dijelaskan sebagai berikut:

Pengumpulan Data

Untuk mendapatkan sebuah dataset membutuhkan proses pengumpulan data yang diambil dari hasil *crawling* pada platform media sosial Twitter menggunakan API dan *library Tweepy* dengan bahasa pemrograman Python. Dari proses *crawling* tersebut dataset yang diperoleh berjumlah 4256 data dengan rentang waktu dari tanggal 09 September – 17 September 2022 menggunakan kata kunci pencarian “bbm naik”[3] dan kemudian diekstraksi menjadi format.csv.

Labeling Data

Pada tahap pelabelan data ini, nantinya data akan dikategorikan menjadi tiga, yaitu data yang bersifat

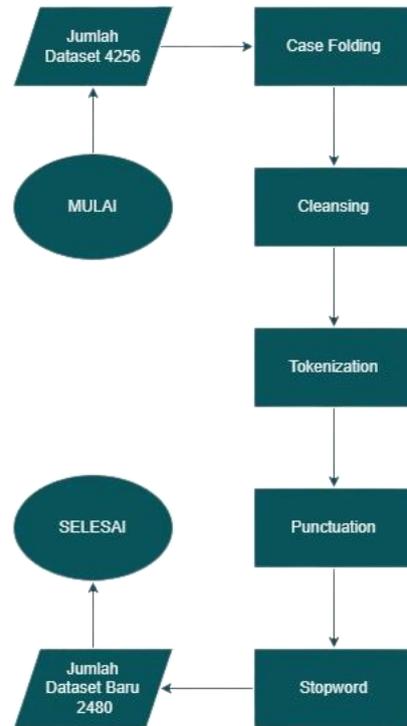
positif, negatif dan netral. Data tersebut akan diproses secara otomatis menggunakan Library pada bahasa pemrograman Python.

Preproses Data

Pada tahap ini, data harus diproses terlebih dahulu untuk memfasilitasi pembelajaran mesin dengan menghapus kata-kata yang tidak perlu seperti kata sambung, titik, dan koma dari *tweet*. Tujuan pra-pemrosesan adalah untuk menyederhanakan desain pembelajaran mesin dan meningkatkan akurasi [4]. Adapun tahapan pre-prosesnya dapat dilihat pada Gambar 2 [5], diantaranya:

- 1) *Case folding*
Case folding adalah proses mengonversi semua teks dalam *tweet* menjadi format standar (huruf kecil) [6].
- 2) *Cleansing*
Cleansing adalah proses menghapus semua simbol, *mention*, taggar, *retweet*, URL, dan emoticon dalam sebuah dataset [7].
- 3) *Tokenization*
Tokenization adalah proses memecah teks menjadi kata, frasa, simbol, atau elemen bermakna lainnya yang disebut token. Tujuan dari *tokenization* adalah untuk membentuk token kata dari kalimat [8].
- 4) *Remove Punctuation*
Remove Punctuation adalah tata cara menghilangkan tanda baca seperti titik(.), koma (,), garis miring (/), angka, dan emoticon. Komponen ini dihilangkan karena tidak digunakan dalam proses analisis sentiment [9].
- 5) *Stopword Removal*
Stopword Removal, yaitu fase yang mengurutkan kata-kata berharga dari hasil tokenisasi dengan mempertahankan kata-kata berharga dan menghapus kata-kata yang tidak layak [10].

Dengan melalui serangkaian tahapan pre-proses ini, data yang digunakan dalam penelitian menjadi lebih terstruktur dan siap untuk digunakan dalam proses pembelajaran mesin. Langkah-langkah ini juga bertujuan untuk meningkatkan kualitas analisis sentimen dengan menghilangkan noise dan memfokuskan pada informasi yang relevan.



Gambar 2. *Flow Pre-processing*

Validasi

Salah satu cara validasi data adalah dengan menggunakan cross validation yaitu metode validasi model dengan memisahkan data menjadi dua subset: data pelatihan dan data uji. Metode ini juga dikenal sebagai k-fold cross-validation. Percobaan validasi dijalankan sebanyak k kali pada model dengan parameter yang sama. Pelatihan model pada data pelatihan menggunakan K-fold Cross Validation (K-Cv). Dalam K-Cv, dapat merekam hingga k data pelatihan, 1 bagian sebagai pengujian, ulangi k kali, memberikan setiap bagian kesempatan untuk digunakan sebagai pengujian[11].

Klasifikasi

Setelah tahapan split data, dimana data dibagi menjadi 2 yaitu data train dan data testing. Selanjutnya akan dilakukan proses klasifikasi pada model *support vector machine* dan *Random Forest* untuk mendapatkan hasil akurasi dari kedua algoritma tersebut dan akan dikomparasi, untuk detail alurnya terlihat pada Gambar 3[12].



Gambar 3. Flow Klasifikasi Algoritma

Evaluasi

Evaluasi model klasifikasi didasarkan pada ukuran kinerja model klasifikasi, yang menggambarkan seberapa baik sistem mengklasifikasikan data. Salah satu metode yang dapat digunakan untuk mengukur kinerja model klasifikasi adalah confusion matrix. Confusion matrix berisi informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Setiap sel berisi angka yang menunjukkan jumlah kasus sebenarnya di kelas yang diamati untuk diprediksi. Kriteria Confusion matrix dapat dilihat pada Tabel 1 [13]. Untuk detail flownya dapat dilihat pada Gambar 4 [14].

Tabel 1. Confusion Matrix

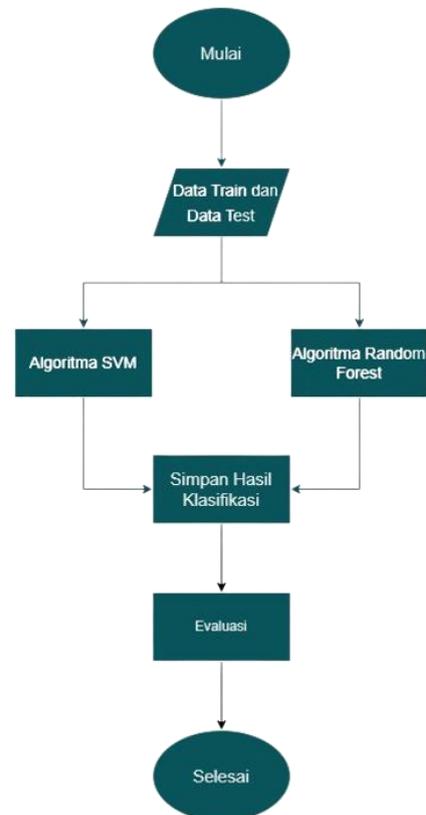
	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positif)	FN (False Negatif)
Negatif	FP (False Positif)	TP (True Positif)

Hasil klasifikasi dapat dihitung berdasarkan tingkat akurasi kinerja matriks. Untuk menghitung presisi matriks dapat menggunakan rumus berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Sensitivity = TPrate = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$



Gambar 4. Flow Evaluasi Model

3. Hasil dan Pembahasan

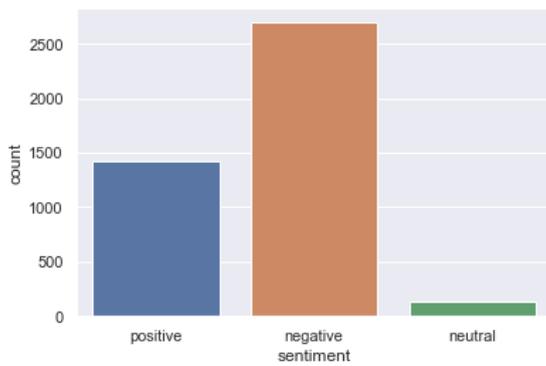
Dataset

Tahap pertama pada penelitian ini, yaitu pengumpulan data, dimana data yang akan dianalisa merupakan data teks yang telah didapatkan dari hasil *crawling* menggunakan API Tweepy pada media sosial Twitter dalam rentang waktu 09 September sampai 17 September 2022 dengan bahasa pemrograman Python. Dari proses tersebut data yang diperoleh sebanyak 4256 dataset dengan kata kunci pencarian “bbm naik”, dataset tersebut terdiri dari 4256 baris dan 5 kolom, seperti terlihat pada Gambar 5.

	Id	Waktu	Username	Tweet	Lokasi
0	158205829342121984	2022-10-17 17:23:51+00:00	Aryo Hendriqono	b'RT @machbeach: "Pidato presiden ini mengatak...	UT -6.215955, 106.865716
1	1582059521832517634	2022-10-17 17:22:21+00:00	Revolusi_Merahpuh	b'RT @machbeach: "Pidato presiden ini mengatak...	Indonesia Tercinta
2	1582059213379231744	2022-10-17 17:21:24+00:00	bucimobuchen	b'@nyyssa yagmanaaa suami w kalo ngechat rand...	Kota Surabaya, Jawa Timur
3	1582059131498024960	2022-10-17 17:21:04+00:00	adi ultras redded capital milanis transmana	b'Semua mahal mbas BBM naik, tapi upah karyaw...	SURABAYA, **
4	1582058615552479232	2022-10-17 17:19:01+00:00	kapten jinglong	b'RT @machbeach: "Pidato presiden ini mengatak...	Jawa Timur, Indonesia
...
4251	1578978670211133445	2022-10-09 05:20:25+00:00	Tusino Mukti	b'RT @Mentari3333333: Di cari subudara ya di...	Jakarta Capital Region
4252	157897793825663744	2022-10-09 05:17:31+00:00	Shanks_Yonko	b'BBM naik jangan panik subsidi tepat sasaran ...	Purwokerto Selatan, Indonesia
4253	157897775972941829	2022-10-09 05:16:52+00:00	fananya ARTIST OF THE YEAR ARIANA	b'RT @meropage_kangen @PDL_Pelautian deno di...	Planet Mars
4254	1578976087669112832	2022-10-09 05:10:09+00:00	#rt_@	b'RT @NellzenBersik2: Dinageti Dunia Terbaik...	Highlands-Baywood Park, CA
4255	1578974571587588096	2022-10-09 05:04:06+00:00	kumparan	b'BBM Partalle terus menjadi polemik setelah ...	Nahi

Gambar 5. Jumlah Dataset

Setelah dataset terkumpul, tahap selanjutnya yaitu memberikan label pada setiap data yang akan dikategorikan menjadi data positif, data negatif, dan netral. Proses awal pelabelan, yaitu menentukan bobot kata dengan kategori positif dan negatif pada setiap *tweet*. Dimana untuk bobot kata dengan kategori positif berjumlah 3609 dan negatif berjumlah 6609 jika ditotal keseluruhan bobot kata antara positif dan negatif berjumlah 10218 kata. Selanjutnya proses pelabelan dilakukan secara otomatis di mana pada setiap *tweet* pada dataset akan diberikan label positif atau negatif sesuai dengan kamus yang sudah tersedia pada library python, Hasil dari pelabelan data tersebut didapatkan jumlah data dengan kategori positif sebesar 1422 data, untuk kategori negatif sebesar 2696 data dan kategori netral sebesar 138 data. Namun pada penelitian ini hanya menggunakan data dengan kategori positif dan negatif. Untuk proses dan hasil dapat dilihat pada Gambar 6.



Gambar 6. Hasil Pelabelan Data

Pre-Processing

Setelah melakukan labeling pada data, selanjutnya akan dilakukan tahap pre-processing guna untuk menghilangkan noise pada data dan meningkatkan hasil akurasi pada model. Terdapat 5 tahapan pre-processing, antara lain:

1) *Case folding*

Tabel 2. Hasil *Case folding*

Sebelum	Sesudah
@Miduk17 Pertamina runtuh kacau balau\nBanyak rakyat mengelluh di sebabkan BBM naik, jadi Ellu jadi Buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk.	@miduk17 buktinya pertamina runtuh kacau balau\nbanyak rakyat mengelluh di sebabkan bbm naik, jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk.

2. *Cleansing*

Tabel 3. Hasil *Cleansing*

Sebelum	Sesudah
@miduk17 buktinya pertamina runtuh kacau balau\nbanyak rakyat mengelluh di sebabkan bbm naik, jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk.	buktinya pertamina runtuh kacau balau\nbanyak rakyat mengelluh di sebabkan bbm naik, jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk

3. *Tokenization*

Tabel 4. Hasil *Tokenization*

Sebelum	Sesudah
buktinya pertamina runtuh kacau balau\nbanyak rakyat mengelluh di sebabkan bbm naik, jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk	['buktinya', 'pertamina', 'runtuh', 'kacau', 'balau', '\\n', 'n', 'banyak', 'rakyat', 'mengelluh', 'di', 'sebabkan', 'bbm', 'jadi', 'ellu', 'buser', 'jangan', 'seperti', 'lalat', 'yang', 'gak', 'bisa', 'membedakan', 'mana', 'yang', 'haru', 'dan', 'mana', 'yang', 'bau', 'busuk']

4. Remove Punctuation

Tabel 5. Hasil Remove Punctuation

Sebelum	Sesudah
[‘buktinya’, ‘pertamina’, ‘runtuh’, ‘kacau’, ‘balau’, ‘\’, ‘n’, ‘banyak’, ‘rakyat’, ‘mengelluh’, ‘di’, ‘sebabkan’, ‘bbm’, ‘naik’, ‘jadi’, ‘ellu’, ‘jadi’, ‘buser’, ‘jangan’, ‘seperti’, ‘lalat’, ‘yang’, ‘gak’, ‘bisa’, ‘membedakan’, ‘mana’, ‘yang’, ‘haru’, ‘dan’, ‘mana’, ‘yang’, ‘bau’, ‘busuk’]	buktinya pertamina runtuh kacau balau n banyak rakyat mengelluh di sebabkan bbm naik jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk

5. Stopword Removal

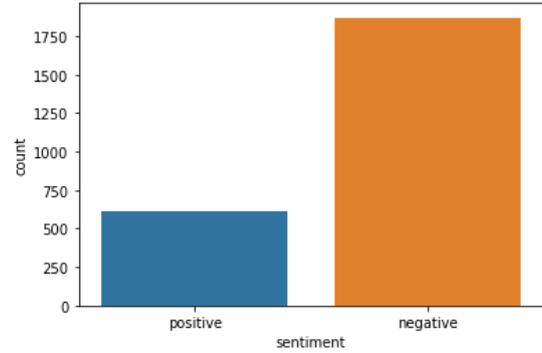
Tabel 6. Hasil Stopword Removal

Sebelum	Sesudah
buktinya pertamina runtuh kacau balau n banyak rakyat mengelluh di sebabkan bbm naik jadi ellu jadi buser jangan seperti lalat yang gak bisa membedakan mana yang haru dan mana yang bau busuk	buktinya pertamina runtuh kacau balau banyak rakyat mengelluh sebabkan bbm naik elu jadi buser jangan seperti lalat yang gak bisa membedakan mana harum dan mana bau busuk

Visualisasi Data

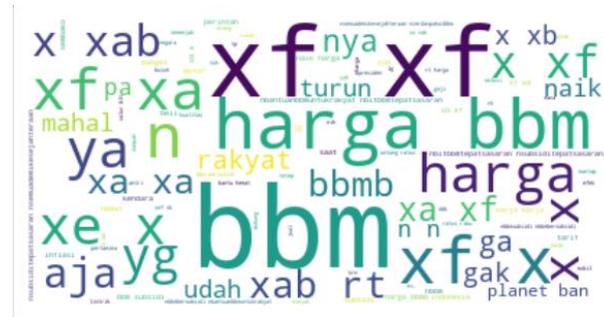
Visualisasi data dilakukan melalui diagram batang dan word cloud. Grafik plot bar menunjukkan hasil analisis sentimen berupa diagram batang yang berisi total sentimen positif dan negatif. Gambar 7 menunjukkan diagram batang yang tidak seimbang antara sentimen positif dan negatif, dimana sentimen negatif lebih dominan [15]. Dengan melihat visualisasi ini, diketahui masyarakat cenderung mengekspresikan lebih banyak sentimen negatif terkait kebijakan kenaikan harga BBM. Analisis ini memberikan gambaran visual tentang persepsi dominan dalam masyarakat terhadap kebijakan tersebut, yang dapat menjadi landasan untuk pemahaman lebih lanjut dalam penentuan kebijakan pemerintah berikutnya. Visualisasi ini memberikan wawasan yang lebih dalam dan mudah dicerna terkait

dengan respons masyarakat terhadap kebijakan pemerintah, yang dapat menjadi dasar untuk pengambilan keputusan lebih lanjut.



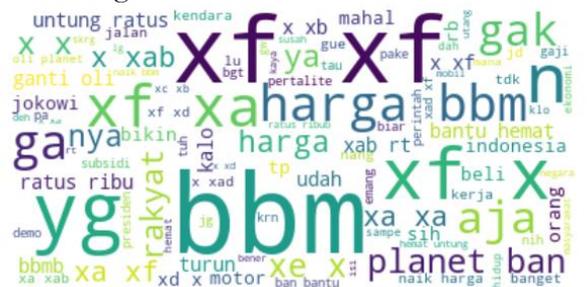
Gambar 7. Visualisasi Diagram Batang

Visualisasi data dengan Word Cloud digunakan untuk menyajikan kata yang paling dominan muncul. Kata dominan muncul akan lebih besar dari kata yang jarang muncul. Pada Gambar 8 Kata-kata dominan yang muncul dalam sentimen positif dari visualisasi data Word Cloud adalah "harga, bbm, turun, mahal, rakyat" sebagai istilah yang paling banyak dibicarakan.



Gambar 8. Visualisasi World Cloud Sentiment Positif

Di sisi lain, pada Gambar 9, kata-kata yang dominan muncul dalam sentimen negatif dari visualisasi data menggunakan Word Cloud adalah kata-kata “gak, kalo, lu, planet” sebagai kata yang paling banyak diperbincangkan.



Gambar 9. Visualisasi World Cloud Sentiment Negatif

Validasi

Sebelum masuk ke pembuatan model, perlu melakukan proses validasi guna untuk mengukur dan mengevaluasi kinerja algoritma. Cara kerja validasi adalah dengan membagi data menjadi dua subset: data test dan data train dengan proporsi 80% dan 20%. Pada proses ini menggunakan dataset yang sudah dibersihkan sebelumnya dan pada penelitian ini untuk hasil akhir menggunakan data uji.

Pengujian

Setelah melalui semua tahapan yang ada mulai dari tahap awal hingga tahap akhir, maka didapatkan hasil nilai akurasi dari kedua metode yang digunakan. Untuk metode *support vector machine* didapatkan nilai akurasi sebesar 77%, sementara untuk metode *Random Forest* didapatkan hasil akurasi sebesar 76%, seperti terlihat pada Gambar 10 dan Gambar 11.

Train report				
	precision	recall	f1-score	support
negative	0.91	0.99	0.95	1497
positive	0.97	0.70	0.81	487
accuracy			0.92	1984
macro avg	0.94	0.85	0.88	1984
weighted avg	0.93	0.92	0.92	1984

Test report				
	precision	recall	f1-score	support
negative	0.80	0.93	0.86	374
positive	0.56	0.28	0.37	122
accuracy			0.77	496
macro avg	0.68	0.60	0.61	496
weighted avg	0.74	0.77	0.74	496

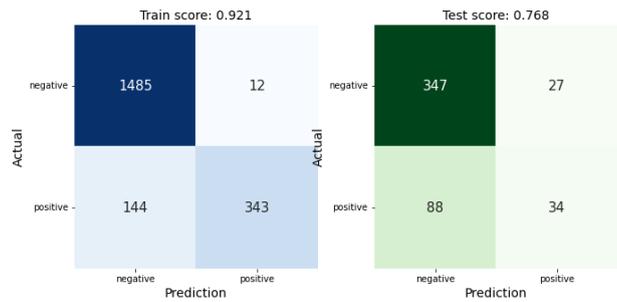
Gambar 10. Hasil Akurasi SVM

Train report				
	precision	recall	f1-score	support
negative	0.97	1.00	0.98	1497
positive	0.99	0.89	0.94	487
accuracy			0.97	1984
macro avg	0.98	0.95	0.96	1984
weighted avg	0.97	0.97	0.97	1984

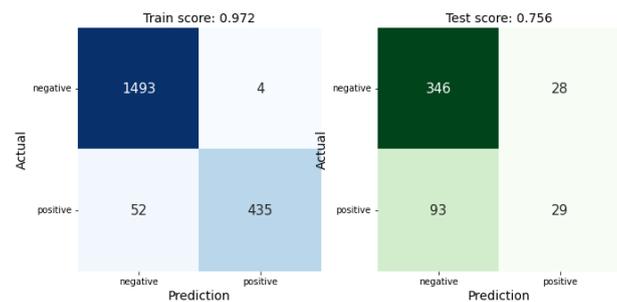
Test report				
	precision	recall	f1-score	support
negative	0.79	0.93	0.85	374
positive	0.51	0.24	0.32	122
accuracy			0.76	496
macro avg	0.65	0.58	0.59	496
weighted avg	0.72	0.76	0.72	496

Gambar 11. Hasil Akurasi RF

Adapun hasil dari kedua metode tersebut dalam bentuk confusion matrix plot dapat dilihat pada Gambar 12 dan Gambar 13.



Gambar 12. Plot Akurasi SVM



Gambar 13. Plot Akurasi RF

Untuk hasil akhir menunjukkan dimana algoritma *support vector machine* (svm) lebih unggul dibandingkan algoritma *Random Forest* dengan hasil akurasi pada *support vector machine* sebesar 77% dan algoritma *Random Forest* sebesar 76%.

Tabel 7. Hasil Akurasi

Algoritma	Sentiment	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	Positif	0.77	0.56	0.28	0.37
	Negatif		0.80	0.93	0.86
RF	Positif	0.76	0.51	0.24	0.32
	Negatif		0.79	0.93	0.85

Pada penelitian sebelumnya yang ditulis oleh Akhmad Miftahusalam, dkk membahas perbandingan algoritma *Random Forest*, dan *support vector machine* mengenai opini masyarakat terhadap penghapusan tenaga honorer dan memperoleh hasil akurasi 62% untuk *Random Forest*, dan 59% untuk *support vector machine*. Sehingga dapat dipastikan *Random Forest* lebih unggul dari *support vector machine* [16]. Oleh karena itu

berdasarkan dari penelitian acuan terdahulu, yang dimana algoritma *Random Forest* memiliki akurasi lebih baik dari *support vector machine* pada klasifikasi sentimen analisis, sehingga pada penelitian ini tertarik untuk membandingkan algoritma *support vector machine* dan *Random Forest*, dimana pada hasil akurasi penelitian ini menyatakan bahwa algoritma *support vector machine* lah yang lebih unggul, dengan beberapa kesimpulan dibawah ini dan dapat dilihat pada Tabel 7 diatas:

- 1) Kedua algoritma yang digunakan dalam penelitian ini yaitu *support vector machine* dan *Random Forest* keduanya bekerja dengan baik.
- 2) Hasil akurasi pada algoritma *support vector machine* sebesar 77% sedangkan algoritma *Random Forest* sebesar 76%
- 3) Dari hasil akurasi menunjukkan algoritma *support vector machine* lebih unggul dari algoritma *Random Forest*.
- 4) Dapat dikatakan persepsi masyarakat pada penelitian ini lebih dominan bersifat negatif

4. Kesimpulan

Penelitian ini mengkaji perbandingan antara algoritma Support Vector Machine (SVM) dan *Random Forest* dalam melakukan analisis sentimen terhadap kebijakan pemerintah Indonesia terkait kenaikan harga Bahan Bakar Minyak (BBM) pada tahun 2022. Keputusan pemerintah untuk mengurangi subsidi BBM sebagai dampak dari kenaikan harga minyak dunia memicu respons dan kritik masyarakat di berbagai media sosial. Tujuan penelitian ini adalah untuk menentukan algoritma yang memiliki tingkat akurasi lebih baik serta memberikan masukan terhadap pandangan masyarakat terhadap kebijakan tersebut. Hasil penelitian menunjukkan bahwa baik algoritma SVM maupun *Random Forest* memberikan kinerja yang baik, dengan akurasi masing-masing sebesar 77% dan 76%. Meskipun perbedaan akurasi tidak signifikan, SVM memiliki tingkat akurasi yang sedikit lebih tinggi. Analisis sentimen ini melibatkan pengumpulan data dari media sosial Twitter dengan kata kunci "bbm naik". Setelah proses pre-processing data, termasuk case folding, cleansing, tokenization, remove punctuation, dan stopword removal, data dilabeli dan dibagi menjadi data latih dan data uji. Visualisasi data menggunakan diagram batang dan Word Cloud memperlihatkan dominasi sentimen

negatif dalam respons masyarakat terhadap kenaikan harga BBM. Selanjutnya, proses validasi menggunakan metode *K-fold Cross Validation* membantu memastikan kinerja model. Evaluasi dilakukan dengan confusion matrix, dan hasil akhir menunjukkan bahwa SVM memiliki akurasi yang sedikit lebih tinggi dibandingkan dengan *Random Forest*. Penelitian ini memberikan pemahaman lebih mendalam tentang persepsi masyarakat terhadap kebijakan pemerintah melalui analisis sentimen menggunakan algoritma SVM dan *Random Forest*. Meskipun keduanya memberikan hasil yang baik, SVM lebih unggul dalam hal akurasi. Penemuan ini dapat memberikan masukan berharga bagi pemerintah untuk memahami opini publik dan meningkatkan interaksi dengan masyarakat terkait kebijakan yang dilaksanakan.

5. Daftar Pustaka

- [1] Kurniasih, U. and Suseno, A.T., 2022. Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM). *Jurnal Media Informatika Budidarma*, 6(4), pp.2335-2340. DOI: <http://dx.doi.org/10.30865/mib.v6i4.4958>.
- [2] Purbolaksono, M.D., Tantowi, M.I., Hidayat, A.I. and Adiwijaya, A., 2021. Perbandingan support vector machine dan modified balanced random forest dalam deteksi pasien penyakit diabetes. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(2), pp.393-399. DOI: <https://doi.org/10.29207/resti.v5i2.3008>.
- [3] Agustian, A. and Nurapriani, F., 2022. Analisis Sentimen, Text Mining Penerapan Analisis Sentimen Dan Naive Bayes Terhadap Opini Penggunaan Kendaraan Listrik Di Twitter. *Jurnal Tika*, 7(3), pp.243-249. DOI: <https://doi.org/10.51179/tika.v7i3.1550>.
- [4] Adrian, M.R., Putra, M.P., Rafialdy, M.H. and Rakhmawati, N.A., 2021. Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*, 7(1), pp. 36-40, 2021, DOI: <https://doi.org/10.26877/jiu.v7i1.7099>.

- [5] Syah, H. and Witanti, A., 2022. Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), pp.59-67. DOI: <https://doi.org/10.47080/simika.v5i1.1411>.
- [6] Lestari, S. and Saepudin, S., 2021, September. Analisis sentimen vaksin sinovac pada twitter menggunakan algoritma Naive Bayes. In *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 163-170).
- [7] Hanif, I.F., Affandi, I.R., Hasan, F.N., Sinduningrum, E. and Halim, Z., 2022. Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression. *Jurnal Linguistik Komputasional*, 5(2), pp.77-84. DOI: <https://doi.org/10.26418/jlk.v5i2.103>.
- [8] Slamet, R., Gata, W., Novtariany, A., Hilyati, K. and Jariyah, F.A., 2022. Analisis sentimen Twitter terhadap penggunaan artis Korea Selatan sebagai brand ambassador produk kecantikan lokal. *INTECOMS: Journal of Information Technology and Computer Science*, 5(1), pp.145-153. DOI: <https://doi.org/10.31539/intecom.v5i1.3933>
- [9] Tanggraeni, A.I. and Sitokdana, M.N., 2022. Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), pp.785-795. DOI: <https://doi.org/10.35957/jatisi.v9i2.1835>.
- [10] Nitami, M.T. and Februariyanti, H., 2022. Analisis Sentimen Ulasan Ekspedisi J&T Express Menggunakan Algoritma Naive Bayes. *Jurnal Manajemen Informatika dan Sistem Informasi*, 5(1), pp.20-29. DOI: <https://doi.org/10.36595/misi.v5i1.396>.
- [11] Aruan, J.D.C., Rahayudi, B. and Ridok, A., 2022. Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(5), pp.2072-2078.
- [12] Pangaribuan, J.J. and Angkasa, V., 2022. Komparasi tingkat akurasi random forest dan knn untuk mendiagnosis penyakit kanker payudara. *Journal Information System Development (ISD)*, 7(1), pp.49-61.
- [13] Putri, N.B. and Wijayanto, A.W., 2022. Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika: Jurnal Sistem Komputer*, 11(1), pp.59-66. DOI: <https://doi.org/10.34010/komputika.v11i1.4350>.
- [14] Saputra, H.W., Rahmadden, R. and Fazri, F., Comparison of Machine Learning Algorithms in Analyzing Public Opinion Sentiments Against Fuel Price Increases. *CESS (Journal of Computer Engineering, System and Science)*, 8(1), pp.138-148.
- [15] Siswanto, D., Nijal, L. and Rajab, S., 2022. Analisa Sentimen Publik Mengenai Perekonomian Indonesia Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Metode Klasifikasi K-NN Dan Svm. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, 2(1), pp.1-9.
- [16] Miftahusalam, A., Nuraini, A.F., Khoirunisa, A.A. and Pratiwi, H., 2022, November. Perbandingan Algoritma Random Forest, Naive Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer. In *Seminar Nasional Official Statistics* (Vol. 2022, No. 1, pp. 563-572).