

Perbandingan Pelabelan Otomatis Dan Manual Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Twitter Menggunakan Algoritma *Support Vector Machine*

Ikhlasul Amal¹, Jayanta²

S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional “Veteran” Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

ikhlasula@upnvj.ac.id¹, jayanta@upnvj.ac.id²

Abstrak. Pada 3 September 2022, pemerintah resmi mengumumkan kenaikan harga BBM. Berbagai tanggapan dan keluhan masyarakat ditumpahkan lewat media sosial *Twitter* dan *trending topic*. Tentu banyak sekali data tersebut yang harus ditampung, maka dari itu diperlukan analisis sentimen untuk mengetahui sentimen pengguna dan perbandingan performa algoritma *Support Vector Machine (SVM)* menggunakan data dari pelabelan otomatis dan manual, melakukan pra-proses untuk mengubah data yang tidak terstruktur menjadi terstruktur. Data tersebut diberi label positif dan negatif secara otomatis menggunakan *Lexicon Based* dan secara manual oleh 2 anotator dengan perhitungan *Kappa Statistic*. Pembobotan kata menggunakan *Term Frequency – Inverse Document Frequency*. Data dibagi menjadi 80:20 data latih dan data uji. Hasil klasifikasi dari data yang dilabelkan secara otomatis dan manual menggunakan algoritma *Support Vector Machine (SVM)* masing-masing memiliki nilai akurasi sebesar 0.83 dan 0.81, presisi sebesar 0.86 dan 0.84, *recall* sebesar 0.92 dan 0.91, *specificity* sebesar 0.58 dan 0.56, dan *F1-Score* sebesar 0.89 dan 0.88.

Kata Kunci: Kenaikan Harga BBM, *Twitter*, Analisis Sentimen, *Support Vector Machine (SVM)*

1 Pendahuluan

Keputusan resmi pemerintah pada 3 September 2022 untuk menaikkan harga Bahan Bakar Minyak (BBM) Solar, Pertalite, dan Pertamax menjadi Rp. 6.800 per liter untuk Solar, Rp. 10.000 per liter untuk Pertalite, dan Rp. 14.500 per liter untuk Pertamax. Bahan Bakar Minyak (BBM) merupakan produk yang berperan penting dalam setiap pergerakan moneter [1]. Saking pentingnya dalam seluruh aktivitas ekonomi, kenaikan harga BBM mendapatkan respons yang sangat cepat dari kalangan masyarakat. Berdasarkan tulisan yang dibuat oleh Suryo di inet.detik.com pada September 2022 bahwa pada hari Sabtu, 3 September 2022, pengumuman kenaikan harga BBM diberitahukan pada pukul 13.00 WIB. Sementara itu kenaikan harga dilakukan pada pukul 14.30 WIB. Kenaikan harga yang mendadak membuat netizen menjerit. Terlebih karena mereka hanya punya sedikit waktu untuk penyesuaian harga yang baru. Berbagai respons dan keluhan masyarakat pun ditumpahkan, salah satunya lewat *Twitter* [2].

Berdasarkan data tersebut tidak heran sekarang *twitter* menjadi tempat untuk memberikan kritik, keluhan, dan saran dalam menanggapi keadaan yang sedang menjadi perbincangan kepada *tweet* perusahaan-perusahaan terkait. Para pengguna selaku masyarakat yang sekarang ini banyak sekali melakukan *tweet* di *twitter* dengan kata kunci atau *hashtag/tagar* (#) BBM Naik, jelas ada banyak data yang perlu dipertimbangkan. Oleh karena itu, analisis sentimen diperlukan untuk mengetahui dan memperoleh opini bagaimana sentimen pengguna terhadap kenaikan harga BBM Pertamina di *twitter* serta mengetahui perbandingan performa dari algoritma *Support Vector Machine (SVM)* menggunakan data dari pelabelan otomatis dan manual. Dalam menganalisis sentimen diperlukan juga algoritma untuk menyelesaikan permasalahan ini, seperti *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine (SVM)*. Pada penelitian oleh Nurrin Muchammad Shiddieqy Hadna, Paulus Insap Santosa, dan Wing Wahyu Winarno mengenai studi literatur yang relevan tentang metode komparatif untuk proses analisis sentimen di *Twitter*. Jelas dari temuan beberapa penelitian sebelumnya bahwa algoritma *Support Vector Machine* memberikan hasil terbaik [3].

Berdasarkan penelitian terdahulu, dalam melakukan analisis sentimen dengan perbandingan beberapa algoritma seperti perbandingan *Logistic Regression* dengan *Support Vector Machine*, perbandingan *Naïve Bayes* dengan *Support Vector Machine*. Algoritma *Support Vector Machine* memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma yang lain. Oleh sebab itu, pada penelitian ini algoritma yang digunakan untuk menganalisis sentimen terhadap kenaikan BBM Pertamina di *Twitter* adalah *Support Vector Machine* dengan

data yang dilabelkan secara otomatis menggunakan *Lexicon Based* dan secara manual oleh 2 anotator menggunakan perhitungan *Kappa Statistic*.

2 Tinjauan Pustaka

2.1 Analisis Sentimen

Analisis sentimen adalah proses yang melibatkan analisis, identifikasi, dan pengelompokan tanggapan atau opini dalam teks, dengan tujuan membedakan apakah opini tersebut bersifat positif atau negatif [4]. Dengan adanya analisis sentimen ini kita dapat melihat opini, sentimen, dan emosi seseorang terhadap isu tertentu apakah bersifat positif atau negatif.

2.2 Praproses Teks

Dalam proses *text mining*, tahapan praproses teks atau *text pre-processing* sangat diperlukan untuk membersihkan dan mempersiapkan data dari informasi yang tidak relevan. Data mentah yang digunakan dalam proses penambangan teks seringkali tidak berada dalam kondisi yang ideal untuk diproses. Oleh karena itu, praproses teks diperlukan untuk mengubah data dokumen teks yang arbitrer atau tidak terstruktur menjadi data yang terstruktur [4].

a) Pembersihan Data

Dalam praproses teks, proses pembersihan data ialah membersihkan data dari noise seperti menghilangkan *URL*, *username*, *hashtag*, akun yang di-*mention*, serta simbol-simbol sehingga data bisa diolah dengan baik dan menghasilkan informasi yang sesuai diharapkan [5].

b) Case Folding

Case folding adalah proses yang berguna ketika semua huruf dalam dokumen teks diubah menjadi huruf kecil (*lowercase*). Tidak semua dokumen teks konsisten dalam penggunaan huruf besar-kecil. Oleh karena itu, langkah *case folding* digunakan untuk melakukan penyeragaman pada seluruh kata yang terdapat dalam dokumen [4].

c) Normalization

Normalization atau normalisasi adalah proses penggunaan kamus *slangword* untuk memperbaiki kata-kata dalam dokumen teks, seperti mengoreksi kesalahan ejaan, mengubah kata-kata yang dibentuk dengan singkatan tertentu, dan menggantikan kata-kata slang atau gaul yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) [4].

d) Tokenization

Proses untuk memisahkan input *text* menjadi potongan-potongan yang bermakna disebut tokenisasi. Token merupakan sebutan dari hasil proses *Tokenization* [6].

e) Stopword Removal

Stopword removal adalah tahap di mana menghapus kata-kata yang kurang penting dari hasil tokenisasi dengan menggunakan algoritma *stoplist* dan *wordlist*. *Stopwords* adalah kata-kata umum yang biasanya muncul dalam jumlah besar dalam teks dan dianggap tidak memiliki makna atau kontribusi signifikan dalam analisis teks [6]. Tujuan dari proses ini adalah untuk meningkatkan kualitas analisis teks dengan fokus pada kata-kata yang memiliki makna atau relevansi yang lebih tinggi.

f) Stemming

Stemming adalah tahap mengubah berbagai bentuk kata menjadi bentuk kata dasar dengan menghilangkan imbuhan [6]. *Library* yang akan digunakan untuk melakukan *stemming* adalah *library* sastrawi.

2.3 Pelabelan *Lexicon Based*

Metode yang digunakan dalam melakukan pelabelan otomatis adalah *Lexicon Based* yang menggunakan kamus sebagai sumber bahasa atau kosakata. *Lexicon Based* sebagai pelabelan sebuah sentimen dari setiap opini sehingga sebuah kalimat sentimen dapat di kelas-kelaskan sesuai kelas positif dan negatif secara otomatis. Metode ini memudahkan klasifikasi dalam menganalisis data sentimen [7]. Proses pelabelan data dilakukan secara otomatis menggunakan bantuan dari *Lexicon* agar proses pelabelan menjadi lebih cepat [8].

2.4 Kappa *Statistic*

Koefisien *kappa Cohen* adalah uji statistik yang menentukan derajat kesepakatan antara dua perkiraan yang berbeda dari variabel respons. Variabel respons harus dalam bentuk kategorikal. Penilaian variabel respons dimonitor baik oleh dua penilai/pengamat yang berbeda atau oleh penilai/pengamat yang sama pada dua waktu yang berbeda [5].

Variabilitas pengamat/penilai dapat diukur dalam berbagai situasi yang melibatkan dua orang atau lebih untuk menilai hal yang sama. Pengukuran variasi antar pengamat dan penilai ini bisa disebut sebagai pengukuran persetujuan.

Perhitungan dilakukan berdasarkan perbedaan antara berapa banyak kesepakatan yang sebenarnya ada (*observed agreement*) dibandingkan dengan berapa banyak kesepakatan yang diharapkan ada (*expected agreement*).

Tabel 1. Variasi Antar-Pengamat/Penilai

		Hasil Pengamat 1		
		Positif	Negatif	Total
Hasil Pengamat 2	Positif	a	b	m ₁
	Negatif	c	d	m ₀
	Total	n ₁	n ₀	n

Penjelasan Tabel 1 diatas:

a dan d: merupakan jumlah dua pengamat setuju.

b dan c: merupakan jumlah dua pengamat tidak setuju.

Jika tidak ada perbedaan atau semuanya setuju maka b dan c bernilai nol dan kesepakatan sebenarnya (*observed agreement*) yang dinotasikan sebagai p_o adalah 1. Sedangkan, jika tidak ada kesepakatan maka a dan d bernilai nol dan kesepakatan sebenarnya (*observed agreement*) yang dinotasikan sebagai p_o adalah 0.

1. Rumus *Observed agreement*:

$$p_o = \frac{a+b}{n} \quad (1)$$

2. Rumus *Expected agreement*:

$$p_e = \left[\left(\frac{n_1}{n} \right) * \left(\frac{m_1}{n} \right) \right] + \left[\left(\frac{n_0}{n} \right) * \left(\frac{m_0}{n} \right) \right] \quad (2)$$

3. Rumus *Kappa*:

$$K = \frac{(p_o - p_e)}{1 - p_e} \quad (3)$$

Hasil perhitungan *kappa* dapat diinterpretasikan menurut Tabel 2 di bawah [5].

Tabel 2. Interpretasi Kappa

Kappa	Persetujuan
< 0	Kurang dari persetujuan
0.01 – 0.20	Persetujuan yang sedikit

0.21 – 0.40	Persetujuan yang agak sedang
0.41 – 0.60	Persetujuan yang sedang
0.61 – 0.80	Persetujuan yang substansial
0.81 – 1.00	Persetujuan yang hampir sempurna

2.5 Term Frequency – Inverse Document Frequency (TF – IDF)

Term Frequency - Inverse Document Frequency (TF-IDF) metode TF-IDF merupakan metode yang menggunakan cara pembobotan pada kata (*term*) yakni menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan sebuah kata dalam sebuah dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut. Semakin sering sebuah kata (*term*) muncul dalam suatu dokumen semakin penting pula kata tersebut [9].

Rumus *TF* (*Term Frequency*) dapat dituliskan pada persamaan (4) dibawah ini:

$$TF(t) = f_{t,d} / \sum t, d \quad (4)$$

$f_{t,d}$ merupakan sebuah kata/*term* (t) yang muncul di dalam dokumen (d).

$\sum t, d$ merupakan total keseluruhan data yang terdapat pada sebuah dokumen.

Sedangkan rumus *IDF* (*Inverse Document Frequency*) dapat dituliskan pada persamaan (5) dibawah ini:

$$IDF(t) = \log N/dft \quad (5)$$

N merupakan jumlah dokumen.

dft merupakan jumlah dokumen dimana kata (*term*) t muncul di dalamnya.

Kemudian untuk *TF-IDF* dapat dituliskan pada persamaan (6) dibawah ini:

$$TF - IDF (t, d) = TF(t) * IDF(t) \quad (6)$$

2.6 Support Vector Machine (SVM)

Support Vector Machine adalah sebuah metode pembelajaran mesin yang beroperasi berdasarkan prinsip *Structural Risk Minimization (SRM)*. *SRM* bertujuan untuk memastikan batas atas generalisasi data uji dengan mengendalikan "kapasitas" atau fleksibilitas hipotesis yang dihasilkan dari pembelajaran. Ada tiga hal yang perlu diperhatikan saat menggunakan algoritma *Support Vector Machine* [4]:

- Support Vectors*, objek data terluar pada tiap kelas terhadap *hyperplane*.
- Hyperplane*, garis pembatas yang berfungsi sebagai pemisah kelas.
- Margin*, jarak antara *support vectors* dengan *hyperplane*.

Pada *support vector machine* ketika menghadapi permasalahan kelas tidak dapat dipisahkan secara linear maka menggunakan teknik yang disebut sebagai *kernel trick* untuk mengubah ruang masukan (input *space*) menjadi ruang dimensi yang lebih tinggi (*higher dimensional space*). *Kernel trick* ini membantu untuk membangun pengklasifikasi (*classifier*) yang lebih akurat. Secara umum, ada empat jenis fungsi *kernel trick* yang dapat digunakan pada *Support Vector Machine*, yaitu [5].

- Kernel Linear*

$$K(x, x_k) = x_k^T x \quad (7)$$

- Kernel Polynomial*

$$K(x, x_k) = (x_k^T x + 1)^d \quad (8)$$

3. Kernel Gaussian (Radial Basis Function/RBF)

$$K(x, x_k) = \exp \left\{ -\frac{\|x - x_k\|_2^2}{\sigma^2} \right\} \quad (9)$$

4. Kernel Sigmoid

$$K(x, x_k) = \tanh [k x_k^T x + \theta] \quad (10)$$

Kernel trick tidak selalu mampu mengubah himpunan data menjadi benar-benar dapat dipisahkan secara linear. Mengklasifikasikan dua kelas yang terpisah secara non-linear, meskipun telah menggunakan *kernel trick* dan ingin memiliki toleransi terhadap derau dan pencilan, dapat dilakukan dengan menggunakan konsep *soft margin*. Konsep ini melibatkan perubahan masalah *Quadratic Programming* dengan menambahkan sebuah variabel *slack* ξ_i yang bernilai lebih dari 0, sehingga permasalahan menjadi:

$$\min \tau(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (11)$$

Dengan batasan

$$y_i (x_i \cdot w + b) \geq 1 - \xi_i \quad (12)$$

Parameter C digunakan untuk mengatur *trade-off* antara margin dan kesalahan klasifikasi dalam proses optimasi. Semakin besar nilai C, semakin besar pula penalti yang diberikan terhadap kesalahan klasifikasi. Berarti jika menggunakan parameter yang besar maka hanya sedikit data yang boleh melanggar batas *hyperplane*.

Proses klasifikasi *Support Vector Machine* dapat diformulasikan dengan persamaan berikut:

$$\begin{aligned} f(\Phi(x)) &= w \cdot \Phi(x) + b \\ &= \sum_{i=1, x_i \in SV}^n \alpha_i y_i \Phi(x) \cdot \Phi(x_i) + b \\ &= \sum_{i=1, x_i \in SV}^n \alpha_i y_i K(x, x_i) + b \end{aligned} \quad (13)$$

Dimana *SV* adalah objek data dalam training dataset yang terpilih sebagai *support vector*.

2.7 Evaluasi

Evaluasi atau ketepatan klasifikasi menggunakan *confusion matrix* untuk mengukur kinerja model klasifikasi dengan melihat nilai pada masing-masing kelas yang diprediksi dan kelas yang sebenarnya. *Confusion matrix* adalah tabel dua dimensi yang menggambarkan hasil prediksi model dengan membandingkan kelas yang diprediksi dengan kelas yang sebenarnya [4].

Tabel 3. *Confusion Matrix*

Prediksi	Aktual	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Keterangan :

True Positive (TP) : jika data aktual positif dan diprediksi positif.
 False Positive (FP) : jika data aktual negatif dan diprediksi positif.
 True Negative (TN) : jika data aktual negatif dan diprediksi negatif.
 False Negative (FN) : jika data aktual positif dan diprediksi negatif.

Terdapat rumus untuk perhitungan evaluasi, berikut adalah rumus-rumus untuk menghitung evaluasi.

a) Akurasi, untuk menghitung seberapa tepat model dalam melakukan klasifikasi. Dengan rumus:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (14)$$

b) Presisi, untuk menghitung rasio prediksi benar positif yang dibandingkan dengan total yang diprediksi positif. Dengan rumus:

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

c) *Recall*, untuk menghitung prediksi benar positif yang dibandingkan dengan keseluruhan aktual positif. Dengan rumus:

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

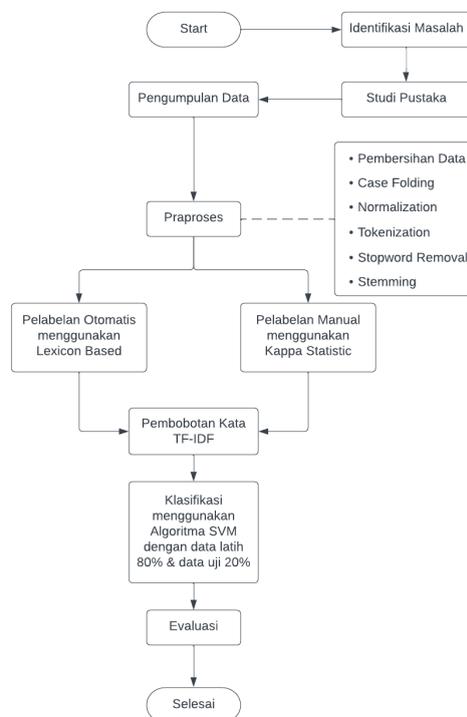
d) *Specificity*, untuk menghitung rasio prediksi benar negatif dengan keseluruhan aktual negatif. Dengan rumus:

$$Specificity = \frac{TN}{TN+FP} \quad (17)$$

e) *F1 Score*, untuk menghitung keseimbangan antara presisi dan *recall*. Dengan rumus:

$$F1\ Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (18)$$

3 Metode Penelitian



Gambar 1. Alur Penelitian

Dari Gambar 1 diatas dapat dijelaskan tahapan/alur penelitian yang dilakukan sebagai berikut.

a. Identifikasi Masalah

Pada tahapan penelitian ini permasalahan yang terjadi adalah permasalahan dalam mengklasifikasikan data tweet kenaikan harga BBM Pertamina dengan metode klasifikasi dan menggunakan algoritma *Support Vector Machine (SVM)* untuk mengetahui perbandingan performa menggunakan data pelabelan otomatis dan Manual.

b. Studi Pustaka

Pada tahapan penelitian ini dilakukan pencarian dan pembelajaran terhadap penelitian-penelitian yang telah dilakukan sebelumnya baik nasional maupun internasional serta jurnal mengenai analisis sentimen, text mining, klasifikasi, dan algoritma *Support Vector Machine (SVM)*. Adapun sumber-sumber lain yang digunakan untuk penelitian ini berupa artikel dan situs resmi, yang nantinya sumber-sumber dan penelitian-penelitian sebelumnya dapat memberikan bantuan berharga dalam menyelesaikan permasalahan yang ada dalam penelitian ini.

c. Pengumpulan Data

Pada proses ini yang dilakukan adalah melakukan pengumpulan data *tweet* dengan cara *crawling* data di twitter dengan cara menggunakan aplikasi *Docker Desktop* dan *VSCode* terminal untuk mengambil sebuah data *tweet* di *twitter* menggunakan *library* yang sudah tersedia di *python* yaitu *library twint* [10]. Data yang diambil adalah data *tweet* dengan kata kunci atau *hashtag* mengenai kenaikan harga BBM, serta berdasarkan tanggal topik tersebut sedang *trending topic* pada 3 – 29 september 2022.

d. Praproses

Pada tahapan praproses, penulis melakukan beberapa tahap, yaitu:

a) Pembersihan Data

Pada proses ini peneliti akan membersihkan data *tweet* yang sudah didapat dengan cara menghapus *URL*, *hashtag* (#), simbol-simbol, dan tanda baca sehingga data tersebut dapat diolah dengan baik.

b) Case Folding

Pada tahapan ini peneliti akan menyeragamkan karakter pada data *text* dari *tweet* yang dimana jika terdapat huruf kapital atau besar akan diubah menjadi huruf kecil.

c) Normalization

Pada tahapan ini peneliti akan merubah kata-kata yang ejaannya masih disingkat atau tidak baku akan diperbaiki supaya mempermudah dalam melakukan pemodelan *machine learning* pada penelitian ini.

d) Tokenization

Pada tahapan ini peneliti akan merubah kalimat yang masih utuh didapat dari *tweet* yang akan dipecah-pecah menjadi kata per kata dengan spasi sebagai pemisahannya.

e) Stopword Removal

Pada tahapan ini, peneliti akan melakukan penghapusan kata-kata yang tidak memiliki makna atau tidak penting untuk memastikan bahwa proses klasifikasi dapat dilakukan secara efisien dan efektif. *Stopwords* adalah kata umum yang sering muncul dalam jumlah besar dalam teks dan tidak memiliki makna yang signifikan dalam analisis teks [6].

f) Stemming

Pada proses ini peneliti melakukan proses mengembalikan kata yang berimbuhan menjadi bentuk dasar kata dengan menghapuskan semua imbuhan.

e. Pelabelan

Pada tahap pelabelan, penulis melakukan perbandingan menggunakan pelabelan otomatis dan manual, yaitu:

a) Pelabelan Otomatis Menggunakan *Lexicon Based*

Pada tahapan ini peneliti akan melakukan pelabelan pada data *tweet* menggunakan *Lexicon Based*. *Lexicon Based* merupakan fitur kata yang memiliki makna positif atau negatif berdasarkan kamus atau leksikon. Proses pelabelan data dilakukan dengan menghitung skor sentimen menggunakan kamus *lexicon based*. Setelah diketahui kata-kata yang memiliki sentimen positif dan negatif dalam sebuah kalimat, langkah selanjutnya adalah menghitung setiap kata yang mengandung sentimen dalam kalimat tersebut dengan menjumlahkan nilai opini [7]. Sentimen positif bernilai 1 atau lebih dan negatif bernilai -1 atau lebih untuk jumlah nilai opini.

b) Pelabelan Manual Perhitungan *Kappa Statistic*

Pada tahap pelabelan manual ini dilakukan dengan memberikan label pada data *tweet* dengan memberikan penilaian terhadap sentimen positif dan negatif. Pelabelan ini dilakukan secara manual oleh 2 orang yang kemudian hasil pelabelan dievaluasi menggunakan statistik *kappa* untuk mengetahui apakah pelabelan sudah baik atau belum. Perhitungan *kappa* ini akan dilakukan dengan mengacu pada bab 2 persamaan (1), (2), dan (3).

f. Pembobotan Kata *TF-IDF*

Pada tahapan ini peneliti akan melakukan pembobotan kata di setiap dokumen yang didapat dari *tweet* sehingga data tersebut memiliki nilai yang dapat membantu dalam proses pemodelan dan klasifikasi. Pembobotan kata *TF-IDF* ini akan dilakukan dengan mengacu pada bab 2 persamaan (6).

g. Klasifikasi Menggunakan Support Vector Machine

Pada tahapan ini peneliti akan melakukan proses klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*. Sebelum melakukan proses klasifikasi menggunakan algoritma *Support Vector Machine (SVM)* dataset akan dibagi menjadi data latih dan data uji. Data latih sebesar 80% dan data uji sebesar 20%. Klasifikasi *Support Vector Machine (SVM)* pada penelitian ini menggunakan *kernel Gaussian (Radial Basis Function/RBF)* karena berdasarkan penelitian yang dilakukan oleh Wiwi Widayani dan Harliana bahwa *kernel Gaussian (Radial Basis Function/RBF)* mendapatkan nilai akurasi lebih tinggi, dibandingkan dengan *Polynomial* dan *Linear* pada pengklasifikasian menggunakan algoritma *SVM* [11].

Perhitungan untuk menggunakan *kernel Gaussian (Radial Basis Function/RBF)* ini digunakan karena mengantisipasi jika dataset yang digunakan tidak dapat dipisahkan secara linear dengan menggunakan persamaan (9), setelah itu perhitungan untuk mencari *hyperplane* optimum dapat menggunakan persamaan (11), lalu perhitungan proses klasifikasi *SVM* menggunakan persamaan (13). Dalam penentuan parameter yang ada pada *kernel Gaussian (Radial Basis Function/RBF)* yaitu menggunakan *Grid Search CV* [12].

h. Evaluasi

Pada tahapan ini peneliti akan melakukan evaluasi terhadap algoritma yang digunakan yaitu *Support Vector Machine (SVM)* yang dilakukan menggunakan tabel *confusion matrix* dan menghitung nilai akurasi, presisi, *recall*, *specificity*, dan *f1-score*. Menggunakan perhitungan yang dilakukan berdasarkan persamaannya yaitu pada persamaan (14), (15), (16), (17), dan (18).

4 Hasil dan Pembahasan

Data yang digunakan pada penelitian ini berupa *tweet* dari pengguna yang diambil melalui platform sosial media *Twitter*. Proses pengumpulan data dilakukan dengan menggunakan *library python* yaitu *twint* yang dijalankan pada aplikasi *Virtual Studio Code* dengan bantuan *Dockerfile*. Dalam pengambilan data, data yang diambil merupakan isi *tweet* pengguna dari tanggal 3 September – 29 September 2022 tentang kenaikan harga BBM Pertamina.

Data yang diambil merupakan data *tweet* dengan kata kunci kenaikan harga bbm atau bbm naik dan data yang didapat sebanyak 6005 data. Hasil dari pengambilan data dapat dilihat pada Gambar 2.

	created_at	date	time	username	tweet
0	2022-09-28 23:28:37 UTC	2022-09-28	23:28:37	ameeranti	@juli_sma53 @CNNIndonesia Apalagi sejak zaman ...
1	2022-09-28 20:20:36 UTC	2022-09-28	20:20:36	terkinidotid	Harga BBM Naik, AHM Bagi Tips Merawat Sejumlah...
2	2022-09-28 16:25:14 UTC	2022-09-28	16:25:14	sunawanajiprsty	Ni gimn to biasanya menjalang kenaikan bbm mi...
3	2022-09-28 16:24:35 UTC	2022-09-28	16:24:35	ceceplagi	@Nada10062503 @jokowi Sejak plandemit aja tari...
4	2022-09-28 15:52:22 UTC	2022-09-28	15:52:22	rmoceh	@Nada10062503 @jokowi Kata supirnya kenaikan o...
...
6000	2022-09-03 12:41:59 UTC	2022-09-03	12:41:59	wongbiasa1818	@detikcom @jokowi Ini namanya bukan kenaikan h...
6001	2022-09-03 12:38:48 UTC	2022-09-03	12:38:48	chuplizz	Ketakuan masyarakat bukan hanya pada kenaikan ...
6002	2022-09-03 12:36:43 UTC	2022-09-03	12:36:43	winarsodki	Pertama, kenaikan BBM tersebut akan menurunkan...
6003	2022-09-03 12:34:28 UTC	2022-09-03	12:34:28	winarsodki	Partai Buruh kembali menegaskan penolakannya t...
6004	2022-09-03 12:31:40 UTC	2022-09-03	12:31:40	p4pikris	@Ekobudi31842141 @detikcom Naah mari tengok se...

6005 rows x 5 columns

Gambar 2. Hasil Crawling Data Twitter

Data sebanyak 6005 merupakan data mentah yang tidak terstruktur, maka dari itu harus dilakukan praproses untuk merubah data menjadi data yang terstruktur. Seperti pada Tabel 4 di bawah.

Tabel 4. Sampel Tweet

Tweet	Sebelum Praproses Teks	Sesudah Praproses Teks
-------	------------------------	------------------------

T1	@juli_sma53 @CNNIndonesia Apalagi sejak zaman pak Jokowi kenaikan gaji PN/GURU/TNI Polri sangat minim sekali. Berbanding terbalik dgn harga BBM yang tiap tahun naik. Ada juga tujangan guru di hapus? So... Harus sabar, pemerintahan skrg belum bisa memberikan perhatian utk kesejahteraan guru/PN.	zaman jokowi naik gaji pegawai negeri guru tni polri minim banding balik harga bbm tahun tujang guru hapus sabar perintah hati sejahtera guru pegawai negeri
T2	Harga BBM Naik, AHM Bagi Tips Merawat Sejumlah Bagian Motor Biar Irit Bahan Bakar: Kenaikan harga bahan bakar minyak tentu memengaruhi pengaturan keuangan masyarakat. Selain itu, para... https://t.co/WtG2hKZU7Z #News #bengkelahassterdekat #ahasshonda #iritbbm #motoriritbbm	harga bbm tips rawat motor biar irit bahan bakar naik harga bahan bakar minyak pengaruh atur uang masyarakat berita
T3	Ni gimn to biasanya menjalang kenaikan bbm minyak langka, la ni dah naik masih ja mgantri, kadang malah g da bbm, gimana to pak bos? https://t.co/KUENlySrOq	biasa jelang naik bbm minyak langka naik antre kadang bbm bos

Setelah dilakukannya praproses dan penghapusan data duplikat, menghasilkan sebanyak 4077 data dari 6005 data yang didapatkan sebelumnya. Data tersebut akan dilabelkan secara otomatis menggunakan *lexicon based* dan manual dengan perhitungan *kappa statistic*.

Pada proses pelabelan otomatis menggunakan *lexicon based* dilakukan perubahan pada data tweet menjadi bahasa Inggris menggunakan *library googletrans* dan *translator*. Setelah itu, dilakukan perhitungan *compound score* secara otomatis menggunakan *library vadersentiment* dan dari *score* ini bisa diklasifikasikan. Jika *score* lebih besar sama dengan 0, maka kalimat tersebut positif dan diberi label 1.

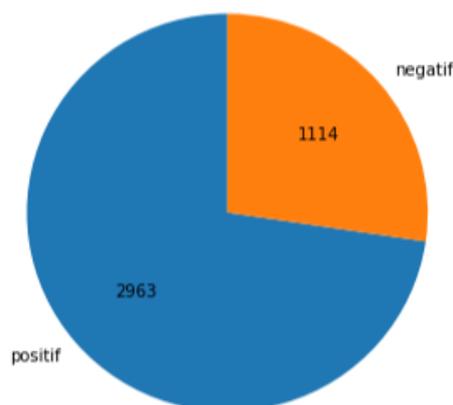
Sedangkan *score* lebih kecil dari pada 0, maka kalimat tersebut negatif dan diberi label 0. Setelah data *tweet* sudah terlabel secara otomatis, peneliti merubah kembali data tweet yang berbahasa Inggris menjadi bahasa Indonesia seperti semula. Hal ini dilakukan karena merujuk pada poin nomer 2 dari ruang lingkup penelitian ini. Hasil dari pelabelan otomatis menggunakan *lexicon based* dapat dilihat pada Tabel 5 di bawah.

Tabel 5. Sampel Pelabelan Otomatis

Sebelum	Sesudah	Compound Score	Label
zaman jokowi naik gaji pegawai negeri guru tni polri minim banding balik harga bbm tahun tujang guru hapus sabar perintah hati sejahtera guru pegawai negeri	during the Jokowi era the salary increase for civil servants the teacher the Indonesian National Police the minimum appeal to return the price of fuel the year of support the teacher remove the patient the commandment of a prosperous heart a civil servant teacher	0.9062	1
harga bbm tips rawat motor biar irit bahan bakar naik harga bahan bakar minyak pengaruh atur uang masyarakat berita	fuel prices tips on maintaining a motorbike so that it is fuel efficient rising fuel prices the influence of managing people's money news	0.4215	1
biasa jelang naik bbm minyak langka naik antre kadang bbm bos	it's normal to ride fuel, rare oil, get in line, sometimes fuel boss	0.0000	1

Data yang dihasilkan dari pelabelan menggunakan *lexicon based* didapat sebanyak 2963 sentimen positif

dilabelkan dengan angka 1 dan 1114 sentimen negatif dilabelkan dengan angka 0. Hasil persebaran label positif dan negatif dapat dilihat pada Gambar 3. total masing-masing sentimen positif dan negatif.



Gambar 3. Hasil Pelabelan Otomatis

Selanjutnya melakukan pelabelan pada data *twitter* yang sama menggunakan pelabelan manual oleh 2 anotator dengan hasil anotasi dapat dilihat pada Tabel 6 di bawah.

Tabel 6. Hasil Anotasi

Penilai	Anotator 1			Total
	Sentimen	Positif	Negatif	
Anotator 2	Positif	2822	29	2851
	Negatif	11	1215	1226
	Total	2833	1244	4077

Berdasarkan hasil anotasi pada Tabel 6, dapat dilihat bahwa anotator 1 dan anotator 2 memiliki kesamaan dalam memberikan label positif dan negatif yang masing-masing berjumlah 2822 positif dan 1215 negatif. Selain kesamaan kedua anotator dalam memberikan label juga terdapat perbedaan yaitu terdapat 29 tupel yang berbeda, ketika anotator 1 menilai negatif tetapi anotator 2 menilai positif. Selain itu, terdapat 11 tupel yang berbeda pendapat dalam memberikan label data tersebut yaitu ketika anotator 1 menilai positif tetapi anotator 2 menilai negatif. Selanjutnya menghitung *Observed agreement* (rumus (1)) dan *Expected agreement* (rumus (2)).

Perhitungan *Observed agreement*:

$$p_o = \frac{2822+1215}{4077} = 0.9901$$

Perhitungan *Expected agreement*:

$$p_e = \left[\left(\frac{2833}{4077} \right) * \left(\frac{2851}{4077} \right) \right] + \left[\left(\frac{1244}{4077} \right) * \left(\frac{1226}{4077} \right) \right] = 0.5776$$

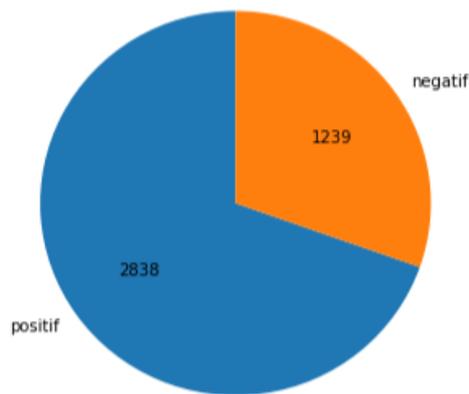
Kemudian menghitung nilai kappa untuk mengetahui reliabilitas persetujuan pelabelan, perhitungan nilai *kappa* menggunakan rumus (3).

Perhitungan *Kappa*:

$$K = \frac{(0.9901-0.5776)}{1-0.5776} = 0.9765$$

Dari hasil perhitungan kappa di atas, didapat nilai kappa sebesar 0.9765. Berdasarkan Tabel 2 Interpretasi *Kappa* menyatakan bahwa rentang nilai 0.81 – 1.00 memiliki tingkat reliabilitas persetujuan yang hampir sempurna. Jika dilihat nilai *kappa* dari hasil perhitungan dan dicocokkan dengan interpretasi *kappa*, maka dapat dikatakan hasil perhitungan nilai *kappa* di atas hampir sempurna dan dapat disimpulkan bahwa hasil pelabelan secara manual yang dilakukan oleh kedua anotator dapat digunakan untuk pemodelan.

Perbedaan dalam memberikan label tersebut kemudian didiskusikan kembali oleh anotator 1 dan anotator 2. Dari hasil diskusi tersebut 29 tupel yang anotator 1 menilai negatif sedangkan anotator 2 menilai positif telah mencapai kesepakatan yaitu 15 negatif dan 14 positif. Pada 11 tupel yang anotator 1 yang menilai positif anotator 2 menilai negatif telah mencapai kesepakatan yaitu 9 negatif dan 2 positif. Sehingga hasil dari pelabelan secara manual didapatkan 2838 positif dan 1239 negatif. Hasil persebaran label positif dan negatif dapat dilihat pada Gambar 4. total masing-masing sentimen positif dan negatif.



Gambar 4. Hasil Pelabelan Manual

Tabel 7. Perbandingan Hasil Pelabelan

Pelabelan	Positif	Negatif	Total
Otomatis	2963 (73%)	1114 (27%)	4077
Manual	2838 (70%)	1239 (30%)	data

Berdasarkan Tabel 7. Perbandingan data yang dilabelkan secara otomatis dan manual, sentimen positif lebih banyak dibandingkan sentimen negatif. Pelabelan secara otomatis pada label positif didapatkan 3% lebih banyak dibandingkan label positif dari pelabelan manual, sedangkan pada label negatif didapatkan 3% lebih banyak pada pelabelan manual dibandingkan label negatif dari pelabelan otomatis.

Setelah dilakukan pelabelan data secara otomatis dan manual, data hasil praproses yaitu 4077 data perlu dilakukan perubahan data dari kata menjadi angka supaya bisa diproses oleh model *machine learning* karena model *machine learning* hanya menerima masukan berupa angka-angka. Oleh karena itu, data teks perlu diubah menjadi angka dengan salah satu teknik yaitu pembobotan dengan *Term Frequency – Inverse Document Frequency (TF-IDF)*. Pembobotan *TF-IDF* yang dilakukan berdasarkan rumus (6) melalui perkalian rumus (4) dan rumus (5). Berikut tabel sampel nilai *TF-IDF* dari tiga kata yang dipilih yaitu antre, harga, minyak :

Tabel 8. Sampel Hasil *TF-IDF*

Kata	TF			df	IDF	TF * IDF		
	D1	D2	D3			D1	D2	D3
antre	0	0	0.090	1	0.477	0	0	0.043
harga	0.04	0.105	0	2	0.176	0.007	0.018	0
minyak	0	0.053	0.090	2	0.176	0	0.009	0.016

Adapun pengklasifikasian yang dilakukan menggunakan algoritma *Support Vector Machine (SVM)* dengan *kernel Gaussian (Radial Basis Function/RBF)*. *Kernel Gaussian (Radial Basis Function/RBF)* memiliki parameter *cost (C)* dan *gamma (γ)*.

Parameter *cost (C)* dalam algoritma *Support Vector Machine (SVM)* memainkan peran penting dalam mengatur keseimbangan antara margin dan kesalahan klasifikasi. Semakin besar nilai *cost (C)*, semakin keras sanksi yang diberikan terhadap kesalahan klasifikasi yang berarti jika parameter *cost (C)* besar maka hanya sedikit data yang boleh melanggar batas *hyperplane*.

Parameter *gamma (γ)* berfungsi untuk menentukan banyaknya lengkungan pada batasan keputusan (*decision boundary*), semakin besar nilai *gamma (γ)* maka semakin banyak lengkungan. Sedangkan, semakin kecil nilai *gamma (γ)* maka semakin sedikit lengkungan.

Dalam menentukan parameter *cost (C)* dan *gamma (γ)* pada *kernel Gaussian (Radial Basis Function/RBF)* menggunakan *hyperparameter tuning Grid Search CV*. *Hyperparameter tuning* tersebut akan mencari parameter *cost (C)* dan *gamma (γ)* satu per satu untuk mendapatkan nilai parameter terbaik untuk model algoritma *Support Vector Machine (SVM)*.

Sebelum dilakukan klasifikasi, dataset dibagi menjadi 80% data latih dan 20% data uji. Seperti pada Tabel 9 dibawah ini.

Tabel 9. Pembagian Data

	80% (Data latih)	20% (Data uji)	Total
Dataset	3261	816	4077

Pada penelitian ini, pencarian parameter *cost* (C) dan *gamma* (γ) dengan menentukan jangkauan parameter, yaitu $C = [0.1, 1, 10, 100, 1000]$ dan $\gamma = [0.1, 1, 10, 100, 1000]$. Proses *Grid Search CV* dalam pencarian parameter *cost* (C) dan *gamma* (γ) seperti pada Tabel 10.

Tabel 10. Proses *Grid Search CV*

		<i>gamma</i> (γ)					
		Nilai	0.1	1	10	100	1000
<i>cost</i> (C)	0.1	(0.1,0.1)	(0.1,1)	(0.1,10)	(0.1,100)	(0.1,1000)	
	1	(1,0.1)	(1,1)	(1,10)	(1,100)	(1,1000)	
	10	(10,0.1)	(10,1)	(10,10)	(10,100)	(10,1000)	
	100	(100,0.1)	(100,1)	(100,10)	(100,100)	(100,1000)	
	1000	(1000,0.1)	(1000,1)	(1000,10)	(1000,100)	(1000,1000)	

Hasil dari proses *Grid Search CV* pada data pelabelan otomatis dan pelabelan manual, sama-sama mendapatkan parameter terbaik yaitu $C = 10$ dan $\gamma = 0.1$. Hasilnya didapat menggunakan kode program python dibawah ini.

```
# menentukan jangkauan parameter
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [0.1, 1, 10, 100, 1000],
              'kernel': ['rbf']}

grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3, cv = 5)

grid.fit(X_train, y_train)
print(grid.best_params_)
```

Dapat dilihat bahwa kode diatas merupakan perintah untuk mencari parameter C dan *gamma* yang optimal menggunakan *Grid Search CV* dengan algoritma *Support Vector Machine*, hasil parameter terbaik dapat dilihat pada Gambar 5 di bawah.

```
{'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}
```

Gambar 5. Hasil Parameter Terbaik *Grid Search CV*

Setelah dilakukan klasifikasi terhadap kedua data yang dilabelkan secara otomatis dan manual dengan data dibagi menjadi 80% data latih dan 20% data uji, selanjutnya dilakukan evaluasi untuk mengetahui bagaimana performa/kinerja model yang telah dilakukan sebelumnya.

Hasil dari klasifikasi terhadap data yang dilabelkan secara otomatis kelas prediksi terdapat 644 positif dan 172 negatif, sedangkan kelas sebenarnya terdapat 606 positif dan 210 negatif. Berikut tabel *confusion matrix* hasil pelabelan otomatis.

Tabel 11. *Confusion Matrix* Hasil Pelabelan Otomatis

		Actual Class	
		Positive	Negative
Predicted Class	Positive	555	89
	Negative	51	121

Hasil dari klasifikasi terhadap data yang dilabelkan secara manual kelas prediksi terdapat 634 positif dan 182 negatif, sedangkan kelas sebenarnya terdapat 582 positif dan 234 negatif. Berikut tabel *confusion matrix* hasil pelabelan manual.

Tabel 12. *Confusion Matrix* Hasil Pelabelan Manual

		Actual Class	
		Positive	Negative
Predicted Class	Positive	532	102
	Negative	50	132

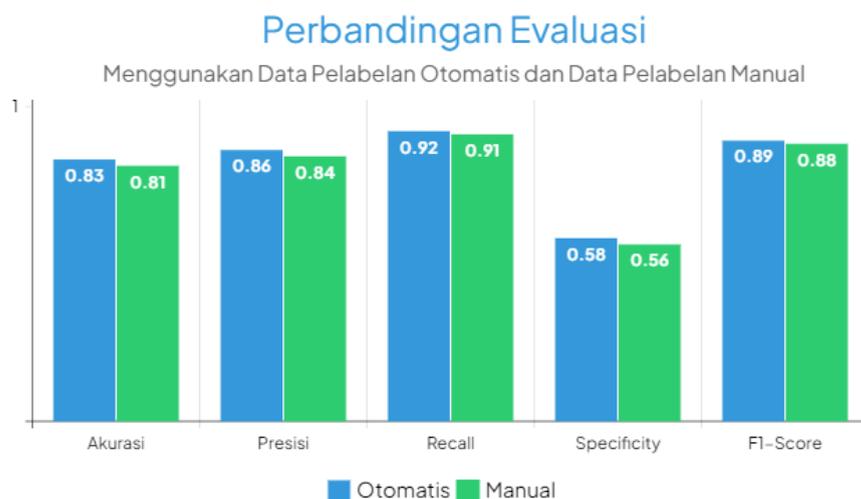
Pada Tabel 11 dan Tabel 12 diatas dapat dilihat *True Positive* atau kelas yang diprediksi sesuai dengan kelas yang sebenarnya terdapat masing-masing 555 dan 532, *False Positive* atau kelas yang diprediksi positif padahal kelas sebenarnya negatif terdapat 89 dan 102, *False Negative* atau kelas yang diprediksi negatif padahal kelas sebenarnya positif terdapat masing-masing 51 dan 50, *True Negative* atau kelas yang diprediksi negatif sesuai dengan kelas sebenarnya terdapat masing-masing 121 dan 132.

Rumus perhitungan evaluasi yang digunakan pada penelitian ini adalah *Accuracy*, *Precision*, *Recall*, *Specificity*, *F1-score*. Perhitungan dari masing-masing evaluasi tersebut menggunakan rumus (14), (15), (16), (17), (18).

Tabel 13. Hasil Perbandingan Evaluasi menggunakan Data Pelabelan Otomatis dengan Pelabelan Manual

Data	Akurasi	Presi	Recall	Specificity	F1-Score
Pelabelan Otomatis	$= \frac{TP+TN}{TP+FN+FP+TN} = \frac{555+121}{555+51+89+}$	$= \frac{TP}{TP+FP} = \frac{5}{555}$	$= \frac{TP}{TP+FN} = \frac{5}{555}$	$= \frac{TN}{TN+FP} = \frac{121}{121+}$	$= 2 \cdot \frac{Precision \times Recall}{Precision+Recall} = 2 \cdot \frac{0.86 \times 0.92}{0.86+0.92}$
Pelabelan Manual	$= \frac{TP+TN}{TP+FN+FP+TN} = \frac{532+132}{532+50+102+}$	$= \frac{TP}{TP+FP} = \frac{53}{532+}$	$= \frac{TP}{TP+FN} = \frac{53}{532+}$	$= \frac{TN}{TN+FP} = \frac{13}{132+}$	$= 2 \cdot \frac{Precision \times Recall}{Precision+Recall} = 2 \cdot \frac{0.84 \times 0.91}{0.84+0.91}$

Berdasarkan perhitungan akurasi, presisi, *precision*, *recall*, *specificity*, dan *f1-score* dapat dilihat bahwa masing-masing dari data pelabelan otomatis dan manual memiliki nilai akurasi sebesar 0.83 dan 0.81, presisi sebesar 0.86 dan 0.84, *recall* sebesar 0.92 dan 0.91, *specificity* sebesar 0.58 dan 0.56, dan *F1-score* sebesar 0.89 dan 0.88.

**Gambar 6.** Perbandingan Hasil Evaluasi

Berdasarkan Gambar 6. Dapat dilihat bahwa performa algoritma *support vector machine* menggunakan data yang dilabelkan secara otomatis sedikit lebih baik daripada data yang dilabelkan secara manual.

5 Kesimpulan dan Saran

5.1 Kesimpulan

- Menganalisis sentimen pengguna/masyarakat di *Twitter* terhadap kenaikan harga bbm pertama berdasarkan data yang diambil dari *Twitter*, sentimen positif lebih banyak dibandingkan sentimen negatif, terdapat 2963 positif dan 1114 negatif pada pelabelan otomatis, terdapat 2838 positif dan 1239 negatif pada pelabelan manual.
- Pelabelan otomatis menggunakan *lexicon based* dan pelabelan manual oleh 2 anotator dapat mempengaruhi kinerja dari performa algoritma *Support Vector Machine*. Data yang dilabelkan secara otomatis menggunakan *lexicon based* mendapatkan performa yang lebih baik dibandingkan pelabelan secara manual, untuk nilai akurasi terdapat perbedaan sebesar 0.02, untuk presisi sebesar 0.02, untuk *recall* sebesar 0.01, untuk *specificity* sebesar 0.02, dan untuk *F1-Score* sebesar 0.01.

5.2 Saran

- Menambahkan 2 anotator lagi dalam melabelkan data secara manual untuk membandingkan hasil performa algoritma *support vector machine* dari data pelabelan manual yang sudah dilakukan pada penelitian ini.

- b) Menggunakan pelabelan otomatis yang lainnya untuk membandingkan pelabelan secara otomatis menggunakan *lexicon based* yang sudah dilakukan pada penelitian ini.
- c) Menambahkan kategori/kelas netral pada pelabelan untuk perbandingan performa dengan kategori/kelas yang sudah dilakukan pada penelitian ini.

Referensi

- [1] Yuliani, D., Saryono, S., Apriani, D., Maghfiroh., & Ro, M. (2022). Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok (Sembako) Di Kecamatan Tambun Selatan Dalam Masa Pandemi. *Jurnal Citizenship Virtues*, 2022, 2(2), 320-326.
- [2] Suryo, A. (2022). BBM Naik, Pertamina Masih Trending di Twitter. Diambil tanggal 20 September 2022, dari <https://inet.detik.com/cyberlife/d-6272319/bbm-naik-pertamina-masih-trending-di-twitter>
- [3] Hadna, N., M., S., Santosa, P., I., & Winarno, W., W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016).
- [4] Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya* (Vol. 2, No. 2, pp. 64-74).
- [5] Arjuna, R., M. (2021). Analisis Sentimen Terhadap Layanan Provider Telkomsel Pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine. Diambil tanggal 27 September 2022, dari <https://repository.upnvj.ac.id/11199/>
- [6] Kelvin., Banjarnahor, J., Indra, E., & Sinurat, S., H. (2022). Analisis Perbandingan Sentimen Corona Virus Disease-2019 (COVID19) Pada Twitter Menggunakan Metode Logistic Regression dan Support Vector Machine (SVM). *JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)*. Vol.5, No. 2.
- [7] Mahendrajaya, R., Buntoro, G., H., & Setyawan, M., B. (2019). Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based dan Support Vector Machine. *KOMPUTEK: Jurnal Teknik Universitas Muhammadiyah Ponorogo*, 3(2), 2019: 52-63.
- [8] Mujahidin, S., Prasetio, B., & Utomo, M., C., C. (2022). Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian Naïve Bayes. *Jurnal Vocational Teknik Elektronika dan Informatika*. Vol. 10, No.3.
- [9] Hariyby, E., Hoiriyah, & Walid, M. (2022). Twitter Text Mining Mengenai Isu Vaksinasi Covid-19 Menggunakan Metode Term Frequency, Inverse Document Frequency (TF-IDF). *Jurnal Mahasiswa Teknik Informatika (JATI)*. Vol. 6, No. 2.
- [10] Satria, H. (2023). Cara Berhasil Mendapatkan Data Tweet dari Twitter — 2023. Diambil tanggal 2 Februari 2023, dari <https://medium.com/@helimisatria/cara-mendapatkan-data-twitter-di-2023-44aaefb616ea>
- [11] Widayani, W., & Harliana. (2021). Perbandingan Kernel Support Vector Machine Dalam Melakukan Klasifikasi Penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains dan Informatika*. Vol. 7, No. 1.
- [12] Yao, L., Fang, Z., Xiao, Y., Hou, J., & Fu, Z. (2021). An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*, 214, 118866.