

Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga BBM dengan Metode Support Vector Machine

Siti Nurhaliza*, Yusra, Muhammad Fikry

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950125222@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammadfikry@uin-suska.ac.id

Email Penulis Korespondensi: 11950125222@students.uin-suska.ac.id

Submitted: 11/06/2023; Accepted: 30/06/2023; Published: 30/06/2023

Abstrak—Kenaikan harga Bahan Bakar Minyak (BBM) di Indonesia selalu menjadi kontroversi yang bisa dilihat dari media-media online seperti Twitter yang berpengaruh terhadap perekonomian Indonesia, dengan adanya masalah tersebut hal itu memberikan dampak perubahan pada ketidakstabilan biaya akan kenaikan harga BBM yang juga akan berpengaruh pada tingkat bertambahnya biaya transportasi dan tingkat inflasi. Pengaruh perubahan tersebut banyak mengarah pada opini-opini masyarakat yang berbeda sehingga menghasilkan pro dan kontra terhadap perubahan tersebut, dengan adanya permasalahan di atas maka proses klasifikasi sangat dibutuhkan. Penelitian ini menggunakan 3000 data tweet yang didapat dari proses crawling. Penelitian ini mendapatkan akurasi sebesar 85% pada perbandingan 90:10, untuk nilai precision sebesar 85%, recall 99% dan f1-score sebesar 91% untuk sentimen negatif, sedangkan nilai precision 83%, recall 19%, f1-score 30% untuk sentimen positif. Kemudian pada percobaan perbandingan 80:20 didapatkan akurasi sebesar 83%, untuk nilai precision sebesar 83%, recall 99% dan f1-score sebesar 91% untuk sentimen negatif, sedangkan nilai precision 82%, recall 16%, f1-score 26% untuk sentimen positif.

Kata Kunci: Support Vector Machine; Sentimen Analisis; Kenaikan Bbm; Twitter; Klasifikasi

Abstract—The increase in the price of fuel oil (BBM) in Indonesia has always been a controversy which can be seen from online media such as Twitter which has an effect on the Indonesian economy, with this problem it has a change in the impact of cost instability due to an increase in fuel prices which will also affect the rate of increase in transportation costs and the rate of inflation. The effect of these changes leads to many different public opinions so as to produce pros and cons of these changes, with the existence of the problems above, the classification process is needed. This study uses 3000 tweet data obtained from the crawling process. This study obtains an accuracy of 85% at a ratio of 90:10, for a precision value of 85%, 99% recall and 91% f1-score for negative sentiment, while 83% precision value, 19% recall, 30% f1-score for positive sentiment. Then in the 80:20 comparison experiment, an accuracy of 83% was obtained, for a precision value of 83%, a recall of 99% and an f1-score of 91% for negative sentiment, while a precision value of 82%, a recall of 16%, an f1-score of 26% for positive sentiment.

Keywords: Support Vector Machine; Sentiment Analysis; Fuel Increase; Twitter; Classification

1. PENDAHULUAN

Pada saat ini perkembangan teknologi disosial media menjadi acuan untuk mendapatkan opini publik, salah satunya adalah Twitter. Pada Google Play Store Twitter adalah salah satu aplikasi dengan jumlah download sebanyak 1 juta pengguna, berdasarkan hasil data yang didapat pada tahun 2021, Twitter berada di peringkat pertama untuk kategori MicroBlogging [1], yaitu pembuatan pesan atau post singkat pada platform penyedia informasi, banyak informasi yang bisa didapat dari platform tersebut salah satunya mengenai kenaikan harga Bahan Bakar Minyak (BBM) yang menjadi sorotan dan perbincangan masyarakat.

Kenaikan harga Bahan Bakar Minyak (BBM) di Indonesia selalu menjadi kontroversi yang bisa dilihat dari media-media online seperti Twitter yang berpengaruh terhadap perekonomian Indonesia. Pada tahun 2011 yaitu dengan harga sebesar 115,55 harga minyak mentah di Indonesia naik dengan sangat drastis pada bulan maret. Kemudian seiring berjalannya waktu, harga tersebut mengalami penurunan yang sangat signifikan yang terjadi pada tahun 2015. Ketidakstabilan akan kenaikan harga BBM tersebut akan memberikan dampak perubahan pada bertambahnya biaya transportasi yang akan berpengaruh juga pada tingkat inflasi, pengaruh perubahan tersebut akan mengarah kepada opini-opini masyarakat. Terdapat beberapa opini dan sentimen berbeda dari masyarakat yaitu yang bersifat positif maupun negatif, sehingga sentimen tersebut harus di klasifikasikan [2].

Proses klasifikasi tersebut dilakukan dengan menggunakan metode Support Vector Machine, dimana bekerja untuk mencari hyperplane dari sebuah data set dengan memaksimalkan jarak antar kelas pada setiap sentiment dan juga merupakan metode yang efisien. Support Vector Machine telah dimodifikasi untuk menyertakan fungsi kernel sehingga hasilnya dapat ditemukan dengan cepat [3]. Selain itu, metode Support Vector Machine juga merupakan model pembelajaran analisis yang memberikan hasil akurasi lebih baik dibanding dengan metode lain [4].

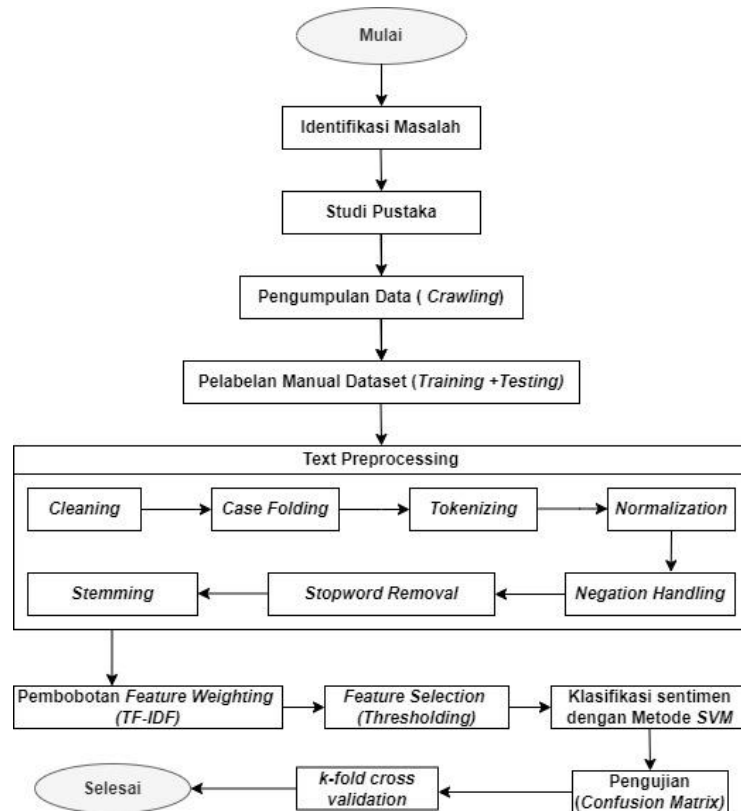
Terdapat beberapa penelitian yang membandingkan metode SVM dan K-NN mengenai analisis sentimen di Twitter, penelitian tersebut menjelaskan bahwa SVM mendapatkan nilai akurasi tertinggi sebanyak 89,7%, sementara K-NN memiliki nilai akurasi sebanyak 76,3%. [5]. Penelitian selanjutnya yang membandingkan antara metode Support Vector Machine dan Naïve Bayes yaitu mengenai kenaikan harga BBM, dimana SVM mendapatkan nilai akurasi tertinggi sebesar 90%, sedangkan Naïve Bayes memperoleh akurasi sebesar 78% [6].

Penelitian selanjutnya yang menentukan akurasi berdasarkan klasifikasi menggunakan Support Vector Machine dan Naïve Bayes, SVM memiliki nilai lebih tinggi dibanding Naïve Bayes yaitu sebesar 93%, sedangkan Naïve Bayes menghasilkan 86% [7]. Penelitian selanjutnya yang membahas mengenai sentimen analisis Twitter mengenai program vaksinasi di Indonesia menggunakan Support Vector Machine yang bertujuan untuk mengetahui tanggapan masyarakat terhadap vaksinasi covid, penelitian tersebut memperoleh hasil akurasi sebesar 90% [8]. Selanjutnya, penelitian yang membandingkan metode Naïve Bayes dan SVM mengenai Detection Traffic, SVM menghasilkan grafik yang lebih besar nilai akurasinya [9].

Oleh karena itu, dari beberapa penelitian di atas diambil kesimpulan bahwa metode Support Vector Machine memiliki nilai akurasi tertinggi dibandingkan dengan algoritma lain. Selain itu penelitian ini mendapatkan data dari Twitter dengan jumlah data sebanyak 3000 data dengan kata kunci "Kenaikan BBM". Tujuan dari penelitian ini yaitu bagaimana menerapkan metode Support Vector Machine (SVM) didalam mengklasifikasikan sentimen terhadap kenaikan harga BBM serta mengetahui tingkat akurasi dari metode tersebut.

2. METODOLOGI PENELITIAN

Pada gambar tahapan pada penelitian ini dimulai dari pengidentifikasian masalah, studi kepustakaan, mengumpulkan data dengan teknik crawling, pelabelan manual data set yang dibagi menjadi dua, selanjutnya masuk pada tahap preprocessing, kemudian melakukan pembobotan kata TF-IDF, feature selection (Thresholding), kemudian pada proses klasifikasi metode yang digunakan yaitu Support Vector Machine, selanjutnya dilakukan pengujian menggunakan confusion matrix dan k-fold cross validation.



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Pada tahap ini merupakan permasalahan utama pada penelitian analisis, dimana suatu proses merumuskan masalah untuk menemukan solusi terbaik dan mendapatkan pemahaman yang merata terhadap masalah tersebut.

2.2 Studi Pustaka

Dalam tahapan ini peneliti melakukan proses tinjauan mengenai analisis sentimen untuk membantu tercapainya hasil yang diharapkan. Maka dari itu dibutuhkanlah suatu proses untuk menyelesaikan masalah tersebut, yaitu dengan cara mengumpulkan beberapa sumber yang berguna untuk menguatkan penelitian ini.

2.3 Pengumpulan Data (Crawling)

Pada penelitian ini data didapat menggunakan teknik crawling yaitu teknik untuk mengekstrak suatu data dengan jumlah besar dari sebuah web dan di simpan dalam format excel atau csv. Penelitian ini menggunakan data

sebanyak 3000 data tweet mengenai kenaikan harga BBM berdasarkan kata kunci “Kenaikan BBM” dan “BBM Naik” pada aplikasi Twitter. Dan yang dimana data nya dikumpulkan dari 1 November 2022 sampai 1 Desember 2022.

2.4 Pelabelan Manual Dataset

Tahap pelabelan dilakukan untuk mengklasifikasikan teks ke dalam kategori untuk mendapatkan sentimen positif maupun negatif yang sesuai, pelabelan manual dataset yang dilakukan oleh validator dosen bahasa Indonesia.

2.5 Text Preprocessing

Pada tahap preprocessing dilakukan beberapa proses didalamnya yang mana data akan dibersihkan dan mengurangi volume kata pada dokumen yang dimulai dari:

1. Cleaning merupakan tahap penghapusan untuk mengurangi karakter [10]. Karakter tersebut berupa tanda baca, tanda hastag (#), tanda tanya (?), tanda seru (!), emoji, link website, username (@), dan lain sebagainya.
2. Case folding yaitu tahapan merubah huruf kapital menjadi huruf kecil [11].
3. Tokenizing yaitu proses pada sebuah data yang berguna sebagai pemecahan kalimat menjadi beberapa kata [12].
4. Normalisasi yaitu proses merubah kata sesuai KBBI yaitu kata yang dianggap tidak baku menjadi baku [13].
5. Negation handling suatu proses perubahan kata negasi menjadi kata yang sebenarnya seperti kata "tidak", "jangan", "bukan".
6. Stopword removal adalah proses penghapusan kata yang dianggap tidak berguna.
7. Stemming merupakan suatu cara pengambilan kata yang berimbunan menjadi kata dasar [13].

2.6 Pembobotan Kata (TF-IDF)

Merupakan metode yang akan digunakan pada tahap memberikan bobot pada fitur dengan mengetahui seberapa berpengaruh pada term. Tahap dalam melakukan pembobotan disebut dengan Term Frequency (TF). Kemudian pada tahap perhitungan disebut dengan Inverse Document Frequency (IDF) [14].

2.7 Feature Selection (Thresholding)

Proses yang digunakan untuk mengurangi ukuran data sehingga mendapatkan hasil yang lebih baik [15].

2.8 Klasifikasi Support Vector Machine

Support Vector Machine adalah metode untuk pengklasifikasian menggunakan support vector dan margin yang berguna untuk mencari Maximum Marginal Hyperplane [16], sehingga dapat menemukan garis pemisah (hyperplane) yang berfungsi untuk membantu memisahkan dua kelas [17]. Penelitian menggunakan kernel RBF dengan pasangan Cost (C) dan Gamma (γ) [18].

2.9 Pengujian (Confusion Matrix)

Pada tahap ini, pengujian dilakukan dengan menggunakan metode Confusion Matrix yang berguna untuk membandingkan kelas asli dan prediksi mengenai informasi nilai klasifikasi [19].

2.10 K-Fold Cross Validation

Merupakan teknik yang digunakan untuk memisahkan data training dan testing agar tidak terjadi perulangan [20].

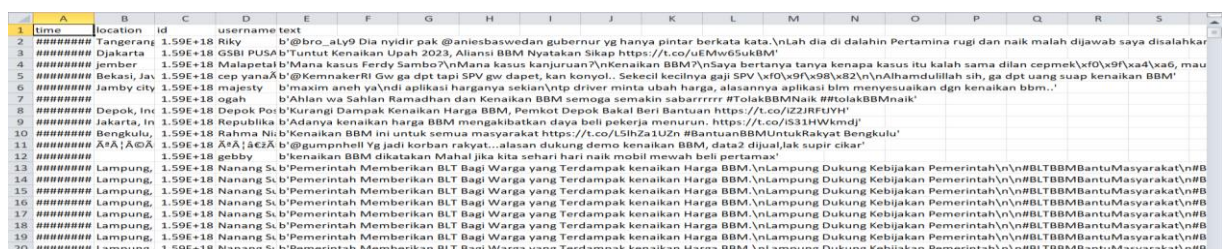
3. HASIL DAN PEMBAHASAN

3.1 Identifikasi Masalah

Proses untuk menjelaskan suatu masalah sehingga proses tersebut dapat diukur dalam penelitian.

3.2 Pengumpulan Data (Crawling)

Penelitian ini menggunakan data sentimen sebanyak 3000 data, yang diperoleh dengan proses crawling pada aplikasi Twitter kemudian disimpan dalam format xlxs. Hasil yang didapat dapat dilihat pada gambar di bawah.



1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2	#####	Tangerang	1.59E+18	Riky	b'@bro_aLy9	Dia nyidir pak @aniesbaswedan gubernur yg hanya pintar berkata kata.\nLah dia di dalahin Pertamina rugi dan naik malah dijawab saya disalahkan													
3	#####	Djakarta	1.59E+18	GSBJ PUSa	b'Tuntut Kenaikan Upah 2023, Aliansi BBM Nyatakan Sikap https://t.co/uE5Me6SukBM'														
4	#####	Jember	1.59E+18	Malapetal	b'Mana kasus Ferdys Sambo?nMana kasus kanjuruhan?nKenaikan BBM?nSaya bertanya tanya kenapa kasus itu kalah sama dilan cepmek.\xf0\x9f\xad\xad' mau														
5	#####	Bekasi, Jawa	1.59E+18	cep yana	b'@KemnakerRI Gw ga dpt tapi SPV gw dapt, kan konyol.. Sekecil kecilnya gaji SPV \xf0\x9f\x98\x82\n\nAlhamdulillah sih, ga dpt uang suap kenaikan BBM'														
6	#####	Jambiy city	1.59E+18	majesty	b'maxim aneh ya\ndi aplikasi harganya sekian\ntp driver minta ubah harga, alasannya aplikasi blm menyesuaikan dgn kenaikan bbbm.'														
7	#####		1.59E+18	ogah	b'Ahlan wa Sahlan Ramadhan dan Kenaikan BBM semoga semakin saberrrrr #TolakBBMkaik #TolakBBMnaik'														
8	#####	Depok, In	1.59E+18	Depok Pos	b'Kurangi Dampak Kenaikan Harga BBM, Pemkot Depok Bakal Beri Bantuan https://t.co/IZ2JRFUyH'														
9	#####	Jakarta, In	1.59E+18	Republika	b'Adanya kenaikan harga BBM mengakibatkan daya beli pekerja menurun. https://t.co/IS31HWkmdj'														
10	#####	Bengkulu,	1.59E+18	Rahma Nil	b'Kenaikan BBM ini untuk semua masyarakat https://t.co/L5ih2a1U2n #BantuanBBMUntukRakyat Bengkulu'														
11	#####	ÅA;A;A;A;A;A	1.59E+18	ÅA;A;A;A;A	b'gumphehl Yg jadi korban rakyat...alasan dukung demo kenaikan BBM, data2 dijual,lak supir cekar'														
12	#####		1.59E+18	gebbly	b'kenaikan BBM dikatakan Mahal jika kita sehari hari naik mobil mewah beli pertamax'														
13	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
14	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
15	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
16	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
17	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
18	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
19	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														
20	#####	Lampung,	1.59E+18	Nanang Su	b'Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM.\nLampung Dukung Kebijakan Pemerintah\n\n#BLTBBMBantuMasyarakat\n#B														

Gambar 2. Hasil Crawling Data

3.3 Pelabelan Data

Selanjutnya, pelabelan data pada penelitian ini diproses secara manual dan dilabelkan oleh validator Dosen Bahasa Indonesia yaitu ibu Elfina, S.Pd.,M.Pd. dengan mengelompokkan sentimen sesuai dengan kelas nya. Hasil pelabelan dapat dilihat pada gambar berikut.

A	B	C	D	E	F
10/11/2022 0:12	Indonesia	1.59E+22	IF	b'sektor riil pun terhantam. terlebih setelah kenaikan harga bbm.\n\ntiapi ketemu kawan2 ukm, keluhannya s	negatif
10/11/2022 0:09	Jakarta Capital Regi	1.59E+22	Rakyan#DukaKanjuruhan	b'@niluhdjelantik yang saya suka dr ahok, orang lain yang selalu salah. kenaikan bbm pun yang salah orang la	negatif
10/11/2022 0:03		1.59E+22	Zulkifli Usman	b'@niluhdjelantik sejarah ahok\njadi gubernur yg suka maki2 rakyat dan bikin statemen konyol yg mendiskrei	negatif
9/11/2022 23:52	Duren Sawit, Indone	1.59E+23	Ã*Ã;Ã%ÃcÃjÃcÃcÃc	b'@f0werbe4n_hanya di jaman jkw kenaikan bbm sampai 13x. jaman cebong hukum terbang pilih'	negatif
9/11/2022 23:22		1.59E+23	Agathis	b'intimidasi kenaikan harga bbm dan gas, aleg pks: bukti pemerintah lindungi oligarki fraksi pks\n\nharusny	negatif
9/11/2022 18:26		1.59E+23	Visioning Prapatun	b'ehehehe...selalu ada alasan untuk menaikkan cukai rokok dan target pajak yes @kemenkeuri ??\nbaru mau	negatif
9/11/2022 18:22	Jakarta	1.59E+23	Andin Setianingrum	b'news - kenaikan harga bbm tak pengaruhi ekonomi seluruh kuartal 3 2022. aceh hadir mendukung &t	positif
9/11/2022 16:17	Cikarang	1.59E+23	rumahburuh	b'bertepatan hari pahlawan, aliansi bbm nyatakan siap berjuang tuntut kenaikan upah 2023 https://t.co/jp0h	negatif
9/11/2022 15:20		1.59E+23	Boge als Mursid	b'@ramilirizal @kedaiPENACOM mengingatkan kpd kaum tua dan kaum muda, marilah bangkit melawan kedlo	negatif
9/11/2022 15:11	Maluku, Indonesia	1.59E+23	Sukrimin	b'pemerintah siapkan lagi bantuan untuk umkm terdampak kenaikan harga bbm. \xf0\x9f\x91\x8d\xf0\x9f\x	positif
9/11/2022 14:20		1.59E+23	fiona23	b'@pksjabar kami menolak kenaikan bbm hemm pro rakyat, kami menolak #trotoardepok bodo amatt... \xf0'	negatif
9/11/2022 14:18		1.59E+23	chavashawqi	b'bbm satu harga bukti pelayanan pertamina untuk masyarakat dapatkan kualitas terbaik dengan harga resm	positif
9/11/2022 14:09		1.59E+23	New_pengguna	b'@muhammad_saewad tidakkah tercium bau korupsi di kenaikan harga bbm ini....?'	negatif
9/11/2022 12:55	FT - UB Malang	1.59E+23	HMP FT-UB	b'ketidaktepatan sasaran dalam pemberian subsidi bbm menjadi salah satu faktor yang mendorong terbentuk	negatif
9/11/2022 12:29	Lampung	1.59E+23	Nothing !!	b'#makinhits itu kalo elo ikutun demo kenaikan harga bbm tapi bawanya spanduk slank, ama ngebakar foto mantan	positif
9/11/2022 11:59	Jln. Sudirman, Pema	1.59E+23	Polres Pematangsiantar	b'bhabin kamtibmas kel. bahkapul polsek siantar martoba melaksanakan pengawasan dan pengamanan kegi	positif
9/11/2022 11:56		1.59E+23	Humas Polsek Sekampung	b'polripresisi\n\npolreslampungtimur\n\nngiat program quick wins presisi polsek sekampung dalam rangka dz	positif
9/11/2022 11:27		1.59E+23	PintasIndonesiaÃ*Ã,ÃcÃjÃ*	b'bps ungkap fakta, ternyata kenaikan harga bbm hanya berdampak selama satu bulan!\n#kebayamerah\nka	positif
9/11/2022 11:07	Indonesia	1.59E+22	Maialah GATRA	b'dari 90 jumlah kota di indonesia. ada 88 kota mengalami inflasi di seotember lalu. laiu akibat kenaikan hare:	oositif

Gambar 3. Hasil Pelabelan Manual Data

3.4 Text Preprocessing

Kemudian pada tahap preprocessing merupakan suatu proses untuk menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data. Dalam tahap preprocessing terbagi menjadi beberapa proses yaitu:

- Cleaning merupakan tahap proses pembersihan kata maupun karakter yang tidak penting atau tidak berkaitan dalam proses klasifikasi. Karakter yang dihilangkan berupa tanda baca, tanda hastag (#), tanda tanya (?), tanda seru (!), emoji, link website, username (@), dan lain sebagainya. Hasil dari cleaning bisa dilihat pada tabel berikut.

Tabel 1. Hasil Cleaning

Sebelum Cleaning	Hasil Cleaning
Hey masyarakat sumatera Utara tak perlu panik dengan kenaikan harga BBM. Yang pastinya klen semua akan dapat bansos	Hey masyarakat sumatera Utara tak perlu panik dengan kenaikan harga BBM Yang pastinya klen semua akan dapat bansos

- Case Folding adalah cara yang dilakukan untuk mengubah semua huruf besar menjadi huruf kecil serta penghapusan tanda baca. Hasil dari case folding dapat dilihat pada tabel dibawah.

Tabel 2. Hasil Case Folding

Sebelum Case Folding	Hasil Case Folding
Hey masyarakat sumatera Utara tak perlu panik dengan kenaikan harga BBM Yang pastinya klen semua akan dapat bansos	hey masyarakat sumatera utara tak perlu panik dengan kenaikan harga bbm yang pastinya klen semua akan dapat bansos

- Tokenizing merupakan tahapan proses pemecahan suatu kalimat menjadi perkata. Hasil dari tokenizing dapat dilihat pada tabel berikut.

Tabel 3. Hasil Tokenizing

Sebelum Tokenizing	Hasil Tokenizing
hey masyarakat sumatera utara tak perlu panik dengan kenaikan harga bbm yang pastinya klen semua akan dapat bansos	['hey', 'masyarakat', 'sumatera', 'utara', 'tak', 'perlu', 'panik', 'dengan', 'kenaikan', 'harga', 'bbm', 'yang', 'pastinya', 'klen', 'semua', 'akan', 'dapat', 'bansos']

- Normalisasi yaitu proses merubah kata sesuai KBBi yaitu kata yang dianggap tidak baku menjadi baku. Bisa dilihat pada tabel berikut.

Tabel 4. Hasil Normalisasi

Sebelum Normalisasi	Hasil Normalisasi
hey	hey
masyarakat	masyarakat
sumatera	sumatera
utara	utara

Sebelum Normalisasi	Hasil Normalisasi
tak	tidak
perlu	perlu
panik	panik
dengan	dengan
kenaikan	kenaikan
harga	harga
bbm	bbm
yang	yang
pastinya	pastinya
klen	kalian
semua	semua
akan	akan
dapat	dapat
bansos	bansos

- e. Negation Handling merupakan tahapan proses mengubah kata yang mengandung negasi menjadi kata yang sebenarnya. Berikut hasil dari negation handling.

Tabel 5. Hasil Tokenizing

Sebelum Negation Handling	Hasil Negation Handling
['hey', 'masyarakat', 'sumatera', 'utara', 'tidak', 'perlu', 'panik', 'dengan', 'kenaikan', 'harga', 'bbm', 'yang', 'pastinya', 'kalian', 'semua', 'akan', 'dapat', 'bansos']	['hey', 'masyarakat', 'sumatera', 'utara', 'abaikan', 'panik', 'dengan', 'kenaikan', 'harga', 'bbm', 'yang', 'pastinya', 'kalian', 'semua', 'akan', 'dapat', 'bansos']

- f. Stopword Removal adalah proses penghapusan kata yang dianggap tidak berguna. Bisa dilihat pada tabel berikut.

Tabel 6. Hasil Stopword Removal

Sebelum Stopword Removal	Hasil Stopword Removal
['hey', 'masyarakat', 'sumatera', 'utara', 'abaikan', 'panik', 'dengan', 'kenaikan', 'harga', 'bbm', 'yang', 'pastinya', 'kalian', 'semua', 'akan', 'dapat', 'bansos']	['hey', 'masyarakat', 'sumatera', 'utara', 'abaikan', 'panik', 'kenaikan', 'harga', 'bbm', 'pastinya', 'bansos']

- g. Stemming, yaitu suatu cara pengambilan kata yang berimbuhan menjadi kata dasar. Bisa dilihat pada tabel berikut.

Tabel 7. Hasil Stemming

Sebelum Stemming	Hasil Stemming
['hey', 'masyarakat', 'sumatera', 'utara', 'abaikan', 'panik', 'kenaikan', 'harga', 'bbm', 'pastinya', 'bansos']	['hey', 'masyarakat', 'sumatera', 'utara', 'abai', 'panik', 'naik', 'harga', 'bbm', 'pasti', 'bansos']

3.5 Feature Selection

Selanjutnya masuk ketahap pemilihan fitur menggunakan thresholding merupakan tahapan yang dilakukan untuk mengurangi dimensi data pada suatu penelitian.

	000	10	2022	abang	aceh	ahok	akibat	alas	allah	angkut	...	untung	upah	usaha	utang	utara	wajar	warga	wkwk	yogyakarta	Kelas					
0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif				
1	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif			
2	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.279229	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif			
3	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif		
4	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif		
...		
1495	0.0	0.0	0.0	0.0	0.0	0.167818	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif		
1496	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	
1497	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	
1498	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
1499	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif

Gambar 4. Hasil Thresholding

3.6 Feature Weighting

Setelah melakukan tahap text preprocessing kemudian dilakukan pembobotan kata menggunakan TF-IDF. Berikut hasil dari perhitungan TF-IDF yang dapat dilihat pada gambar berikut.

	000	09	10	100	1001	100k	100km	100persen	100rb	1075	...	ywdh	zainal	zalim	zaman	zefanya	zerr	zholimi	zico	zona	Kelas			
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif		
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif

Gambar 5. Hasil TF-IDF

3.7 Klasifikasi dengan Metode Support Vector Machine

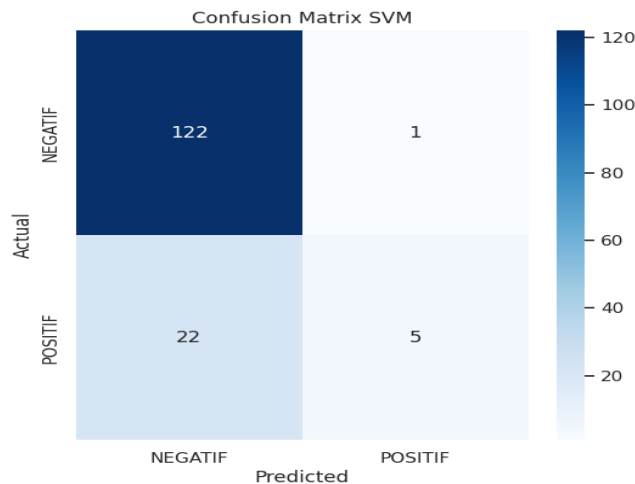
Setelah melakukan pembobotan dengan TF-IDF, tahap berikutnya masuk kedalam klasifikasi. Dengan menggunakan klasifikasi metode Support Vector Machine, pada tahap ini data akan di proses melalui fase training dan testing, kemudian dilakukan modeling dengan metode SVM untuk memperoleh nilai accuracy, precision, recall, dan f1 score. Berikut hasil klasifikasi metode SVM dengan menggunakan tools Google Colllab.

	000	2022	adang	adeh	aja	akibat	alhamdulillah	alun	aman	angkut	...	urah	usaha	utara	wajar	waluku	warna	ya	yogyakarta	Kelas	Target	Hasil	Klasifikasi	
0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
1	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	positif	positif	
2	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	positif	positif	
3	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
4	0.293272	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
...
325	0.000000	0.0	0.0	0.0	0.081305	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.123558	0.0	negatif	negatif	positif	positif	
326	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
327	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
328	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	negatif	negatif	negatif	negatif	
329	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.300318	0.0	0.0	0.0	0.0	0.000000	0.0	positif	positif	positif	positif	

Gambar 6. Hasil Klasifikasi

3.8 Pengujian

Hasil pengujian untuk data training dan testing menggunakan confusion matrix. Pada penelitian ini melakukan 3 pengujian yaitu untuk data opini – data opini, data balance, dan data opini. Dengan pengujian perbandingan 90:10 dan 80:20. Berikut hasil pada gambar dibawah ini.



Gambar 7. Hasil Confussion Matrix

Pada perbandingan 90:10 didapatkan akurasi sebesar 85%, untuk nilai precission sebesar 85%, recall 99% dan f1-score sebesar 91% untuk sentimen negatif, sedangkan nilai precission 83%, recall 19%, f1-score 30% untuk sentimen positif. Kemudian pada percobaan perbandingan 80:20 didapatkan akurasi sebesar 83%, untuk nilai

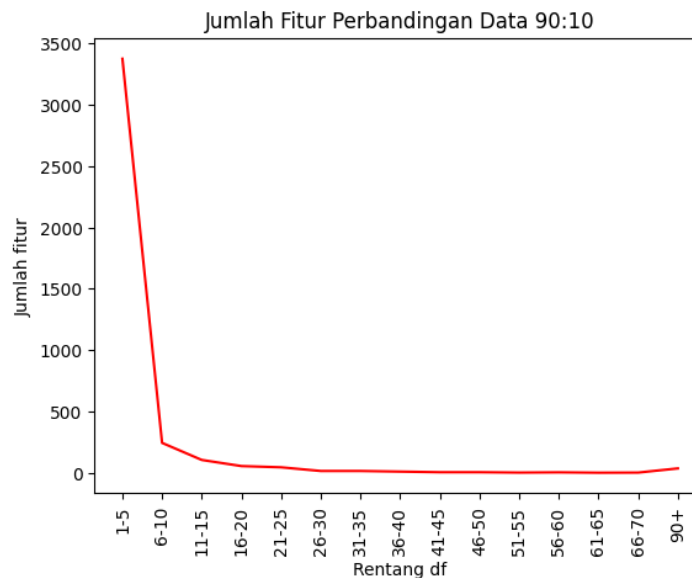
precision sebesar 83%, recall 99% dan f1-score sebesar 91% untuk sentimen negatif, sedangkan nilai precision 82%, recall 16%, f1-score 26% untuk sentimen positif.

a. Ekstraksi Fitur DF, yang ditampilkan pada Tabel 8 dibawah ini.

Tabel 8. Tabel Distribusi Nilai DF

No	Nilai DF	Jumlah Kata
1	1-5	3374
2	6-10	244
3	11-15	105
4	16-20	55
5	21-25	45
6	26-30	15
7	31-35	15
8	36-40	10
9	41-45	5
10	46-50	5
11	51-55	2
12	56-60	4
13	61-65	1
14	66-70	2
15	70-inf	36

Pada tabel di atas nilai yang dipilih berdasarkan jumlah kata yang terdapat pada grafik tersebut.



Gambar 6. Fitur yang Terpilih untuk Nilai DF

3.9 K-fold Cross Validation

K-fold cross validation didapat dari pasangan parameter C dan gamma pada metode Support Vector Machine (SVM). Ditunjukkan pada Tabel 9 sebagai berikut.

Tabel 9. Tabel 10 K-fold Cross Validation

c/γ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,1	0,7755	0,7755	0,7755	0,7755	0,7807	0,7888	0,7925	0,7948	0,7970	0,8000
0,2	0,7755	0,7755	0,7829	0,7940	0,7977	0,8007	0,8022	0,8029	0,8029	0,8051
0,3	0,7755	0,7807	0,7948	0,8000	0,8022	0,8029	0,8044	0,8059	0,8074	0,8088
0,4	0,7755	0,7896	0,7970	0,8014	0,8029	0,8037	0,8051	0,8051	0,8081	0,8088
0,5	0,7755	0,7933	0,7992	0,8029	0,8037	0,8051	0,8074	0,8074	0,8081	0,8103
0,6	0,7755	0,7948	0,8000	0,8037	0,8044	0,8059	0,8074	0,8081	0,8096	0,8103
0,7	0,7762	0,7955	0,8007	0,8044	0,8059	0,8074	0,8074	0,8081	0,8096	0,8088
0,8	0,7762	0,7955	0,8014	0,8037	0,8059	0,8081	0,8081	0,8081	0,8088	0,8096
0,9	0,7762	0,7955	0,8022	0,8044	0,8059	0,8081	0,8081	0,8074	0,8096	0,8096
1,0	0,7762	0,7962	0,8022	0,8044	0,8059	0,8074	0,8081	0,8074	0,8088	0,8103

Berdasarkan pada tabel diatas, yang menjadi parameter terbaik berada pada pasangan parameter $C = 1,0$ dan $\gamma = 1,0$ didapatkan hasil akurasi sebesar 81% yang merupakan model dengan hasil tertinggi.

4. KESIMPULAN

Berdasarkan hasil yang dilakukan pada penelitian ini dengan metode Support Vector Machine pada proses pengklasifikasian sentimen yang diambil dari Twitter, yaitu mengenai kenaikan harga BBM berhasil dilakukan, yaitu dengan menggunakan dua pengujian berbeda dan menggunakan data opini sebanyak 1500 data. Hasil pengujian dari data opini pada perbandingan 90:10 didapat dengan akurasi yaitu sebesar 85%, untuk nilai precision sebesar 85%, recall 99% dan f1-score sebesar 91% untuk sentimen negatif, sedangkan nilai precision 83%, recall 19%, f1-score 30% untuk sentimen positif. Pasangan parameter terbaik berada pada pasangan $C = 1,0$ dan $\gamma = 1,0$ yaitu dengan akurasi sebesar 81%. Hasil diatas membuktikan bahwa metode Support Vector Machine menghasilkan nilai atau tingkat akurasi yang cukup baik.

REFERENCES

- [1] Appannie.com, "State of Mobile 2022," appannie.com, 2022. .
- [2] S. Merlinda, R. Aniq, H. Alam, F. Ekonomi, and D. Bisnis, "Peranan Pemerintah melalui Intervensi Pasar dalam Perspektif Islam (Studi Kasus: Inflasi Harga BBM di Indonesia) Qorry Anggita Rishaq," vol. 20, pp. 114–134, 2022.
- [3] B. Pamungkas, M. E. Purbaya, and D. J. A.K, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," J. Informatics, Inf. Syst. Softw. Eng. Appl., vol. 3, no. 2, pp. 10–20, 2021.
- [4] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," Edutic - Sci. J. Informatics Educ., vol. 7, no. 1, pp. 1–11, 2020.
- [5] 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.
- [6] Ferdi and Vina Ayumi, "Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes dan Support Vector Machine," JSAI (Journal Sci. Appl. Informatics), vol. 6, no. 1, pp. 1–10, 2023, doi: 10.36085/jsai.v6i1.4628.
- [7] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Classifier di Twitter," J. TeknoKompak, vol. 17, no. 1, pp. 25–38, 2023, doi: 10.30865/json.v4i2.5381.
- [8] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.
- [9] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly dengan Metode Naive Bayes dan Support Vector Machine (SVM)," Ilk. J. Ilm., vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [10] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," J. Inform. J. Pengemb. IT, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [11] A. D. Hartanto and N. K. Fitriyani, "Analisis Sentimen Terhadap Tokoh Publik Menggunakan Support Vector Machine," MEANS (Media Inf. Anal. dan Sist., vol. 5, no. 1, pp. 8–12, 2020.
- [12] A. S. Rahayu and A. Fauzi, "Komparasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen Spotify," J. Sist. Komput. dan Inform. Hal 349–, vol. 354, no. 2, pp. 349–354, 2022, doi: 10.30865/json.v4i2.5398.
- [13] L. Mutawalli, M. T. A. Zaen, and W. Bagye, "Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)," J. Inform. dan Rekayasa Elektron., vol. 2, no. 2, p. 43, 2019, doi: 10.36595/jire.v2i2.117.
- [14] D. W. Syahputra, B. Rahayudi, and L. Muflikhah, "Analisis Sentimen Twitter terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat menggunakan Metode Support Vector Machine," Fak. Ilmu Komputer, Univ. Brawijaya, vol. 6, no. 3, pp. 1067–1072, 2022.
- [15] M. Metode, S. Vector, M. Svm, and D. A. N. Asosiasi, "Analisis Sentimen Ulasan Aplikasi Tiktok di Google Play menggunakan Metode Support Vector Machine (SVM) dan Asosiasi," vol. 10, pp. 346–358, 2021.
- [16] W. Paulina, F. A. Bachtiar, and A. N. Rusydi, "Analisis Sentimen Berbasis Aspek Ulasan Pelanggan terhadap Kertanegara Premium Guest House Menggunakan Support Vector Machine," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 4, no. 4, pp. 1141–1149, 2020.
- [17] H. C. Husada and A. S. Paramita, "Analisis Sentimen pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," Teknika, vol. 10, no. 1, pp. 18–26, 2021.
- [18] Y. Yusra and M. Fikry, "Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine," J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf., vol. 4, no. 2, p. 50, 2018, doi: 10.24014/coreit.v4i2.5205.
- [19] Styawati., N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat terhadap Program Kartu Prakerja pada Twitter dengan Metode Support Vector Machine," J. Inform. J. Pengemb. IT, vol. 6, no. 3, pp. 150–155, 2021.
- [20] R. Darmawan, I. Indra, and A. Surahmat, "Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) pada Analisis Sentimen Terhadap Official Account Ruang Guru di Twitter," J. Kaji. Ilm., vol. 22, no. 2, pp. 143–152, 2022, doi: 10.31599/jki.v22i2.1130.