

JURNAL RATIH PUSPITASARI

by Alfian Indra Kusuma

Submission date: 15-Feb-2023 03:13PM (UTC+0900)

Submission ID: 2014647519

File name: 801-Article_Text-4283-1-9-20230119_UMSIDA.pdf (982.95K)

Word count: 5601

Character count: 34411

SENTIMENT ANALYSIS OF POST-COVID-19 INFLATION BASED ON TWITTER USING THE K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE CLASSIFICATION METHODS

Ratih Puspitasari¹, Yulian Findawati^{*2}, Mochamad Alfian Rosid³

^{1,2,3}Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo, Indonesia
Email: ¹Ratihpsari@umsida.ac.id, ²yulianfindawati@umsida.ac.id, ³alfanrosid@umsida.ac.id

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstract

The COVID-19 pandemic caused a crisis in global economic growth. The impact of injuries due to the COVID-19 pandemic has also caused price increases and an increase in the inflation rate. Inflation is a price increase caused by a certain factor so that it has an impact on the prices of nearby goods which increase the circulation of money in society to increase. Many people expressed their various opinions or criticisms of the post-COVID-19 price increase policy on social media, one of which was via Twitter. Sentiment analysis was carried out to see how public sentiment is towards the price increase policy after the COVID-19 pandemic, and these sentiments are combined into multiclass, namely positive, negative and neutral sentiments. So that this sentiment can later be used as material for evaluation regarding the post-COVID-19 price increase policy. This study aims to see and compare the accuracy of the two classification methods, namely K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) in the sentiment classification process. The data used was 5989 tweets with the keywords "Stuffs Go Up Post-Pandemic", "Fuel Goes Up", "Inflation 2022", "Covid19 Inflation", "Inflation Post-Pandemic" with a data collection period from August to October 2022. The data obtained then enter the text preprocessing stage before later entering the classification stage. The results obtained after carrying out the classification using the K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) methods show that the Support Vector Machine (SVM) method has a higher accuracy of 79%, while the K-Nearest Neighbor (K-NN) has an accuracy of 54%.

Keywords: COVID-19, Inflation, K-Nearest Neighbor, Sentiment Analysis, Support Vector Machine

ANALISIS SENTIMEN TERHADAP INFLASI PASCA COVID-19 BERDASARKAN TWITTER DENGAN METODE KLASIFIKASI K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE

Abstrak

Pandemi COVID-19 menyebabkan krisis pada pertumbuhan ekonomi secara global. Dampak luka akibat pandemi COVID-19 juga menyebabkan harga naik dan kenaikan angka inflasi. Inflasi merupakan kenaikan harga yang disebabkan oleh suatu faktor tertentu sehingga berdampak pada harga barang terdekat yang meningkatkan peredaran uang di masyarakat menjadi meningkat. Banyak masyarakat menyampaikan berbagai opini atau kritik mereka terhadap kebijakan kenaikan harga pasca pandemi COVID-19 di media sosial, salah satunya melalui Twitter. Sentimen analisis dilakukan untuk melihat bagaimana sentiment masyarakat terhadap kebijakan kenaikan harga pasca pandemi COVID-19, dan sentimen tersebut dikategorikan menjadi multiclass yaitu sentimen positif, negatif dan netral. Sehingga sentimen tersebut nantinya bisa menjadi bahan evaluasi terkait kebijakan kenaikan harga pasca pandemi COVID-19. Penelitian ini bertujuan untuk melihat dan membandingkan akurasi dari dua metode klasifikasi yaitu K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) pada proses klasifikasi analisis sentimen. Data yang digunakan sebanyak 5989 tweet dengan kata kunci "Bahan Pokok Naik Pasca Pandemi", "BBM Naik", "Inflasi 2022", "Inflasi Covid19", "Inflasi Pasca Pandemi" dengan periode pengumpulan data dari bulan Agustus hingga Oktober 2022. Data yang didapatkan kemudian masuk ke tahap text preprocessing sebelum nantinya masuk ke tahap klasifikasi. Hasil yang didapatkan setelah melakukan klasifikasi menggunakan metode K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) menunjukkan bahwa metode Support Vector Machine (SVM) memiliki akurasi yang lebih tinggi yaitu 79%, sedangkan metode K-Nearest Neighbor (K-NN) memiliki akurasi sebesar 54%.

Kata kunci: Analisis Sentimen, COVID-19, Inflasi, K-Nearest Neighbor, Support Vector Machine

1. PENDAHULUAN

Organisasi Kesehatan Dunia (*World Health Organization*) menggambarkan coronavirus (Cov) sebagai virus yang menginfeksi saluran pernapasan. Wabah virus corona yang telah menyebar ke wilayah berbeda di dunia berdampak pada perekonomian Indonesia baik dari bidang perdagangan, investasi maupun pariwisata. Jutaan orang akan terjerumus ke dalam kemiskinan karena jumlah pengangguran akan meningkat karena beberapa aktivitas manufaktur akan terhenti karena kurangnya permintaan untuk merangsang aktivitas manufaktur [1].

Secara khusus, pertumbuhan ekonomi Indonesia saat ini berada di angka yang bisa dikatakan berkembang ke arah yang positif dibandingkan 2 tahun atau saat wabah COVID-19 melanda setiap pelosok tanah air. Indikator kunci untuk mengetahui keadaan perekonomian Indonesia adalah mengetahui Produk Domestik Bruto (PDB) yang menunjukkan penurunan tajam sebesar 2,07% year-on-year pada tahun 2020. Pada tahun 2021, diproyeksikan meningkat sebesar, atau 3,69%. Salah satunya adalah inflasi kenaikan sejak awal 2022 dan berada di atas 2% pada awal tahun yakni 2,18%, tetapi mencapai 3,55% pada akhir Mei 2022 (data dari Bank Indonesia)[2]. Proyeksi pertumbuhan perekonomian secara Global (%*y-on-y*) yang berkaitan dengan latar belakang dari kondisi perekonomian global dapat dilihat pada Gambar 1[3].



Gambar 1. Kondisi Perekonomian Global.

Inflasi adalah kenaikan harga, yang bisa dikatakan menjadi penurunan daya beli dari waktu ke waktu. Kenaikan harga yang seringkali dinyatakan menjadi persentase, mengindikasikan bahwa satu unit yang dibeli dengan jumlah uang yang sama hanya bisa membeli lebih sedikit menurut dalam periode sebelumnya[4]. Pada awal bulan September 2022 pemerintah memutuskan kebijakan kenaikan harga Bahan Bakar Minyak (BBM) pada Indonesia. Kenaikan BBM yang terjadi pada Indonesia akan berdampak dalam inflasi sebagai akibatnya akan berpengaruh terhadap mendasar makro ekonomi Indonesia[5].

Saat ini, pengguna internet paling sering mengunjungi situs media sosial (191 juta orang pada Januari 2022, dari seluruh responden), dan salah satu dari media sosial terbaru adalah Twitter. Media sosial Twitter adalah salah satu media terhubung yang dibutuhkan oleh orang di seluruh dunia. Media sosial twitter adalah salah satu media terhubung yang dibutuhkan oleh masyarakat di seluruh dunia.

Pengguna Twitter memberikan berita dan komentar terbaru tentang apa yang sedang tren di dunia saat ini. Pertumbuhan Twitter terus meningkat setiap waktu, sebagai akibatnya hal tersebut dimanfaatkan oleh pengguna Twitter untuk mengungkapkan kabar berupa komentar kritik juga saran pada kebijakan kenaikan harga pasca pandemi Covid-19[6].

Analisis sentimen salah satu dalam bidang Natural Language Processing (NLP) dan merupakan proses yang digunakan membantu mengidentifikasi isi catatan opini atau opini (perasaan) dalam bentuk teks tentang suatu masalah atau peristiwa itu termasuk kelas positif, negatif atau netral. Pada umumnya, analisis sentimen diterapkan pada perspektif yang berbeda, seperti prediksi harga saham, isu politik, kepuasan ke produk atau layanan, analisis reputasi dan sebagainya. Salah satu analisis sentimen yang bisa dilakukan adalah analisis sentiment terhadap isu Twitter[7]. Salah satu cirinya jika tweet digunakan akan sangat tidak berguna jika tidak digunakan dengan benar, sehingga dibutuhkan teknologi yang dapat memproses tweet tersebut untuk menghasilkan hasil nantinya menjadi informasi berharga[8].

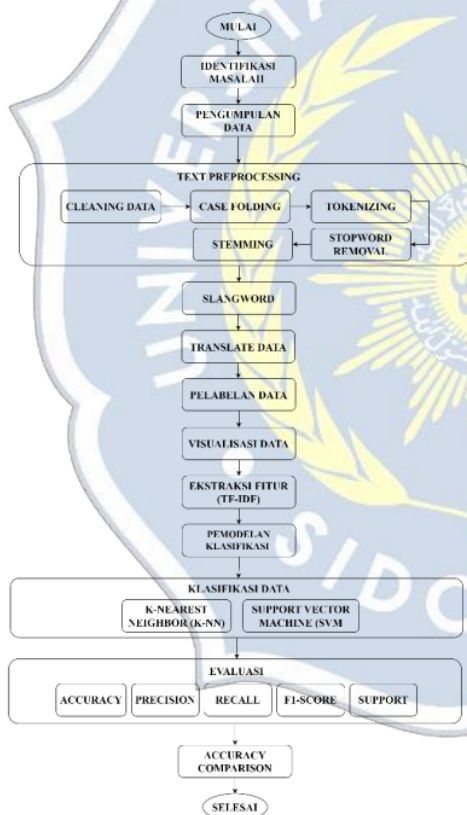
Pada penelitian ini menggunakan metode klasifikasi, yaitu sebuah algoritma *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM). Algoritma masing-masing klasifikasi ini memiliki kelebihan dan kekurangan klasifikasi data dalam bentuk teks. Terdapat beberapa penelitian yang merujuk pada penelitian ini. Pertama adalah penelitian mengenai perbandingan tiga algoritma, yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbor* (K-NN). Dari pengujian ini, algoritma SVM mendapatkan skor akurasi 85.71%, Algoritma *Naïve Bayes* mendapatkan skor akurasi 76.70%, Algoritma K-NN mendapatkan skor akurasi 52.74% [9]. Kedua adalah penelitian mengenai tiga metode termasuk *Naïve Bayes*, K-NN, dan *Decision Tree*. Data yang digunakan dalam penelitian ini adalah data persepsi siswa menggunakan Twitter tentang pelaksanaan kuliah daring yang diperoleh dengan cara crawling data. Dari pengujian diketahui bahwa metode *Decision Tree* memiliki prioritas yang lebih tinggi antara lain dengan akurasi 61.92%, presisi 73.63%, dan recall 11.42%[10]. Ketiga adalah penelitian tanggapan dari masyarakat terhadap tindakan vaksinasi yang beragam di jejaring sosial Twitter, ada yang mendukung dan tidak setuju. Dataset yang digunakan hingga 845 Tweet dengan dua kata kunci, yaitu "vaksinmerahputih" dan "vaksininovac". Kemudian informasi tersebut dibagikan menjadi 253 untuk data *training* dan 592 data *testing*. Hasil klasifikasi menggunakan metode *Naïve Bayes* mendapatkan nilai akurasi 85,59%, sedangkan SVM sebesar 84,41%[11].

Penelitian sebelumnya dapat membuktikan bahwa analisis sentimen dapat dilakukan dengan *Machine Learning*, salah satunya metode adalah

klasifikasi. Masalah apa yang harus dilihat jika pemerintah menetapkan kebijakan kenaikan harga pasca pandemi covid-19. Tujuan penelitian ini untuk mengetahui sentiment masyarakat terhadap inflasi pasca pandemi covid-19 apakah menerima ulasan positif, negatif atau netral dengan metode *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) dan melihat perbandingan keakuratan kedua metode tersebut.

2. METODE PENELITIAN

Penelitian dilakukan berdasarkan data yang diambil melalui media sosial Twitter menggunakan emulator Google Colab. Diagram alir penelitian yang diusulkan disajikan dalam bentuk bagan, sebagaimana dapat dilihat pada Gambar 2.



Gambar 2. Diagram Alir Penelitian

Berdasarkan Gambar 2 dapat dilihat merupakan proses dalam tahapan penelitian. Tahapan tersebut digunakan untuk memudahkan pada proses penelitian yang dilakukan.

2.1. Identifikasi Masalah

Penelitian yang dilakukan berdasarkan permasalahan yang telah dijelaskan pada sebelumnya yaitu :

1. Penerapan metode *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) pada analisis isi teks Twitter terkait Inflasi pasca pandemic covid-19 jika teks termasuk teks berlabel positif, negatif, dan netral.
2. Untuk mengetahui seberapa akurat dengan melakukan perbandingan antara metode *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) saat menganalisis sentimen terkait inflasi pasca pandemi covid-19.

2.2. Pengumpulan Data

Setelah proses mengidentifikasi masalah, proses selanjutnya yaitu pengumpulan data dengan menggunakan emulator Google Colab. Proses dari pengumpulan data dapat dilihat pada Gambar 3.

```
c = twint.Config()
searchstr = "Inflasi 2022 until:2022-08-30 since:2022-08-01"
c.Search = searchstr
c.Lang = "id"
c.Pandas = True
c.Store_csv = True
c.Output = "Inflasi 2022 - tweet.csv"
twint.run.Search(c)
```

Gambar 3. Source Code Pengambilan Data Twitter

Pada Gambar 3 merupakan *library twint* yang digunakan pada proses pengumpulan data yaitu *import twint*. Twint merupakan *tools scrapping* data tweets yang terdapat pada python tanpa menggunakan API Twitter. Pada tahap ini, dilakukan pencarian *keyword* berdasarkan kata kunci “Bahan Pokok Naik Pasca Pandemi”, “BBM Naik”, “Inflasi 2022”, “Inflasi Covid19”, “Inflasi Pasca Pandemi” dari bulan Agustus hingga Oktober 2022. Setelah pengambilan data tweet dari Twitter kemudian data tersebut disimpan dalam bentuk format .csv.

2.3. Text Mining

Text mining merupakan salah satu metode yang dapat digunakan untuk mengatasi masalah *classification*, *clustering*, *information extraction*, dan *information retrieval*. *Text mining* adalah fase pencarian informasi yang dapat berinteraksi dengan beberapa dokumen dari waktu ke waktu dengan alat analisis. Prosedur utama untuk *text mining* adalah memperoleh kata-kata yang dapat menggantikan isi dokumen dan kemudian dilakukan proses verifikasi untuk melihat bagaimana setiap dokumen memiliki keterkaitan[12].

2.4. Analisis Sentimen

Analisis sentimen adalah proses penentuan sentimen dan klasifikasi polaritas teks dari suatu dokumen atau kalimat untuk mendefinisikan kategori sebagai sentimen positif, negatif, atau netral. Jaringan media sosial seperti Twitter sering digunakan dalam analisis sentimen untuk menentukan persepsi publik. Analisis sentimen juga dapat dibandingkan dengan opini mining karena berfokus pada pendapat yang bersifat opini diekspresikan secara positif atau negatif. Analisis sentimen melibatkan penambangan data untuk menganalisis, mengekstrak data tekstual dalam suatu entitas, seperti jasa, produk, individu, atau topik tertentu[13].

2.5. K-Nearest Neighbor Classifier

Algoritma *K-Nearest Neighbor* (KNN) digunakan untuk mengklasifikasikan objek berdasarkan pembelajaran yang paling dekat dengan objek. Pengukuran jarak dalam penelitian ini menggunakan metode *Euclidean Distance*. Rumus perhitungan jarak Euclidean dapat dilihat pada persamaan 1. Dimana $dis(x,y)$ adalah jarak antara data x dan data y , x_i adalah data test ke- i dan y_i adalah data train ke- i [14].

$$dis(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Algoritma *K-Nearest Neighbors* adalah penentu klasifikasi berdasarkan contoh fundamental yang tidak membangun representasi kategori yang eksplisit dan deklaratif, tetapi bergantung pada label kelas yang dilampirkan pada dokumen train sesuai dengan dokumen test. Algoritma *K-Nearest Neighbors* adalah metode untuk mengklasifikasikan objek berdasarkan data training yang terdekat. *K-Nearest Neighbors* memiliki prinsip yang sederhana, bekerja berdasarkan jarak terdekat dari sampel test ke sampel train. Algoritma *K-Nearest Neighbors* memiliki tahapan :

1. Mendefinisikan parameter K (jumlah tetangga terdekat).
2. Menentukan bobot setiap term dengan menggunakan Term Weighting TF-IDF.
3. Menghitung kemiripan antar dokumen menggunakan cosinus similarity dengan menggunakan persamaan 2 sebagai berikut.

$$\cos(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (2)$$

4. Menghitung kemiripan antar dokumen menggunakan cosinus similarity
5. Ambil sebanyak K dengan kemiripan tertinggi dengan dokumen yang sudah diklasifikasikan, mendefinisikan sebuah kelas[15].

2.6. Support Vector Machine Classifier

Algoritma *Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi yang kuat untuk masalah karakter non-linier. SVM juga berfungsi setelah menemukan batasan (*hyperplane*) dari suatu fungsi atau proses memaksimalkan jarak antar kelas untuk setiap sentimen. Menangani kasus non-linear, SVM dimodifikasi untuk menyertakan fungsi kernel untuk menemukan hasil dengan cepat. Berdasarkan kinerja kernel SVM yang digunakan dalam penelitian ini adalah :

1. Radial Basis Function (RBF)

RBF adalah metode mengklasifikasikan kumpulan data yang tidak terpisah secara linear. Demikian RBF ini memiliki keuntungan tingkat akurasi training dan prediksi yang sangat baik. Dapat diimplementasikan menggunakan persamaan 3 sebagai berikut[16]:

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}\right) \quad (3)$$

2. Linear Kernel

Linear kernel adalah fungsi kernel yang baik digunakan ketika data sudah terpisah secara linier. Dapat diimplementasikan menggunakan persamaan 4 sebagai berikut:

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^t \vec{x}_j \quad (4)$$

2.7. Text Preprocessing

Text preprocessing dalam penelitian ini, menerapkan serangkaian tahapan yang berurutan yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*. Penjelasan singkat dari masing-masing tahapan tersebut disajikan berikut ini :

1. *Cleaning* merupakan tahapan untuk membersihkan karakter yang tidak dibutuhkan seperti URL, @, #, https:, RT (*Retweet*), angka, simbol, dan emoticon.
2. *Case folding* merupakan tahapan untuk mengubah huruf kapital dalam dokumen menjadi standar, yaitu huruf kecil (*lowercase*).
3. *Tokenizing* merupakan tahapan untuk memecah kalimat dalam dokumen menjadi kata. Sebuah proses untuk menghilangkan tanda baca, simbol, dan karakter bacaan yang tidak berharga.
4. *Stopword removal* merupakan tahapan untuk menghilangkan sebuah kata informatif rendah. Stopword dilakukan jika kalimat tersebut mengandung kata umum dan tidak signifikan seperti waktu, penghubung, dan sebagainya.
5. *Stemming* merupakan tahapan untuk menghapus imbuhan awalan dan akhiran, sehingga menjadi bentuk kata dasar. *Stemming* dilakukan menggunakan library Sastrawi pada python

Text processing dilakukan untuk merubah tweet atau data teks tidak terstruktur sehingga data dapat disusun sesuai kebutuhan analisis sentimen tentang inflasi pasca pandemi COVID-19[17].

2.8. Slangword

Konversi kata slang adalah proses mengubah kata tidak baku menjadi kata baku. Tahapan ini dilakukan dengan bantuan kamus kata slang dan setara dalam kata baku. Tahapan ini memeriksa apakah kata-kata tersebut mengandung unsur kata-kata slang atau tidak. Jika kata tidak baku ada dalam kamus kata slang, kata tidak baku akan diubah menjadi kata baku[18]. Istilah, singkatan, atau kesalahan ejaan yang tidak standar diklasifikasikan sebagai kategori slang. Tahap konversi slangword mengubah bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI)[19]. Dalam penelitian ini jumlah kamus slangwords sebanyak 1.500 kata.

2.9. Translate Data

Tahap translate data adalah tahap untuk menerjemahkan kata berbahasa Indonesia ke dalam bahasa Inggris. Tahap ini wajib dilakukan apabila pelabelan data menggunakan *library Vader Sentiment Lexicon*. *Library Vader Sentiment Lexicon* hanya mengandung kata berbahasa Inggris. Tahap translate dilakukan setelah melalui tahap slangword pengubahan kata tidak baku menjadi baku. Proses translate data menggunakan *library google translate* dengan python dapat dilihat pada Gambar 4.

```
import googletrans
from googletrans import Translator
translator = Translator()
translations = {}
for column in data.columns:
    unique elements = data[column].unique()
    for element in unique elements:
        translations[element] =
translator.translate(element).text
translations
```

Gambar 4. Source Code Translate Data

2.10. Pelabelan Sentimen

Tahapan pelabelan sentimen menggunakan *library Vader Sentiment* dengan python. Sebelum menggunakan *Vader Sentiment*, data tweet diterjemahkan terlebih dahulu ke dalam bahasa Inggris karena *library Vader Sentiment* hanya bisa menggunakan bahasa Inggris saat menentukan apakah sifat tweet tersebut masuk ke dalam kelas positif, negatif, atau netral[20].

2.11. Evaluasi

Pada penelitian ini menggunakan evaluasi *fold-cross* validasi sebesar 10. Proses validasi memiliki dua subproses untuk melatih data, yaitu sub-proses *training* dan sub-proses *testing*. Setelah melakukan *training* data, dengan bantuan tools Google Colab akan menghasilkan evaluasi kinerja model klasifikasi pada penelitian yaitu *accuracy*, *precision*, *recall*, dan *f1-score*. Rumus perhitungan evaluasi kinerja model klasifikasi dapat dilihat pada persamaan (5), (6), (7), dan (8).

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)} \times 100\% \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$F - score = \frac{2 \times precision \times recall}{precision + recall} \quad (8)$$

Skor model klasifikasi terdapat TP (True Positive) adalah hasil klasifikasi yang benar, TN (True Negative) adalah hasil dari klasifikasi yang tidak tepat, FP (False Positive), yang merupakan hasil klasifikasi yang benar tapi faktanya kurang tepat, FN (False Negative) yaitu hasil dari klasifikasi yang salah, tetapi faktanya tepat[21].

3. HASIL DAN PEMBAHASAN

Pelaksanaan analisis sentimen dalam penelitian ini dilakukan secara bertahap, meliputi pengumpulan data (dengan *crawling* data di Twitter), *text preprocessing*, pelabelan data, sampai dengan hasil pengujian dan evaluasi.

3.1. Dataset

Penelitian ini menggunakan data yang diambil melalui bantuan *library twint* di python untuk mengekstrak *tweet* tentang inflasi pasca pandemi COVID-19 dari Twitter. Data tweet tersebut akan dilihat keakuratan kedua algoritma tersebut selanjutnya akan dibandingkan menggunakan data tweet setelah analisis sentimen menggunakan metode *K-Nearest Neighbors* dan *Support Vector Machine*. Data tweet yang dihasilkan berjumlah 7.899 dari tanggal 2022-08-01 hingga 2022-10-31, memiliki 36 *attribute* antara lain *id*, *created_at*, *date*, *time*, *username*, *tweet*, dan sebagainya. Tweet tersebut diambil dalam bentuk teks bahasa Indonesia berdasarkan kata kunci masing-masing kalimat. Teknik *crawling* ini menghasilkan tweet data dalam format *.csv*. Data kemudian beralih ke tahap *text preprocessing* untuk meningkatkan struktur dari data. Hasil *crawling tweet* ditunjukkan pada Tabel 1.

Tabel 1. Sample Tweet Hasil Crawling.

No	Tweet
1.	Cukup dong bikin rakyat susah, masa pemulihan ekonomi pasca pandemi aja msh banyak yg nganggur, apalagi bbm naik imbas pasti harga bahan pokok jg naik 😞 https://t.co/rNWkMVMcmC #PKStolakkenaikanBBM https://t.co/0GprYTDAwD
2.	Negara ini akan hancur bila BBM naik #TolakBBMnaik
7899 Nilai Tukar Rupiah Hari Ini, 11 Agustus 2022: Bikin Dolar AS dan Mata Uang Dunia Keok Parah! https://t.co/ryPeMpTxYp

Tabel 3. Hasil Slangword

Tahapan	Hasil
preprocessing	[T1] masyarakat napas pasca pandemi udah engap n ngosan barang bahan pokok efek bbm perintah solusi rakyat bencana tambah beban [T2] bikin rakyat susah pulih ekonomi pasca pandemi aja msh yg nganggur bbm imbas harga bahan pokok jg
slangword	[T1] masyarakat nafas pasca pandemi sudah pengap dan memburu barang bahan pokok efek bbm perintah solusi rakyat bencana semakin beban [T2] bikin rakyat susah pulih ekonomi pasca pandemi saja masih yang nganggur bbm imbas harga bahan pokok juga

3.2. Text preprocessing

Tahap dasar yang digunakan dalam analisis sentimen adalah preprocessing yang membantu untuk meningkatkan kinerja dan merupakan langkah penting dalam penambangan data. Dalam penelitian ini, *clearning teks, case folding, tokenization, stopword removal* dan *stemming*.

Tujuan dari tahapan ini adalah untuk menganalisis hanya kata-kata penting. Setelah menyelesaikan tahap preprocessing jumlah data *tweet* menjadi 5989. Hasil dari tahapan preprocessing dapat dilihat pada Tabel 2.

Tabel 2. Hasil dari Text Preprocessing

No	Tahapan	Hasil text preprocessing
1	Kalimat asli	@ghozhanif22 @shidicious Kirain gw doang ngerasain gini. Boros banget pas awal2 bbm naik, pas berita heboh terus ngisi lagi kerasa iritnya kaya sebelum pertalet naik anjir
2	Cleaning text	kirain gw doang ngerasain gini boros banget pas awal bbm naik pas berita heboh terus ngisi lagi kerasa iritnya kaya sebelum pertalet naik anjir
3	Tokenization	['kirain', 'gw', 'doang', 'ngerasain', 'gini', 'boros', 'banget', 'pas', 'awal', 'bbm', 'naik', 'pas', 'berita', 'heboh', 'terus', 'ngisi', 'lagi', 'kerasa', 'iritnya', 'kaya', 'sebelum', 'pertalet', 'naik', 'anjir']
4	Stopword Removal	['kirain', 'gw', 'doang', 'ngerasain', 'gini', 'boros', 'banget', 'bbm', 'berita', 'heboh', 'ngisi', 'kerasa', 'iritnya', 'kaya', 'pertalet', 'anjir']
5	Stemming	kirain gw doang ngerasain gin boros banget bbm berita heboh ngisi rasa irit kaya pertalet anjir

3.3. Tahap slangword

Slangword dimodifikasi pada tahapan ini. Selama tahapan ini, kata-kata tidak baku diubah menjadi baku sesuai dengan KBBI yang relevan. Proses Tabel 3 mengilustrasikan bagaimana kata slang telah berubah.

3.4. Pelabelan Data

Setelah melalui tahap preprocessing selanjutnya pada tahap ini dilakukan proses pelabelan data, dengan menggunakan *library vader* setiap tweet akan di berikan skor polaritas yang akan menunjukkan apakah sentimen *negative, positive* atau *neutral*. Jika skor total atau *compound score* lebih dari 0,05 maka akan diberi label positive jika kurang dari 0,05 akan menerima label negative selain dari itu akan mendapatkan label neutral. Hasil pelabelan data ditampilkan pada Tabel 4.

Tabel 4. Sample Hasil Pelabelan Data

Tweet_Clean	Pos	Neg	Neu	Compound	Sentiment
gojek, shopeefood bad gojek bbm, meter different, shopeefood increase meter, still thousands kilometers target, incentive bonus down, quite sad bbm poor people told save salaries dpr price fuel put million poor gods	0.284	0.236	0.48	0.1045	Positive
	0.283	0.274	0.442	0.0258	Neutral
	0	0.383	0.617	-0.4767	Negative

Jumlah data setelah melakukan pelabelan dapat dilihat *persentase* dari jumlah data 5989 terdapat 2508 data yang memiliki sentimen *positive*, 1804 data sentimen *negative* dan 1677 data sentimen *neutral*. Setelah didapatkan semua sentimen pada data tweet, kemudian di visualisasikan dengan menampilkan informasi menggunakan diagram lingkaran dan bar plot. Visualisasi data menggunakan diagram lingkaran dari hasil pelabelan data dapat dilihat pada Gambar 6. Sedangkan visualisasi data menggunakan bar chart dari hasil tahapan pelabelan data dapat dilihat pada Gambar 7.

banyak penyediaan barang berkurang karena kebijakan pembatasan.

3.6. Ekstraksi Fitur

Dokumen teks yang sebelumnya telah melalui banyak proses kemudian ditambahkan ke dalam proses pembobotan kata dengan menggunakan algoritma *Term Frequency – Invers Document Frequency* (TF-IDF). Dimana jika nilai yang didapatkan dari tahap pembobotan kata semakin tinggi bobot katanya maka mengindikasikan bahwa kata tersebut semakin layak digunakan sebagai keyword terhadap kalimat tersebut. Tahapan ekstraksi fitur menggunakan TF-IDF dapat dilihat pada Gambar 11.

```
vectorizer = TfidfVectorizer()
train_vectors = vectorizer.fit_transform(data_train['Tweet Clean'])
test_vectors = vectorizer.transform(data_test['Tweet Clean'])
print (train_vectors)
```

Gambar 11. Source Code TF-IDF

Pada Gambar 11 baris pertama kode tersebut berfungsi untuk membuat variabel vectorizer berisi `TfidfVectorizer()`, kemudian pada baris kedua variabel tersebut akan diterapkan pada data frame pada kolom `Tweet_Clean` yang berisi teks tweet yang sudah melalui proses *Text Preprocessing*. Pada baris ketiga digunakan untuk membuat variabel baru yaitu `test_vectors` yang berisi data yang sudah ditransformasi menggunakan `tfidf`.

3.7. Klasifikasi K-Nearest Neighbor (K-NN)

Tahapan sebelum melakukan klasifikasi data dibagi menjadi data training dan data testing. Proses pembagian data menggunakan modul `train_test_split` dari *library sklearn.model_selection* di python. Rasio pembagian data training dan data testing adalah 90:10, dimana 90% untuk data training dan 10% untuk data testing. Data training digunakan untuk menilai model sedangkan data training digunakan untuk membangun model atau mengidentifikasi pola. Setelah data telah dibagi kemudian dapat digunakan dalam proses klasifikasi. Pengujian pertama untuk melakukan klasifikasi pengujian menggunakan algoritma *K-Nearest Neighbor* (K-NN). Hasil didapatkan dari klasifikasi algoritma *K-Nearest Neighbor* (K-NN) dapat dilihat pada Gambar 13. Hasil dari algoritma *K-Nearest Neighbor* (K-NN) menyatakan bahwa tingkat *accuracy* pada metode ini adalah sebesar 54%, dimana nilai *precision negative* 60%, nilai *precision positive* 73%, nilai *precision neutral* 39%, nilai *recall negative* 51%, nilai *recall positive* 45%, nilai *recall neutral* 71%, dan nilai *f1-score positive* 56%, *f1-score negative* 55%, dan *f1-score neutral* 50%. Hasil klasifikasi dari K-NN dapat dilihat pada Gambar 12.

	precision	recall	f1-score	support
Negative	0.60	0.51	0.55	184
Neutral	0.39	0.71	0.50	156
Positive	0.73	0.45	0.56	259
accuracy			0.54	599
macro avg	0.57	0.56	0.54	599
weighted avg	0.60	0.54	0.54	599

Hasil Klasifikasi Sentimen Analisis Inflasi:
 Accuracy: 0.5358931552587646
 Precision: 0.5743229738047702
 Recall: 0.5567614915441003
 f1_score: 0.537503518153673
 error_rate: 0.4641068447412354

Gambar 12. Hasil Klasifikasi K-NN

3.8. Klasifikasi Support Vector Machine (SVM)

Klasifikasi kedua dilakukan dengan algoritma *Support Vector Machine* (SVM) dengan menggunakan *library sklearn* di python. Pada model pendekatan yang digunakan dalam penelitian ini adalah kernel linear dan radial. Hasil dari algoritma SVM menyatakan bahwa tingkat *accuracy* pada metode ini paling tinggi yaitu kernel linear sebesar 79%, dimana nilai *precision negative* 76%, nilai *precision positive* 89%, nilai *precision neutral* 70%, nilai *recall negative* 72%, nilai *recall positive* 83%, nilai *recall neutral* 82%, dan nilai *f1-score positive* 86%, *f1-score negative* 74%, dan *f1-score neutral* 75%.

Pada penelitian ini hasil klasifikasi SVM dengan menggunakan kernel linear mendapatkan akurasi yang paling tinggi yaitu 79% dapat dilihat dari Gambar 14.

Here is the classification report:

	precision	recall	f1-score	support
Negative	0.76	0.72	0.74	180
Neutral	0.70	0.82	0.75	168
Positive	0.89	0.83	0.86	251
accuracy			0.79	599
macro avg	0.79	0.79	0.79	599
weighted avg	0.80	0.79	0.79	599

Hasil Klasifikasi Sentimen Analisis Inflasi:
 Accuracy Linear : 79.29883138564274 %
 Precision Linear : 0.7851672944512339
 Recall Linear : 0.788794557220852

Gambar 13. Hasil Klasifikasi SVM Kernel Linear

Sedangkan kernel RBF mendapatkan akurasi sebesar 75% memiliki sedikit selisih dari kernel Linear yaitu 4% dapat dilihat pada Gambar 15. Dari penelitian ini bahwa klasifikasi SVM menggunakan kernel RBF menghasilkan akurasi yang rendah dibandingkan akurasi menggunakan kernel Linear.

```

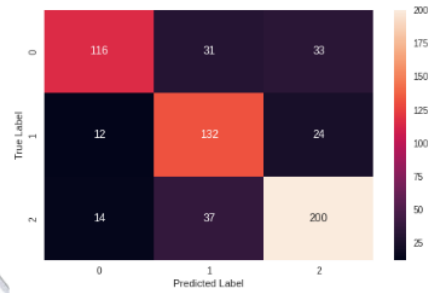
Here is the classification report:
              precision    recall  f1-score   support

 Negative     0.82     0.64     0.72     180
 Neutral      0.66     0.79     0.72     168
 Positive     0.78     0.80     0.79     251

 accuracy          0.75     599
 macro avg         0.75     0.74     0.74     599
 weighted avg      0.76     0.75     0.75     599
    
```

Hasil Klasifikasi Sentimen Analisis Inflasi:
 Accuracy RBF : 74.79131886477462 %
 Precision RBF : 0.7517038417274072
 Recall RBF : 0.7423238263875714

Gambar 14. Hasil Klasifikasi SVM Kernel Radial



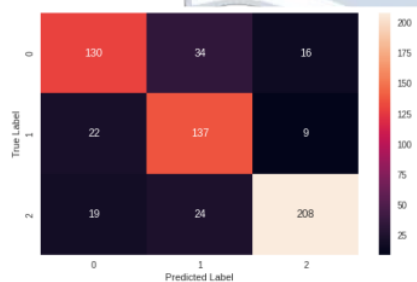
Gambar 17. Hasil Confusion Matrix SVM Kernel Radial

3.9. Validasi dan Evaluasi

Data training dan data testing adalah dua sub-proses yang membentuk proses validasi. Model algoritma yang dipilih selama tahap pemodelan dilatih menggunakan sub-proses data training. Setelah pemodelan algoritma dilatih pada tahap sub-proses training, selanjutnya akan dilakukan testing. Adapun evaluasi atau pengujian hasil klasifikasi *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) menggunakan *confusion matrix* yang merupakan alat untuk melakukan analisis terhadap seberapa baik klasifikasi yang telah dihasilkan dan mengenali *tuple* dari kelas yang berbeda. Hasil *confusion matrix* metode *K-Nearest Neighbor* dapat dilihat pada Gambar 15 dan *confusion matrix* metode SVM dengan kernel linear pada Gambar 16 dan hasil dari kernel radial dapat dilihat pada Gambar 17.



Gambar 15. Hasil Confusion Matrix K-NN



Gambar 16. Hasil Confusion Matrix SVM Kernel Linear

Pada Gambar confusion matrix tersebut merupakan hasil dari matrix yang dihasilkan dan divisualisasikan dalam bentuk confusional matrix. Setelah proses yang sudah dilakukan maka dapat digambarkan dalam bentuk tabel matrix evaluasi dapat dilihat pada Tabel 5 sebagai berikut :

Tabel 5. Matrix Hasil Evaluasi

Model Algoritma	K-NN	SVM (Kernel Linear)	SVM (Kernel RBF)
Akurasi	54%	79%	75%
Precision	57%	78%	75%
Recall	55%	78%	74%

4. DISKUSI

Penelitian sebelumnya yang menghasilkan banyak sumber hasil atau analisis konseptual menjadi dasar untuk penelitian ini. Untuk melakukan penelitian ini, akan sangat penting untuk memiliki data-data dari penelitian sebelumnya. Data pendukung tersebut dapat dijadikan pertimbangan untuk mengeksplorasi dalam penelitian tentang inflasi pasca COVID-19 berdasarkan media sosial Twitter menggunakan teknik klasifikasi text mining dengan algoritma *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM).

Pada penelitian terdahulu oleh Fuji Astuti dan Resi Taufan dengan judul penelitian "Sentimen Analisis Vaksinasi Covid-19 pada Twitter Menggunakan Algoritma Klasifikasi Berbasis PSO" menjelaskan dari hasil pengujian akhir dengan dua algoritma klasifikasi yaitu Naive Bayes (NB) dan Support Vector Machine (SVM) kemudian dibandingkan menggunakan Particle Swarm Optimization (PSO). Hasil akurasi NB dan SVM adalah 64,04% dan 72,55%, sedangkan hasil akurasi setelah menggunakan PSO pada algoritma SVM dan NB adalah 70,43% dan 76,38%[22].

Arief Asro dan Herny Februariyanti dengan judul penelitian "Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor" menjelaskan bahwa penelitian

tersebut menggunakan algoritma K-Nearest Neighbor dengan jumlah data yang diperoleh sebanyak 6408 data tweet. Hasil training model membuktikan bahwa skor akurasi 69,5%, recall 69,5%, dan presisi 68,7%[23].

Didik Siswanto, Zamzami, Lasri Nijal, dan Safrul Rajab dengan judul penelitian "Analisa Sentimen Publik Mengenai Perekonomian Indonesia Pada Masa Pandemi COVID-19 di Twitter menggunakan Metode Klasifikasi K-NN dan SVM" menjelaskan bahwa metode yang dipakai dalam penelitian ini adalah metode K-Nearest Neighbors (K-NN) dan Support Vector Machine (SVM). Hasil pengujian menggunakan confusion matrix didapatkan akurasi dari analisis sentimen menggunakan metode K-NN sebanyak 76%. Sedangkan akurasi dari analisis sentimen menggunakan metode SVM sebanyak 78%[24].

5. KESIMPULAN

Pada penelitian ini terkumpul 5390 data training dan 599 data testing. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa nilai akurasi dari metode *K-Nearest Neighbor* sebesar 54%, sedangkan nilai akurasi dari metode *Support Vector Machine* dengan kernel sebesar 79%. Hasil yang dari kedua metode klasifikasi tersebut dalam penelitian dapat disimpulkan bahwa *Support Vector Machine* menghasilkan nilai akurasi yang lebih tinggi dari pada metode *K-Nearest Neighbor* dengan perbedaan selisih 25%. Namun, data yang lebih baik atau lebih bersih dapat digunakan untuk mendapatkan akurasi maksimum. Dapat juga melakukannya dengan menggabungkan lebih banyak data atau menggunakan metode lain untuk mendapatkan hasil yang lebih akurat.

Adapun saran mengenai ide untuk memperluas penelitian yang serupa adalah penentuan data training dapat mempengaruhi hasil klasifikasi, maka dari itu untuk penelitian diharapkan menambahkan atribut atau data training yang lebih lengkap karena penentuan tingkat akurat dapat dibentuk oleh data training. Karena data training dapat mempengaruhi hasil klasifikasi. Selain itu, pada saat pelabelan data juga ketika menggunakan vader sentiment hanya mengandung kata dalam bahasa Inggris juga mempunyai pengaruh ketika kata dalam bahasa Indonesia diterjemahkan kedalam bahasa Inggris dalam menyatakan sebuah sentimen pada kalimat.

DAFTAR PUSTAKA

- [1] A. I. Fahrika and J. Roy, "Dampak pandemi covid 19 terhadap perkembangan makro ekonomi di indonesia dan respon kebijakan yang ditempuh," *Inovasi*, vol. 16, no. 2, pp. 206–213, 2020.
- [2] D. Anggraeni, H. Sirait, D. Rahandri, U. D.

Persada, and U. Muhammadiyah, "DAMPAK INFLASI TERHADAP SEKTOR EKONOMI PASCAPANDEMI COVID-19," vol. 7, no. 7, 2022.

- [3] B. P. Statistik, "Ekonomi Indonesia Triwulan III-2022 Tumbuh 5,72 Persen (y-on-y)," *Badan Pusat Statistik*, 2022. <https://www.bps.go.id/>.
- [4] S. Bunga, D. A. N. Resesi, T. Kinerja, and M. H. Saputra, "SAHAM PERUSAHAAN PROPERTI DAN REAL ESTATE," vol. 11, no. 04, pp. 981–992, 2022.
- [5] U. Kurniasih and A. T. Suseno, "Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM)," vol. 6, pp. 2335–2340, 2022, doi: 10.30865/mib.v6i4.4958.
- [6] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [7] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," vol. 10, pp. 71–76, 2020.
- [8] A. Muzaki and A. Witanti, "SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER SENTIMEN ANALISIS MASYARAKAT DI TWITTER TERHADAP PILKADA 2020 DITENGAH PANDEMIC COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER," vol. 2, no. 2, pp. 101–107, 2021.
- [9] S. Lestari, M. Mupaat, and A. Erfina, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter," vol. 8, no. 1, 2022.
- [10] J. Jtik, J. Teknologi, T. W. Putra, and A. Triayudi, "Analisis Sentimen Pembelajaran Daring menggunakan Metode Naive Bayes, KNN, dan Decision Tree," vol. 6, no. 1, 2022.
- [11] B. Laurensz and Eko Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [12] M. Afdal and L. Rahma Elita, "Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, pp. 78–87, 2022.
- [13] U. Verawardina, F. Edi, and R. Watrianthos,

- “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [14] N. Hidayati, J. Suntoro, and G. G. Setiaji, “Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM,” *J. Sains dan Inform.*, vol. 7, no. 2, pp. 117–126, 2021, doi: 10.34128/jsi.v7i2.313.
- [15] A. Dwiki *et al.*, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN,” vol. 8, no. 2, pp. 636–646, 2021.
- [16] B. Pamungkas, M. E. Purbaya, and D. J. A. K, “Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada,” vol. 8106, pp. 10–20, 2021.
- [17] P. Arsi, R. Wahyudi, and R. Waluyo, “Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.
- [18] M. A. Nurrohmat and A. SN, “Sentiment Analysis of Novel Review Using Long Short-Term Memory Method,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.
- [19] A. Rahman *et al.*, “Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 8, no. 2, pp. 299–303, 2022, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/articel/view/54704>.
- [20] I. Yunanto and S. Yulianto, “TWITTER SENTIMENT ANALYSIS PEDULILINDUNGI APPLICATION USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE ANALISIS SENTIMEN TWITTER APLIKASI PEDULILINDUNGI,” vol. 3, no. 4, pp. 807–814, 2022.
- [21] Iskandar, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 158, pp. 1120–1126, 2021.
- [22] F. Astuti and R. Taufan, “Sentimen Analisis Vaksinasi Covid - 19 pada Twitter Menggunakan Algoritma Klasifikasi Berbasis PSO/ Sentiment Analysis of Covid-19 Vaccination on Twitter using Classification Algorithms based on PSO,” *Sist. J. Sist. InformasiEMASI*, vol. 11, pp. 364–376, 2022.
- [23] A. Asro’i and H. Februariyanti, “Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan Ppkm Menggunakan Metode K-Nearest Neighbor,” *J. Khatulistiwa Inform.*, vol. 10, no. 1, pp. 17–24, 2022, doi: 10.31294/jki.v10i1.12624.
- [24] D. Siswanto, L. Nijal, and S. Rajab, “Analisa Sentimen Publik Mengenai Perekonomian Indonesia Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Metode Klasifikasi K-Nn Dan Svm,” vol. 2, no. 1, pp. 1–9, 2022.

JURNAL RATIH PUSPITASARI

ORIGINALITY REPORT

13%

SIMILARITY INDEX

12%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	jutif.if.unsoed.ac.id Internet Source	8%
2	Submitted to Ciputra University Student Paper	2%
3	jurnal.iaii.or.id Internet Source	2%
4	pdfs.semanticscholar.org Internet Source	2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On