

Perbandingan Kinerja *Multinomial Naïve Bayes Classifier* dan *Naïve Forecasting* dalam Klasifikasi dan Peramalan Jumlah Opini Kenaikan Harga BBM

Nita Oktaviani¹, Embay Rohaeti^{2*}, Maya Widyastiti³, Ani Andriyati⁴

^{1,2,3,4} Program Studi Matematika, FMIPA, Universitas Pakuan, Bogor, Indonesia
Jl. Pakuan RT.02/RW.06 Tegallega, Kecamatan Bogor Tengah, Bogor, Indonesia
Korespondensi penulis: embay.rohaeti@unpak.ac.id

Abstrak

Data opini pengguna X terhadap topik kenaikan harga Bahan Bakar Minyak (BBM) memiliki nilai sentimen yang dapat menentukan kelas opini dominan sebagai gambaran penilaian dari sisi masyarakat pengguna X. Opini pengguna X akan diklasifikasi dalam tiga kelas yaitu kelas opini positif, negatif dan netral menggunakan model *Multinomial Naïve Bayes Classifier* (MNBC). Hasil klasifikasi yang diperoleh dilanjutkan pada tahapan peramalan jumlah opini dengan metode *Naïve Forecasting* (NF). Tujuan dari penelitian ini yaitu mengklasifikasikan opini dengan MNBC, meramalkan jumlah opini untuk jangka waktu satu minggu kedepan dengan NF, dan mengevaluasi hasil klasifikasi serta hasil peramalan. MNBC merupakan salah satu algoritma machine learning yang digunakan untuk klasifikasi teks. NF merupakan salah satu metode peramalan untuk data time series. Perhitungan pada penelitian ini dilakukan dengan penggunaan bantuan software R. Data yang digunakan berupa data sekunder sebanyak 2500 data. Periode pengambilan data selama satu minggu dimulai dari 20 Oktober 2022 hingga 27 Oktober 2022. Hasil dari pemodelan MNBC didapatkan tiga kelas yaitu sebanyak 775 dokumen diklasifikasikan sebagai opini negatif, 475 dokumen diklasifikasikan sebagai opini netral, dan 581 dokumen diklasifikasikan sebagai opini positif. Akurasi model MNBC dikategorikan sangat baik sebesar 92% untuk keseluruhan kelas. Hasil peramalan tiga kelas klasifikasi dengan NF yaitu jumlah opini kelas positif sebanyak 44 opini dengan nilai RMSE sebesar 8,96, jumlah opini kelas negatif sebanyak 25 opini dengan nilai RMSE sebesar 14,87, dan jumlah opini kelas netral sebanyak 21 opini dengan nilai RMSE sebesar 11,45. Hal ini menunjukkan Peramalan dengan NF dikategorikan cukup baik.

Kata Kunci: MNBC, NF, Opini, Klasifikasi, Peramalan.

Abstract

X user opinion data on the topic of fuel oil (BBM) price increases has a sentiment value that can determine the dominant opinion class as an illustration of the assessment from the X user community. X users' opinions will be classified into three classes, namely positive, negative and neutral opinion classes using the Multinomial Naïve Bayes Classifier (MNBC) model. The classification results obtained were continued at the stage of forecasting the number of opinions using the naïve forecasting (NF) method. The aim of this research is to classify opinions with MNBC, predict the number of opinions for the next week with NF, and disseminate the classification results and forecasting results. MNBC is a machine learning algorithm used for text classification. NF is a forecasting method for time series data. Calculations in this research were carried out with the help of R software. The data used was 2500 secondary data. The one week data collection period starts from October 20 2022 to October 27 2022. The results of the MNBC modeling obtained three classes, namely 775 documents classified as negative opinion, 475 documents classified as neutral opinion, and 581 documents classified as positive opinion. The accuracy of the MNBC model is rated as very good at 92% for the entire class. The results of forecasting three classification classes with NF are the number of positive class opinions of 44 opinions with an RMSE value of 8.96, the number of negative class opinions of 25 opinions with an RMSE value of 14.87, and the number of neutral class opinions of 21 opinions with an RMSE value of 11.45. This shows that forecasting with NF is quite good.

Keywords: MNBC, NF, Opinion, Classification, Forecasting.

1. Pendahuluan

Subsidi bahan bakar minyak merupakan penurunan harga bahan bakar minyak yang ditetapkan oleh pemerintah. Hadirnya subsidi bahan bakar berpengaruh pada beberapa aspek ekonomi seperti harga kebutuhan pokok melalui biaya produksi, dan tarif angkutan umum. Sasaran subsidi bahan bakar minyak yaitu untuk membantu masyarakat kurang mampu, namun diatas 70% subsidi bahan bakar minyak dipakai oleh masyarakat mampu dengan mobil pribadi [1], [2]. Kondisi ini berakibat pada penyaluran subsidi bahan bakar minyak yang tidak tepat sasaran [3], [4]. Oleh karena itu melalui siaran pers yang digelar di Istana Negara, pemerintah menetapkan kebijakan pengalihan subsidi bahan bakar minyak. Kebijakan yang ditetapkan yaitu menaikkan harga bahan bakar minyak sejak 3 September 2022 untuk mengatasi permasalahan tersebut [5].

Perubahan harga bahan bakar minyak menjadi ramai dibicarakan masyarakat terutama melalui media sosial salah satunya Twitter yang saat ini berganti nama menjadi X. Media sosial X menjadi incaran masyarakat Indonesia dalam menyebarkan opini, maupun memverifikasi ketepatan informasi yang sedang terjadi secara *real time*. Oleh karena itu Indonesia menempati peringkat keenam sebagai negara dengan jumlah pengguna X terbesar di dunia. Jumlah pengguna X aktif di Indonesia mencapai 14,8 juta pada April 2023 [6]. Melalui media sosial X pengguna dapat menemukan informasi populer melalui fitur top trending. Pengguna dapat saling memberi komentar maupun beradu opini dengan pengguna lain secara *real time*. Masyarakat yang biasa saja akan perubahan harga bahan bakar minyak menyuarakan opini netral, namun ada pula masyarakat yang menyuarakan rasa kecewa atau penolakan melalui opini negatif, ataupun menyatakan dukungan melalui opini positif yang saling memancing komentar

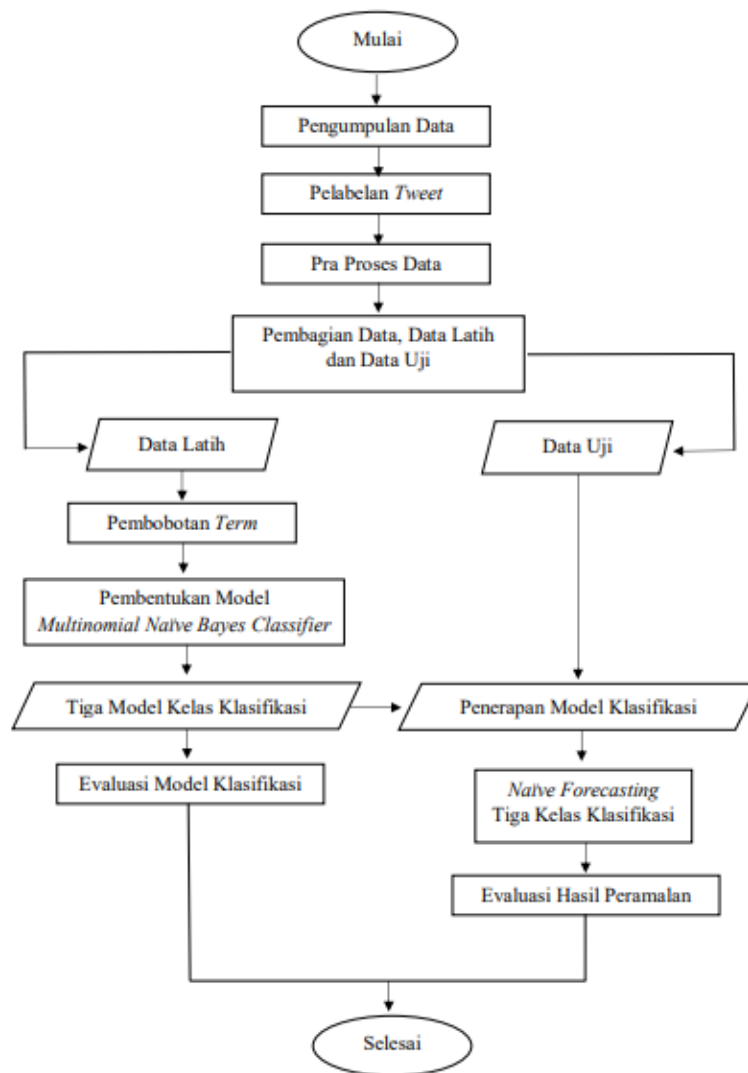
dari pengguna lain untuk berpendapat. Kumpulan opini masyarakat baik positif, negatif maupun netral mengenai kenaikan harga bahan bakar minyak dapat memberikan gambaran berupa dampak dari sisi masyarakat, namun tidak bisa dinilai secara langsung dikarenakan keberagaman opini yang ada.

Berdasarkan hal tersebut, perlu dilakukan penelitian untuk mengetahui topik pembicaraan kenaikan harga bahan bakar minyak didominasi oleh tanggapan masyarakat yang positif, negatif atau netral berdasarkan opini di media sosial X. Beberapa penelitian terdahulu telah melakukan ini, yaitu dengan menggunakan berbagai jenis metode untuk mengklasifikasikan tanggapan masyarakat. [7], [8], [9] mengklasifikasikan sentimen masyarakat di X terhadap kenaikan harga BBM menggunakan *Support Vector Machine* (SVM). Pada topik yang sama, [10], [11], [12] menggunakan *K-nearest neighbors* (KNN), sementara [13], [14], [15], [16], [17] menggunakan Metode Naïve Bayes Classifier. [18] menggunakan pendekatan yang sedikit berbeda, yaitu dengan menggunakan model *adaptive boosting*. Beberapa penelitian bahkan membandingkan beberapa metode, seperti [19] yang membandingkan Naïve Bayes dan Random Forest dan [20] yang membandingkan Naïve Bayes, SVM, dan KNN untuk mengklasifikasikan sentimen masyarakat.

Berbeda dengan penelitian terdahulu, selain mengklasifikasikan tanggapan masyarakat, penelitian ini juga melakukan peramalan terkait persebaran opini di waktu yang akan datang. Pada penelitian ini digunakan peubah bebas yaitu opini pengguna X terhadap topik kenaikan harga bahan bakar minyak, sedangkan peubah tak bebas yaitu label positif, negatif, dan netral dari setiap *tweet*. Metode yang digunakan dalam klasifikasi jumlah opini pada penelitian ini yaitu *Multinomial Naïve Bayes Classifier* (MNBC), MNBC adalah salah satu metode dalam pengklasifikasian dengan dasar metode probabilitas bersyarat dan metode statistik [21], [22], [23]. Dalam peramalan pada penelitian ini digunakan metode *Naïve Forecasting* (NF), NF merupakan metode peramalan *time series* yang berfokus pada variabel masa lalu untuk memprediksi nilai masa kini [24], [25], [26]. Kinerja metode NF efektif dalam melakukan peramalan walaupun metode ini termasuk dalam kategori sederhana [27].

2. Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa *tweet* pengguna X pada tanggal 20 sampai 27 Oktober 2022. Data diambil dalam jangka waktu satu bulan setelah harga bahan bakar minyak dinaikkan dan topik kenaikan harga bahan bakar minyak sedang menjadi *trending* topik di media sosial X. Data diambil sebanyak 2500 data yang mengandung kata kunci kenaikan harga bahan bakar minyak dan diambil melalui *crawling* data X. Data yang digunakan dibagi menjadi 80% data latih dan 20% data uji. Data latih digunakan untuk klasifikasi jumlah opini dan data uji digunakan untuk menerapkan model klasifikasi dalam peramalan jumlah opini. Secara umum proses klasifikasi dan peramalan opini X terhadap topik kenaikan harga bahan bakar minyak menggunakan metode MNBC dan NF digambarkan secara skematik pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Analisis

Berikut merupakan beberapa teori pendukung dalam penelitian ini.

Lexicon Based

Lexicon based merupakan suatu pendekatan dalam penentuan nilai suatu sentimen [28]. Nilai yang dihasilkan akan merujuk pada suatu kategori seperti positif, negatif dan netral [29]. Bentuk persamaan dalam perhitungan nilai sentimen sebagai berikut.

$$\text{NilaiPos}(u) = \sum_{i=1}^m \text{NilaiTerm}(T_i) \quad (1)$$

$$\text{NilaiNeg}(u) = \sum_{j=1}^n \text{NilaiTerm}(T_j) \quad (2)$$

$$\text{NilaiSentimen}(u) = \text{NilaiPos}(u) + \text{NilaiNeg}(u) \quad (3)$$

Suatu ulasan (u) memiliki nilai ulasan positif dihitung berdasarkan jumlah nilai term positif (T_i) dengan (m) term positif. Ulasan (u) memiliki nilai ulasan negatif dihitung berdasarkan jumlah nilai term negatif (T_j) dengan (n) term negatif. Nilai sentimen yang dihasilkan diperoleh melalui penjumlahan ulasan bernilai positif dengan ulasan bernilai negatif.

Term Frequency - Inverse Document Frequency (TF – IDF)

Pembobotan dengan TF-IDF akan menghasilkan matriks pembobot pada setiap *term* yang bergantung pada jumlah kemunculan *term* pada sebuah dokumen [30]. Pada Persamaan 4, Persamaan 5, dan Persamaan 6 merupakan skema dari TF-IDF [31].

$$tf_{(t,d)} = \begin{cases} 1 & , \text{if } tf_{(t,d)} > 0 \\ 0 & , \text{lainnya} \end{cases} \quad (4)$$

$$idf_t = \log \frac{N}{df_t} \quad (5)$$

$$tf - idf_{t,d} = tf_{t,d} \times idf_t \quad (6)$$

$tf_{(t,d)}$ merupakan nilai *term frequency* dari suatu *term* (t) pada dokumen (d). $tf_{(t,d)}$ bernilai 1 jika suatu *term* (t) termasuk ke dalam dokumen (d) dan sebaliknya. Sementara idf_t merupakan nilai *inverse document frequency* dari suatu *term* (t). idf_t diperoleh dari hasil logaritma antara jumlah dokumen pada dataset (N) dibagi dengan nilai *document frequency* dari suatu *term* (t) yang dilambangkan dengan df_t .

Multinomial Naïve Bayes Classifier (MNBC)

Pada Persamaan 7 menunjukkan MNBC untuk suatu dokumen d yang berada di kelas c [30].

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (7)$$

Untuk menemukan probabilitas dokumen (d) pada kelas (c) $P(c|d)$ sebagai kelas terbaik dokumen diperoleh dengan memaksimalkan probabilitas suatu dokumen yang berada di kelas (c) atau disebut probabilitas prior $P(c)$ dikalikan dengan probabilitas bersyarat *term* t_k yang berada dalam dokumen pada kelas (c).

Naïve Forecasting (NF)

Pada Persamaan 8 menunjukkan NF dengan asumsi bahwa kejadian sebelumnya merupakan peramalan terbaik untuk hari ini [32].

$$Naïve = X_{t-1} \quad (8)$$

Data aktual pada periode waktu (t) dilambangkan dengan X_t . Data sebelumnya sebelum data aktual pada NF dilambangkan dengan X_{t-1}

Confusion Matrix

Confusion matrix merupakan matriks yang berisi nilai prediksi benar dan salah [33], [34]. Model confusion matrix untuk multikelas klasifikasi disajikan pada Tabel 1.

Tabel 1. *Confusion Matrix* Klasifikasi Multikelas

		Kelas Prediksi			
		Kelas 1	Kelas 2	...	Kelas k
Kelas Aktual	Kelas 1	f_{11}	f_{12}	...	f_{1k}
	Kelas 2	f_{21}	f_{22}	...	f_{2k}
	⋮	⋮	⋮	⋮	
	Kelas k	f_{k1}	f_{k2}	...	f_{kk}

Pada Persamaan (9), Persamaan (10), Persamaan (11) dan Persamaan (12) merupakan ukuran akurasi yang dipakai dalam mengukur performa model klasifikasi yang akan dievaluasi [35].

$$Accuracy = \frac{\sum_{j=1}^m f_{jj}}{\sum_{k=1}^m \sum_{j=1}^m f_{jk}} \quad (9)$$

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (10)$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (11)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (12)$$

Root Mean Squared Error (RMSE)

Nilai RMSE berkisar dari nol hingga positif tak hingga, semakin besar nilai RMSE kesalahan model dalam memprediksi semakin tinggi [36]. Pada Persamaan (13) merupakan bentuk umum dari RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

Berdasarkan persamaan diatas, Nilai data sebenarnya dilambangkan dengan y_i dan nilai hasil peramalan dilambangkan dengan \hat{y}_i . Jumlah observasi yang dilakukan dilambangkan dengan N.

3. Hasil dan Pembahasan

3.1 Pelabelan Tweet

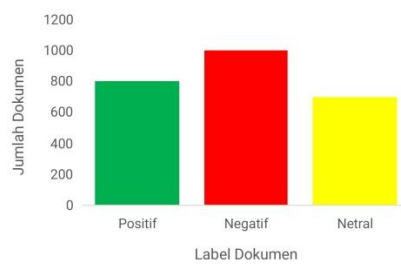
Pada Tabel 2 disajikan hasil pelabelan dari contoh unggahan pengguna X terkait kenaikan harga bahan bakar minyak sebagai berikut.

Tabel 2. Hasil Pelabelan dari Contoh Unggahan Pengguna X

Nama Pengguna	Unggahan pengguna	Label
NanangSuhen32	Pemerintah Memberikan BLT Bagi Warga yang Terdampak kenaikan Harga BBM. Lampung Dukung Kebijakan Pemerintah	Netral
Irfqn_	Hingga kini kemiskinan masih menjadi problem utama di Indonesia yang belum terselesaikan, terlebih di tengah naiknya berbagai bahan pokok, seperti kenaikan beras, telur, tarif listrik, hingga yang terbaru adalah kenaikan BBM, yang sudah pasti mendongkrak kenaikan harga-harga lain	Negatif
BorobudurNews_	RT @BorobudurNews_: Polres Magelang Kota Salurkan Bantuan 1 Ton Beras ke Warga Terdampak Kenaikan Harga BBM https://t.co/aDQ1yMwp9E	Positif
Harmawa17	@democrazymedia Mayoritas puas atas kenaikan bbm dan kenaikan harga bahan pokok	Negatif
kurangbayar	@prastow kenaikan penerimaan diakui sebagai keberhasilan pemerintah, kenaikan bebannya diserahkan untuk ditanggung oleh rakyat (ngurangi subsidi, kenaikan bbm, pelemahan rupiah) ... emang keren ... konoha ...	Positif
hariankompas	Pengeluaran konsumsi penduduk per kapita per bulan selalu meningkat. Pada tahun ini, peningkatannya bisa lebih tinggi lantaran inflasi tinggi yang disebabkan oleh kenaikan harga pangan dan BBM.	Netral
pak_ogah77	Ahlan wa Sahlan Ramadhan dan Kenaikan BBM semoga semakin sabarrrrrr #TolakBBMNaik ##tolakBBMnaik	Netral

abengkriiss	Meski, di sisi lain, harga kebutuhan pokok memang sempat terpengaruh kenaikan harga bahan bakar minyak (BBM) September lalu.	Negatif
ipunge_koben	@ArmiSatria8 @Muhammad_Saewad Iya.. Saya menikmati, menikmati kenaikan BBM maksudnya.	Positif
WulanPuspitaz	RT @OposisiCerdas: Beda Sri Mulyani, Core: Penyebab Inflasi Bukan Rantai Pasok Pangan, Tapi Kenaikan Harga BBM	Negatif

Hasil pelabelan *tweet* pada contoh unggahan pengguna X menunjukkan setiap unggahan yang tulis oleh pengguna X memiliki nilai sentimen yang dikelompokkan menjadi tiga kelas yaitu kelas positif, negatif dan netral. Grafik hasil pelabelan *tweet* secara keseluruhan data disajikan pada Gambar 2 sebagai berikut.

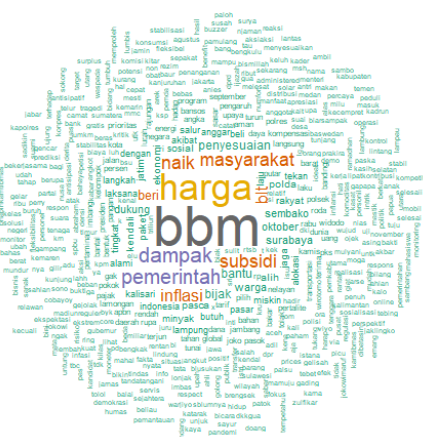


Gambar 2. Grafik Hasil Pelabelan *Tweet*

Hasil pelabelan dokumen diperoleh bahwa label dokumen terbanyak ialah kelas negatif sebanyak 1001, lalu kelas positif sebanyak 786 dan kelas netral sebanyak 713.

3.2 Pra Proses Data

Tahapan pra proses data yang dilakukan pada penelitian ini meliputi pembersihan data, *case folding*, penghapusan *stopwords*, *stemming* dan tokenisasi. Hasil tokenisasi yang telah divisualisasikan dengan *wordcloud* disajikan pada Gambar 3.



Gambar 3. Visualisasi Term Terpopuler dalam *Wordcloud*

Gambar 3 menunjukkan hasil visualisasi frekuensi kemunculan *term* secara keseluruhan menggunakan *wordcloud*. *Term* 'bbm' dengan ukuran paling besar merupakan *term* yang paling banyak dibicarakan. *Term* 'harga', 'dampak' dan 'pemerintah' juga termasuk *term* yang banyak dibicarakan setelah *term* 'bbm'.

3.3 Pembobotan Data

Pada Tabel 3 disajikan hasil perhitungan TF-IDF pada data latih.

Tabel 3. Hasil TF-IDF Pada Data Latih

Dokumen	Term								
	polda	jatim	paket	sembako	nelayan	...	dampak	naik	bbm
1	0,071	0,072	0,071	0,065	0,089	...	0,028	0,033	0
2	0	0	0	0	0	...	0	0	0
3	0,071	0,072	0,071	0,065	0,089	...	0,028	0,033	0
4	0,071	0,072	0,071	0,065	0,089	...	0,028	0,033	0
...
1998	0	0	0	0	0	...	0	0	0
1999	0	0	0	0	0	...	0,018	0,021	0
2000	0	0	0	0	0	...	0	0	0,001
2001	0	0	0	0	0	...	0	0	0

Berdasarkan Tabel 3 menunjukkan seluruh *term* pada data latih telah memiliki bobot sesuai dengan frekuensi kemunculan *term* pada dokumen. Hasil pembobotan TF-IDF bernilai rendah ketika *term* muncul di hampir semua dokumen dan bernilai tinggi ketika frekuensi kemunculan relatif sedikit.

3.3 Model Multinomial Naïve Bayes Classifier

Model *multinomial naïve bayes classifier* merupakan model probabilistik. Model *multinomial naïve bayes* yang akan diperoleh terdiri dari probabilitas *prior* dan probabilitas kelas. Probabilitas *prior* dan probabilitas kelas merupakan parameter berbeda yang akan diestimasi dari data latih.

3.3.1 Perhitungan Probabilitas Prior

Pada Tabel 4 disajikan hasil perhitungan probabilitas *prior* sebagai berikut.

Tabel 4. Hasil Perhitungan Probabilitas Prior

Label	Probabilitas Prior
	$\hat{P}(c) = \frac{N_c}{N}$
Positif	$\hat{P}(positif) = \frac{629}{2001} = 0,314$
Negatif	$\hat{P}(negatif) = \frac{801}{2001} = 0,400$
Netral	$\hat{P}(netral) = \frac{571}{2001} = 0,285$

Tabel 3 menunjukkan hasil perhitungan probabilitas *prior* untuk ketiga kelas. Kelas negatif memiliki hasil perhitungan tertinggi sebesar 0,400, sementara kelas netral memiliki hasil perhitungan terendah sebesar 0,285.

3.3.2 Perhitungan Probabilitas Kelas

Hasil dari perhitungan probabilitas kelas disajikan pada Tabel 5.

Tabel 5. Hasil Perhitungan Probabilitas Kelas

Term	Probabilitas Kelas		
	Negatif	Netral	Positif
polda	0,0003	0,003	0,002
jatim	0,0002	0,003	0,002
paket	0,0003	0,003	0,002
sembako	0,001	0,003	0,002
nelayan	0,0003	0,0004	0,003
kalisari	0,0004	0,0004	0,005
surabaya	0,0003	0,003	0,004
dampak	0,001	0,004	0,003
naik	0,001	0,004	0,003
bbm	0,000	0,0005	0,0004
lurah	0,0003	0,0004	0,003
kamis	0,0003	0,0004	0,003
oktober	0,0009	0,003	0,002
⋮	⋮	⋮	⋮
warna	0,0003	0,0003	0,0005
saksi	0,0003	0,0003	0,0004

Pada Tabel 5 menunjukkan hasil perhitungan probabilitas kelas pada data latih. Hasil probabilitas *prior* dan probabilitas kelas akan diaplikasikan sebagai model probabilistik *multinomial naïve bayes* dalam melakukan klasifikasi.

3.3.3 Proses Klasifikasi Model MNBC

Pada Tabel 6 menunjukkan contoh tiga dokumen sebelum proses klasifikasi.

Tabel 6. Tiga Dokumen Sebelum Proses Klasifikasi

	Dokumen	Label	Klasifikasi
1	'polda', 'jatim', 'paket', 'sembako', 'nelayani', 'kalisari', 'surabaya', 'dampak', 'naik', 'bbm', 'lurah', 'kalisari', 'surabaya', 'kamis', 'oktober'	Positif	?
2	'efek', 'naik', 'bbm'	Netral	?
3	'laju', 'inflasi', 'ri', 'relatif', 'kendali', 'bbm', 'alami'	Negatif	?

Pada proses ini, dihitung probabilitas setiap kategori kelas dan dicari probabilitas tertingginya. Hasil perhitungan yang didapatkan disajikan pada Tabel 7.

Tabel 7. Hasil Klasifikasi Pada Tiga Data Latih

Dokumen	$\hat{P}(c d)$	Klasifikasi
1	Positif $\hat{P}(c_{positif} d_1) = 1,91886e^{-40}$	Positif
	Negatif $\hat{P}(c_{negatif} d_1) = 1,41293e^{-51}$	
	Netral $\hat{P}(c_{netral} d_1) = 3,23138e^{-44}$	
2	Netral $\hat{P}(c_{netral} d_2) = 7,5791e^{-10}$	Netral
	Positif $\hat{P}(c_{positif} d_2) = 1,40131e^{-10}$	
	Negatif $\hat{P}(c_{negatif} d_2) = 1,78656e^{-10}$	
3	Negatif $\hat{P}(c_{negatif} d_3) = 5,19979e^{-20}$	Negatif
	Positif $\hat{P}(c_{positif} d_3) = 6,72351e^{-24}$	
	Netral $\hat{P}(c_{netral} d_3) = 8,04461e^{-21}$	

Proses klasifikasi dilakukan dengan memilih nilai probabilitas kelas dokumen terbesar sebagai hasil klasifikasi. Hasil prediksi menunjukkan kesesuaian antara label data aktual dengan label hasil klasifikasi terpilih. Probabilitas kelas dokumen terpilih ditandai dengan warna merah pada penulisan.

3.4 Evaluasi Model MNBC

Hasil dari model confusion matrix keseluruhan dokumen terdapat pada Tabel 8 dan Tabel 9 menunjukkan performa model MNBC pada setiap kelas.

Tabel 8. Hasil Confusion Matrix Keseluruhan Kelas

Kelas Prediksi	Kelas Aktual		
	Negatif	Netral	Positif
Negatif	775	16	10
Netral	80	475	16
Positif	36	12	581

Pada Tabel 8 hasil prediksi benar atau terklasifikasi sama dengan kelas data sebenarnya ditandai dengan warna biru, sementara untuk kolom lainnya menunjukkan kesalahan hasil prediksi yang tidak sesuai dengan kelas data sebenarnya.

Confusion matrix digunakan dalam perhitungan performa model *multinomial naïve bayes classifier*. Ukuran akurasi yang digunakan yaitu *accuracy*, *precision*, *recall* dan *f1-score*. *Accuracy* dihitung untuk keseluruhan kelas, sementara *precision*, *recall* dan *f1-score* dihitung untuk masing-masing kelas. Perhitungan *accuracy* model *multinomial naïve bayes classifier* menggunakan persamaan (9) dengan hasil perhitungan *accuracy* sebagai berikut.

$$Accuracy = \frac{\sum_{j=1}^m f_{jj}}{\sum_{k=1}^m \sum_{j=1}^m f_{jk}}$$

$$Accuracy = \frac{775 + 475 + 581}{775 + 16 + 10 + 80 + 475 + 16 + 36 + 12 + 581}$$

$$Accuracy = 0,915042 \approx 92\%$$

Pada perhitungan *accuracy* diperoleh hasil sebesar 92%. Hasil *accuracy* yang diperoleh menunjukkan ketepatan akurasi yang dihasilkan model *multinomial naïve bayes classifier* dalam melakukan klasifikasi sebesar 92%.

Tabel 9. Performa Model MNBC Setiap Kelas

Ukuran	Kelas		
	Negatif	Netral	Positif
<i>Precision</i>	0,968	0,831	0,924
<i>Recall</i>	0,870	0,944	0,957
<i>F1-score</i>	0,916	0,885	0,940

Berdasarkan Tabel 9 menunjukkan hasil pengukuran performa model *multinomial naïve bayes classifier* pada setiap kelas. Hasil perhitungan *precision* menunjukkan tingkat kebaikan model dapat melakukan klasifikasi dengan benar, sementara *recall* menunjukkan tingkat kebaikan model dapat melakukan klasifikasi dengan benar untuk kelas positif, dan *f1-score* menunjukkan keseimbangan antara *recall* dan *precision*.

3.4 Peramalan Tiga Kelas Klasifikasi

Peramalan akan dibagi menjadi tiga kelas yaitu peramalan opini kelas positif, kelas negatif dan kelas netral. Data yang akan digunakan dalam peramalan yaitu data uji sejumlah 499 dokumen.

3.4.1 Tabulasi Data Jumlah Opini

Tabel 10 menunjukkan hasil tabulasi data yang dilakukan dengan pembuatan tabel frekuensi.

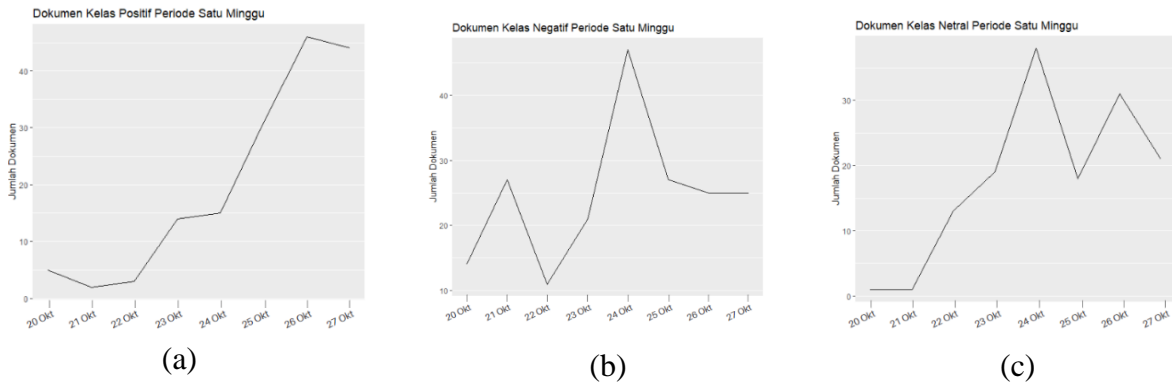
Tabel 10. Distribusi Frekuensi Data Jumlah Opini

Tanggal	Jumlah Dokumen		
	Positif	Negatif	Netral
20/10/2022	5	14	1
21/10/2022	2	27	1
22/10/2022	3	11	13
23/10/2022	14	21	19
24/10/2022	15	47	38
25/10/2022	31	27	18
26/10/2022	46	25	31
27/10/2022	44	25	21

Hasil tabulasi data menunjukkan frekuensi kemunculan dokumen yang termasuk pada kelas tertentu. Data setelah ditabulasi akan melalui proses peramalan dengan NF. Peramalan akan dilakukan pada masing-masing dokumen.

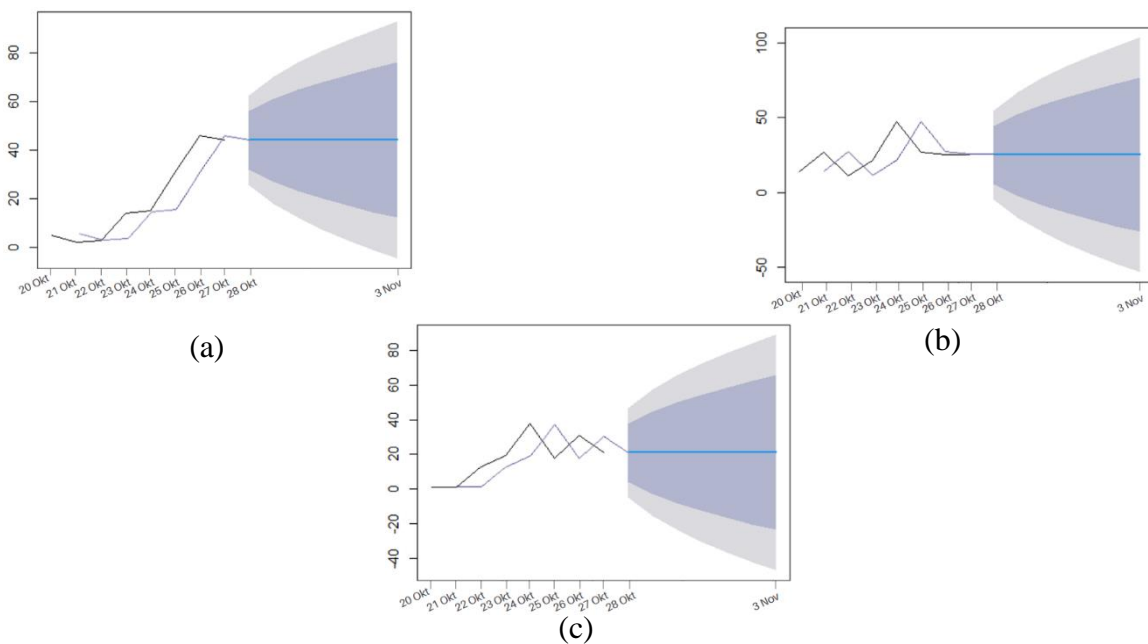
3.4.1 Proses Peramalan

NF merupakan salah satu metode peramalan runtun waktu. Data yang telah ditabulasi akan diubah kedalam data runtun waktu sebelum dilakukan peramalan. Data hasil klasifikasi setelah menjadi data runtun waktu disajikan pada Gambar 4.



Gambar 4. Grafik Pola Data Runtun Waktu Periode Satu Minggu

Gambar 4 menunjukkan grafik pola data runtun waktu periode waktu satu minggu. Pada grafik a) menunjukkan pola data *trend*, sementara grafik b) dan c) tidak menunjukkan pola tertentu. Frekuensi kemunculan dokumen dari setiap kelas selama periode waktu satu minggu memiliki kecenderungan meningkat. Untuk Grafik Hasil peramalan menggunakan *naïve forecasting* disajikan pada Gambar 5.



Gambar 5. Grafik Hasil Peramalan

Pada Gambar 5, garis berwarna hitam menunjukkan data aktual, sementara garis berwarna biru menunjukkan data hasil peramalan. Grafik a) merupakan hasil peramalan kelas positif, b) merupakan hasil peramalan kelas negatif dan c) merupakan hasil peramalan kelas netral. Berdasarkan ketiga grafik hasil peramalan tersebut, metode NF tidak memperhitungkan *trend*. NF hanya mengambil nilai pada siklus terakhir. Untuk hasil perhitungan NF disajikan pada Tabel 11.

Tabel 11. Hasil Perhitungan *Naïve Forecasting*

Tanggal	Data Ramal (X_{t-1})			Data Aktual (X_t)		
	Positif	Negatif	Netral	Positif	Negatif	Netral
20/10/2022	NA	NA	NA	5	14	1
21/10/2022	5	14	1	2	27	1
22/10/2022	2	27	1	3	11	13
23/10/2022	3	11	13	14	21	19
24/10/2022	14	21	19	15	47	38
25/10/2022	15	47	38	31	27	18
26/10/2022	31	27	18	46	25	31
27/10/2022	46	25	31	44	25	21
28/10/2022	44	25	21			
29/10/2022	44	25	21			
30/10/2022	44	25	21			
31/10/2022	44	25	21			
01/11/2022	44	25	21			
02/11/2022	44	25	21			
03/11/2022	44	25	21			

3.4.1 Evaluasi Hasil Peramalan

Hasil Perhitungan RMSE disajikan pada Tabel 12.

Tabel 12. Nilai RMSE Setiap Dokumen Hasil Peramalan

Kelas Dokumen	Nilai <i>root mean square error</i> (RMSE)
Positif	8,958
Negatif	14,866
Netral	11,451

Hasil yang diperoleh menunjukkan besaran error yang dihasilkan dari proses peramalan. Semakin kecil RMSE menunjukkan bahwa akurasi peramalan semakin baik. Berdasarkan Tabel 12 ditunjukkan bahwa akurasi peramalan pada kelas positif paling tinggi diantara ketiga kelas karena nilai RMSE yang dihasilkan paling kecil sebesar 8,958, sementara peramalan pada kelas negatif memperoleh nilai RMSE paling besar sebesar 14,866. Hal ini menunjukkan akurasi peramalan kelas negatif lebih rendah dibandingkan dengan ketiga kelas.

4. Kesimpulan

Berdasarkan pemodelan MNBC diperoleh tiga klasifikasi yaitu kelas positif, kelas negatif dan kelas netral. Hasil klasifikasi yang diperoleh sebanyak 775 dokumen diklasifikasikan sebagai opini negatif, 475 dokumen diklasifikasikan sebagai opini netral, dan 581 dokumen diklasifikasikan sebagai opini positif. Hal ini menunjukkan bahwa opini negatif lebih banyak dibandingkan opini positif dan netral.

Hasil peramalan tiga kelas klasifikasi selama periode satu minggu sejumlah 44 opini kelas positif, 25 opini kelas negatif dan 21 opini kelas netral. Hal ini menunjukkan bahwa bertambahnya jumlah opini positif selama periode satu minggu kedepan akan lebih banyak dibandingkan dengan opini negatif dan opini netral.

Model MNBC memperoleh *accuracy* sebesar 92% secara keseluruhan kelas. Hal ini menunjukkan bahwa tingkat akurasi model MNBC dalam melakukan klasifikasi pada penelitian ini dapat dikatakan sangat baik. Model NF memperoleh nilai RMSE yang telah diperoleh sebesar 8,958 untuk peramalan kelas positif, 14,866 untuk peramalan kelas negatif, dan 11,450. Hal ini menunjukkan peramalan dengan metode NF dikatakan cukup baik.

Adapun saran yang dapat dikembangkan berdasarkan penelitian ini yaitu pelabelan data dilakukan dengan lebih baik serta dilakukan pelabelan ulang agar didapatkan data berkualitas tinggi dan jangkauan waktu pengambilan data diperpanjang sehingga jangka waktu untuk peramalan lebih panjang dan pola data runtun waktu terdeteksi jelas.

Daftar Pustaka

- [1] F. C. Farisa, "Harga BBM Naik, Jokowi: 70 Persen Subsidi Justru Dinikmati Kelompok Mampu Pemilik Mobil," *Kompas*. Accessed: Jun. 14, 2024. [Online]. Available: <https://nasional.kompas.com/read/2022/09/03/14561231/harga-bbm-naik-jokowi-70-persen-subsidi-justeru-dinikmati-kelompok-mampu>
- [2] F. Romaito Pohan, C. Wafa Al-Farizi, and S. Tinggi Ilmu Administrasi Karya Dharma Merauke, "Dampak Implementasi Kebijakan Kenaikan Harga Bahan Bakar Minyak (BBM) Pada Sektor Transportasi," *Jurnal Administrasi Karya Dharma*, vol. 2, no. 2, 2023.
- [3] A. Lutfi, A. Annisa Fitriani, I. Ramadani, N. Azahra Putri, and Y. Shizuka Nelsi, "Efektivitas Penggunaan Aplikasi My Pertamina Di Era Kenaikan BBM Bersubsidi," *Prosiding Seminar Nasional Pendidikan, Bahasa, Sastra, Seni, dan Budaya (Mateandrau)*, vol. 1, no. 2, pp. 244–253, 2022.
- [4] Darmayuda, E. A. Pailis, A. F. Aulia, and M. Hasnan, "Analisis Pergeseran Pola Konsumsi Energi dan Dampak Ekonominya Pada Saat COVID-19 di Provinsi Riau," *Jurnal Ekonomi dan Bisnis*, vol. 10, no. 1, pp. 203–216, 2022.
- [5] A. Pribadi, "Siaran Pers : Pengalihan Subsidi BBM, Demi Subsidi Tepat Sasaran dan Antisipasi Gejolak Minyak Dunia," Kementerian Energi Dan Sumber Daya Mineral Direktorat Jenderal Minyak Dan Gas Bumi.
- [6] S. Kemp, "Twitter Users, Stats, Data & Trends," Data Reportal.
- [7] S. Nurhaliza, Y. Yusra, and M. Fikry, "Klasifikasi Sentimen Masyarakat di Twitter Terhadap Kenaikan Harga BBM dengan Metode Support Vector Machine," *Jurnal*

- Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, p. 586, Jul. 2023, doi: 10.30865/json.v4i4.6322.
- [8] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [9] R. Savira, A. Solichin, and M. Syafrullah, “Analisis Sentimen Pada Twitter Terhadap Kenaikan BBM 2022 Dengan Lexicon dan Support Vector Machine,” *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 2, no. 1, 2023.
- [10] M. Fikry, L. Oktavia, T. Dwi Arista, U. Islam Negeri Sultan Syarif Kasim Riau, J. H. Soebrantas NoKm, and S. Baru, “Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga BBM dengan Metode K-NN,” *JUKI: Jurnal Komputer dan Informatika*, vol. 5, no. 1, 2023.
- [11] C. Heltroyce, G. Feoh, and I. Made Dwi Ardiada, “Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Kesehatan, Sains, dan Teknologi (JAKASAKTI)*, vol. 3, no. 1, 2024, [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jakasakti/index>
- [12] A. J. Arifin and A. Nugroho, “Uji Akurasi Penggunaan Metode KNN dalam Analisis Sentimen Kenaikan Harga BBM pada Media Twitter,” *Progresif: Jurnal Ilmiah Komputer*, vol. 19, no. 2, pp. 700–708, 2023.
- [13] S. Lina, M. Sitio, and R. Nadiyah, “Analisis Sentimen Kenaikan Harga BBM Pertamina Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier,” *Technology and Science (BITS)*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2331.
- [14] S. Mujahidin, B. Prasetyo, and M. C. C. Utomo, “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, pp. 17–24, 2022, [Online]. Available: <http://ejournal.unp.ac.id/index.php/voteknika/index>
- [15] W. Maria Sonia Nadiyah Hutagalung and N. Jaya Perdana, “Analisis Sentimen Pada Opini Kenaikan Harga Bahan Bakar Minyak Pada Media Sosial Twitter,” *Jurnal Sistem Informasi dan Teknik Komputer*, vol. 8, no. 2, 2023.
- [16] P. Samotana Zalukhu, T. Handhayani, and M. Sitorus, “Analisis Sentimen Terhadap Kenaikan BBM di Indonesia Pada Media Sosial Twitter Menggunakan Metode Naive Bayes,” *JURNAL SISTEM INFORMASI DAN TEKNIK KOMPUTER*, vol. 8, no. 1, 2023.
- [17] Ardiansyah and Nur’aini, “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Dengan Metode Naive Bayes,” *Journal Of Information Systems And Informatics Engineering*, vol. 8, no. 1, pp. 1–9, 2024, doi: 10.35145/joisiej.v8i1.3838.
- [18] A. Wahyu *et al.*, “Penerapan Metode Adaptive Boosting Pada Analisis Sentimen Kenaikan BBM Pertamina,” *Jurnal Teknik Informatika Unika ST. Thomas (JTIUST)*, vol. 8, no. 2, pp. 2657–1501, 2023.
- [19] M. L. Pratama, Y. V. Via, and E. P. Mandyartha, “Analisis Performansi Naive Bayes dan Random Forest Terhadap Sentimen Kenaikan Harga BBM di Indonesia,” *SCAN: Jurnal Teknologi Informasi dan Komunikasi*, vol. 18, no. 1, 2023.

- [20] N. Basidit, E. Supriyadi, and A. Susilo, “Perbandingan Algoritma Klasifikasi dalam Analisis Sentimen Opini Masyarakat tentang Kenaikan Harga BBM”, doi: 10.31331/joined.v3i1.kodeartikel.
- [21] T. Sanubari, C. Prianto, and N. Riza, *Odol (one desa one product unggulan online) penerapan metode Naive Bayes pada pengembangan aplikasi e-commerce menggunakan Codeigniter*. Bandung: Kreatif, 2020.
- [22] S. Mujahidin, M. N. Hasyim, and B. M. Pratama, “Implementasi Analisis Sentimen Opini Publik Mengenai Sirkuit Internasional Mandalika Pada Twitter Menggunakan Metode Multinomial Naïve Bayes Classifier,” *Bianglala Informatika*, vol. 10, no. 2, 2022, doi: 10.31294/bi.v10i2.13544.
- [23] S. Mandasari, B. H. Hayadi, and R. Gunawan, “Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier,” *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, vol. 5, no. 2, 2022, doi: 10.53513/jsk.v5i2.5635.
- [24] G. M. Paldino, J. De Stefani, F. De Caro, and G. Bontempi, “Does AutoML Outperform Naive Forecasting? †,” *Engineering Proceedings*, vol. 5, no. 1, 2021, doi: 10.3390/engproc2021005036.
- [25] F. Qeadan *et al.*, “Naive forecast for COVID-19 in utah based on the South Korea and Italy models-the fluctuation between two extremes,” *Int J Environ Res Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082750.
- [26] A. P. Kirana and A. Bhawiyuga, “Coronavirus (COVID-19) Pandemic in Indonesia: Cases Overview and Daily Data Time Series using Naïve Forecast Method,” *Indonesian Journal of electronics, electromedical engineering, and medical informatics*, vol. 3, no. 1, 2021, doi: 10.35882/ijeemi.v3i1.1.
- [27] A. P. Kirana and A. Bhawiyuga, “Coronavirus (COVID-19) Pandemic in Indonesia: Cases Overview and Daily Data Time Series using Naïve Forecast Method,” *Indonesian Journal of electronics, electromedical engineering, and medical informatics*, vol. 3, no. 1, pp. 1–8, Feb. 2021, doi: 10.35882/ijeemi.v3i1.1.
- [28] Y. Fauziah, B. Yuwono, and A. S. Aribowo, “Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review,” *RSF Conference Series: Engineering and Technology*, vol. 1, no. 1, pp. 363–367, Dec. 2021, doi: 10.31098/cset.v1i1.397.
- [29] H. Han, Y. Zhang, J. Zhang, J. Yang, and X. Zou, “Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias,” *PLoS One*, vol. 13, no. 8, p. e0202523, Aug. 2018, doi: 10.1371/journal.pone.0202523.
- [30] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction To Information Retrieval*. England: Cambridge University press, 2009.
- [31] A. Hussein, F. K. Ahmad, and S. S. Kamaruddin, “Cluster Analysis on Covid-19 Outbreak Sentiments from Twitter Data using K-means Algorithm,” *Journal of System and Management Sciences*, vol. 11, no. 4, 2021, doi: 10.33168/JSMS.2021.0409.
- [32] A. Kumila, B. Sholihah, E. Evizia, N. Safitri, and S. Fitri, “Perbandingan Metode Moving Average dan Metode Naïve Dalam Peramalan Data Kemiskinan,” *JTAM / Jurnal Teori dan Aplikasi Matematika*, vol. 3, no. 1, p. 65, Apr. 2019, doi: 10.31764/jtam.v3i1.764.
- [33] S. A. A. Kharis, I. Hadi, and K. A. Hasanah, “Multiclass Classification of Brain Cancer with Multiple Multiclass Artificial Bee Colony Feature Selection and

- Support Vector Machine,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1417/1/012015.
- [34] G. Mutanov, V. Karyukin, and Z. Mamykova, “Multi-class sentiment analysis of social media data with machine learning algorithms,” *Computers, Materials and Continua*, vol. 69, no. 1, 2021, doi: 10.32604/cmc.2021.017827.
- [35] D. E. Cahyani and D. E. Cahyani, “PENERAPAN MACHINE LEARNING UNTUK PREDIKSI PENYAKIT STROKE,” *Jurnal Kajian Matematika dan Aplikasinya VOLUME*, vol. 3, no. 1, 2022, doi: 10.17977/um055v3i1p15-22.
- [36] T. O. Hodson, “Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not,” *Geosci Model Dev*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.