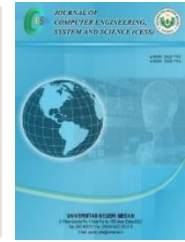


Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Analisis Sentimen Terhadap Tindakan Pemerintah Indonesia Untuk
Menampung Sementara Pengungsi Etnis Rohingya Menggunakan Naïve Bayes
Classifier**

***Sentiment Analysis of the Indonesian Government's Actions to Temporarily
Accommodate Rohingya Ethnic Refugees Using the Naïve Bayes Classifier***

Irwan Gunawan^{1*}, Mhd. Furqan²

^{1,2}Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara
Jl Lap Golf No 120, Kp Tengah, Kec Pancur Batu, Kab Deli Serdang, Provinsi Sumatera Utara, Indonesia
email: ¹irwan0701201002@uinsu.ac.id, ²mfurqan@uinsu.ac.id

ABSTRAK

Etnis Rohingya merupakan penduduk asli di negara myanmar yang sebagian besar mayoritasnya beragama muslim. Konflik yang terjadi pada etnis tersebut dimulai sejak ditetapkannya kebijakan *Burma Citizen Law* oleh pemerintah myanmar. kebijakan ini berisi terkait penolakan pemerintah myanmar terhadap etnis Rohingya sebagai etnis resmi dan memutuskan jika etnis tersebut tidak termasuk dari negara Myanmar. Indonesia merupakan salah satu negara di ASEAN yang masih menampung sementara pengungsi Rohingya, tindakan ini dilakukan berdasarkan konsep *Human Security* dan mengacu pada Peraturan Presiden Republik Indonesia Nomor 125 Tahun 2016 Tentang Penanganan Pengungsi Dari Luar Negeri Pasal 4 Ayat 2 mengenai koordinasi penanganan pengungsi yang meliputi Penemuan, Penampungan, Pengamanan dan Pengawasan. Akibatnya, terjadinya cemburu sosial yang berdampak pada keberagamannya opini masyarakat dan menjadi isu yang sering dibicarakan. Penelitian ini bertujuan untuk mengetahui kecenderungan opini berdasarkan klasifikasi sentimen yang diperoleh melalui video *YouTube*. Manfaat dari penelitian ini adalah agar pemerintah indonesia dapat mengetahui tindakan tersebut cenderung positif atau negatif. Dalam penelitian ini menerapkan algoritma *Naïve Bayes Classifier* dengan dataset berjumlah 7547 yang dibagi menjadi 6037 data latih dan 1510 data uji. Hasil *Confussion Matrix* pada penelitian ini menunjukkan akurasi 93%.

Kata Kunci: *Analisis Sentimen; Menampung; Pemerintah; Rohingya; Penambahan Teks; Youtube; Naive Bayes Clasifier*

ABSTRACT

The Rohingya ethnic group is a native population in the country of myanmar, where the majority are muslim. The conflict that occurred in this ethnic group began when the Burma

*Penulis Korespondensi:
email: irwan05072002@gmail.com

Citizen Law policy was implemented by the Myanmar government. This policy contains the Myanmar government's rejection of the Rohingya ethnicity as an official ethnicity and deciding that this ethnicity is not included in the state of Myanmar. Indonesia is one of the countries in ASEAN which is still temporarily hosting Rohingya refugees. This action is carried out based on the concept of Human Security and refers to Presidential Regulation of the Republic of Indonesia Number 125 of 2016 concerning Handling of Refugees from Abroad Article 4 Paragraph 2 concerning coordination of refugee handling which includes Discovery, Shelter, Security and Supervision. As a result, social jealousy occurs which has an impact on the diversity of public opinion and becomes an issue that is often discussed. To find out this, this research aims to find out opinion trends based on sentiment testimonials obtained through YouTube videos. The benefit of this research is that the Indonesian government can find out whether these actions tend to be positive or negative. In this research, the Naïve Bayes Classifier algorithm was applied with a dataset of 7547 which was divided into 6037 training data and 1510 test data. The results of the Confusion Matrix in this research show that the resulting accuracy level reached 93.

Keywords: *Sentiment Analysis; Accommodate; Government's; Rohingya; Text Mining; Youtube; Naive Bayes Classifier*

1. PENDAHULUAN

Etnis Rohingya merupakan penduduk asli di negara Myanmar yang sebagian besar mayoritasnya beragama Muslim, konflik yang terjadi pada etnis tersebut dimulai sejak ditetapkannya kebijakan *Burma Citizen Law* oleh pemerintah Myanmar pada tahun 1982 [1], kebijakan tersebut membahas terkait penolakan dari pemerintah Myanmar terhadap keberadaan Rohingya sebagai etnis resmi dan mengklaim bahwasanya etnis tersebut bukan bagian dari negara Myanmar. Indonesia termasuk salah satu negara di ASEAN (*Asean South East Nations*) yang menampung pengungsi Rohingya. Penyebaran etnis tersebut mencapai kurang lebih 1000 orang di Indonesia [2]. Selain itu, berdasarkan data dari lembaga UNHCR (*United Nations High Commissioner for Refugees*) bulan Maret tahun 2021 memperkirakan total pengungsi Rohingya di Indonesia mencapai 5% dari 13.497 keseluruhan pengungsi dari berbagai negara yang datang ke Indonesia [3].

Penyebab pemerintah Indonesia menampung sementara pengungsi Rohingya dikarenakan kemanusiaan. Menurut Tirza Listiarni, Indonesia memiliki kebijakan luar negeri yang terbuka terhadap pengungsi Rohingya. Dalam perencanaannya dilakukan dengan pendekatan *Human Security* bukan dengan *National Security* [4]. Selain itu, berdasarkan Peraturan Presiden Republik Indonesia Nomor 125 Tahun 2016 Tentang Penanganan Pengungsi Dari Luar Negeri yang terdapat pada Pasal 4 Ayat 2 mengenai koordinasi penanganan pengungsi yang meliputi Penemuan, Penampungan, Pengamanan dan Pengawasan. Pada pasal 26 Ayat 5 pemerintah juga memberikan bantuan kepada pengungsi Rohingya mencakup penyediaan air bersih, kebutuhan makanan, pakaian, pelayanan kesehatan dan kebersihan serta fasilitas ibadah.

Akibatnya, terjadi cemburu sosial yang berdampak pada beragamnya opini masyarakat yang menjadi isu yang sering di bicarakan di media sosial. Media sosial telah menjadi sarana komunikasi dan interaksi dalam kehidupan masyarakat dunia [5]. *YouTube* menjadi salah satu media sosial yang menyebarkan berita melalui konten video terkait tindakan pemerintah Indonesia yang menampung sementara pengungsi Rohingya. Menurut Hermawan, sejak tahun 2022 diperkirakan 139 juta masyarakat Indonesia menggunakan *YouTube* [6]. Maka dari itu,

YouTube menjadi media sosial yang sering digunakan oleh masyarakat Indonesia, serta menghasilkan data yang cukup banyak dan komentar pro dan kontra yang dapat di amati untuk di analisis. Maka, untuk mengetahui opini masyarakat terhadap tindakan pemerintah Indonesia untuk menampung sementara pengungsi etnis Rohingya, diperlukan Analisis Sentimen yang bertujuan untuk mengetahui kecenderungan opini berdasarkan klasifikasi sentimen melalui komentar video *youtube*. Analisis sentimen adalah pengolahan bahasa natural yang berperan menganalisis pendapat [7], dalam penerapannya analisis sentimen mengacu pada kemampuan perangkat lunak untuk memahami sikap, pendapat, atau pendirian pembicara atau penulis [8], analisis sentimen juga digunakan untuk menentukan sentimen tentang isu, peristiwa, kepuasan layanan, harga sahan dan analisis persaingan berdasarkan data teks [9].

Analisis sentimen yang dimaksud dalam penelitian ini akan menerapkan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan data teks komentar. Menurut Adelia, *Naïve Bayes Classifier* merupakan hasil probabilitas masa depan yang di ambil dari pengalaman masa sebelumnya [10]. Dan *Term Frequency-Inverse Document Frequency* untuk pembobotan kata (*term*). Menurut Mulyani TFIDF termasuk salah satu pembobotan kata terbaik dibandingkan dengan pembobotan kata lainnya [11].

Kajian terkait analisis sentimen menggunakan algoritma *Naïve Bayes Classifier* sebelumnya telah dilakukan oleh beberapa penelitian terdahulu. Seperti pada penelitian berjudul "Analisis Sentimen Tanggapan *Public* Mengenai E-Tilang Melalui Media Sosial *Youtube* Menggunakan Algoritma *Naïve Bayes*" yang dilakukan oleh [12]. Berdasarkan hasil penelitiannya opini yang dikumpulkan diambil dari media *Youtube*, karena media tersebut menjadi salah satu objek untuk menyampaikan opini karena bersifat *user friendly* dan terbuka dalam mengakses komentar. Namun, data yang digunakan selama penelitian berlangsung hanya berjumlah 500 komentar dan menghasilkan akurasi sebesar 79.44%. Selanjutnya, penelitian yang berjudul "Analisis Sentimen Terhadap Opini Feminisme Menggunakan Metode *Naïve Bayes*" yang dilakukan oleh [13], berdasarkan hasil penelitiannya digunakan teknik pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* dengan pola data 80:20 atau 80% data latih dan 20% data uji yang menghasilkan akurasi sebesar 88%. Selanjutnya, penelitian yang berjudul "Analisis Sentimen Masyarakat Terhadap Tindakan Vaksinasi Dalam Upaya Mengatasi Pandemic Covid-19" yang dilakukan oleh [14]. Berdasarkan hasil penelitiannya pemberian label dilakukan secara manual dan menghasilkan akurasi sebesar 85.59%. Penelitian analisis sentimen ini akan menggunakan data berjumlah 7.547 komentar yang diambil dari *platform youtube* dan pemberian label dilakukan secara manual, Lalu sentimen yang akan digunakan hanya Positif dan Negatif, serta pembobotan kata yang akan diterapkan menggunakan teknik *Term Frequency-Inverse Document Frequency* dengan pola 70% data *training* dan 30% data *testing*

2. DASAR/TINJAUAN TEORI

2.1. Etnis Rohingya

Etnis Rohingya merupakan salah satu etnis dari sejumlah etnis yang berada di negara Myanmar, penduduk Myanmar yang beragama Islam sebagian besar berasal dari etnis Rohingya dan etnis Rohingya termasuk etnis yang menganut agama Islam di negara tersebut [15]. Keberadaan etnis Rohingya di negara Myanmar merupakan minoritas Islam dan minoritas yang terisolasi di Myanmar, Rohingya adalah bangsa yang terbentuk selama peristiwa yang panjang. Mereka ada di wilayah Rakhine sebelum Inggris berkuasa di Rakhine dan Burma

antara tahun 1824 dan 1948 [16]. Pelanggaran dan tindakan kejahatan kemanusiaan yang dilakukan oleh pemerintah Myanmar terhadap etnis rohingya berawal dari di tetapkannya Undang-Undang status kewarganegaraan tahun 1962. Peraturan tersebut menghapus status kewarganegaraan kelompok rohingya sehingga sebagian dari mereka kehilangan hak-hak dan perlindungannya dari negara tersebut [17].

2.2. Analisis Sentimen

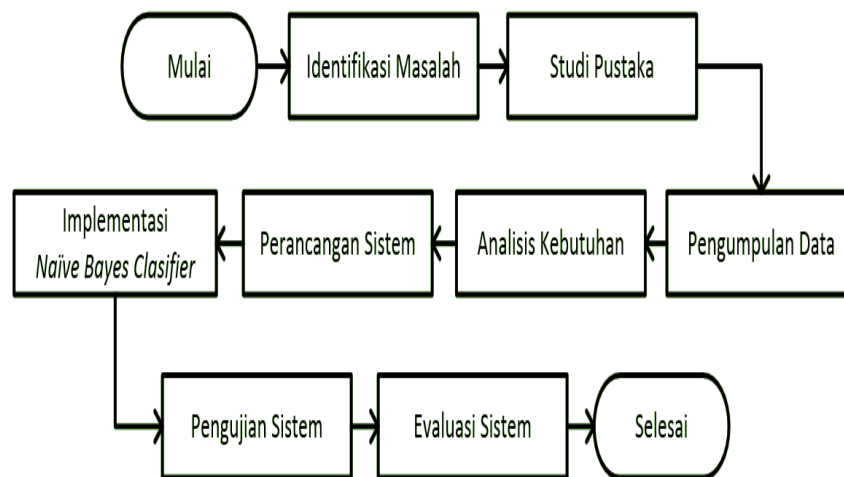
Analisis sentimen merupakan penambangan teks yang bertujuan untuk menganalisis dan mengekstraksi suatu informasi yang bersifat subjektif yang dapat membantu pihak yang membutuhkannya [19]. Analisis sentimen juga digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung di sebuah opini. Sentimen yang di kategorikan terbagi menjadi sentimen positif, netral dan negatif [20].

2.3. Text Mining

Text Mining merupakan proses penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, dan tidak diketahui sebelumnya atau menemukan kembali informasi-informasi yang tersirat secara implisit, temuan tersebut berasal dari informasi yang di ekstrak secara otomatis dari sumber-sumber data teks yang berbeda [20].

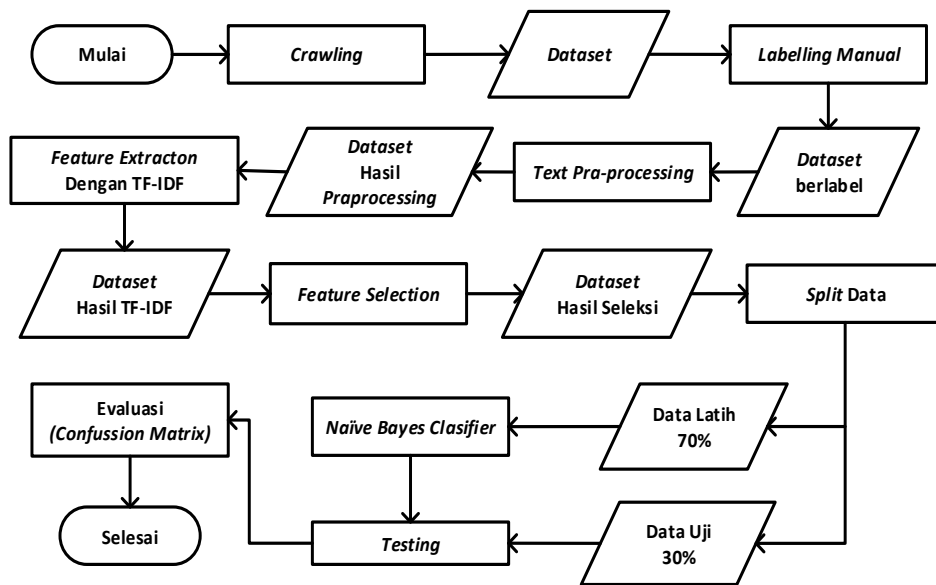
3. METODE

Metode penelitian yang digunakan dalam penelitian ini menggunakan metode Kualitatif, dan variabel yang diteliti adalah variabel *review*, karena data yang akan diteliti diperoleh berdasarkan hasil tanggapan masyarakat. Tahapan penelitian yang akan dilaksanakan disajikan dalam bentuk *flowchart* Gambar 1.



Gambar 1. Kerangka Penelitian

Perancangan model yang akan menggunakan *Naïve Bayes Classifier* juga disajikan dalam *flowchart* sistem Gambar 2.



Gambar 2. Flowchart Sistem Analisis Sentimen

3.1 Term Frequency Inverse Document Frequency

TFIDF adalah suatu pembobotan kata untuk menghitung kata umum pada *information retrieval*, pembobotan ini dikenal sebagai gabungan *term frequency (TF)* dan *inverse document frequency (IDF)*, yang mana *TF* merupakan total keberadaan kata pada dokumen. Sedangkan *IDF* adalah cara untuk mengetahui seberapa penting kata pada suatu dokumen [22]

Rumus perhitungan pembobotan *TF-IDF* dalam [23] disajikan dalam persamaan berikut :
 Persamaan untuk menghitung *TF (Term Frequency)*

$$TF = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}}$$

Keterangan persamaan:

TF = *Term Frequency*

d = dokumen

t = *term* (kata/token)

$f_{t,d}$ = jumlah *term* (kata/token) pada setiap dokumen

$\sum_{t \in d} f_{t,d}$ = jumlah dokumen yang memuat *t*

Persamaan untuk menghitung *IDF (Inverse Document Frequency)*

$$idf_t = \text{Log} \frac{N}{dft}$$

Keterangan persamaan:

idf_t = Nilai *idf* dari *term* (token) *t*

N = Banyaknya dokumen yang tersedia

dft = Intensitas kemunculan kata pada dokumen

3.2 Naïve Bayes Classifier

Naïve Bayes Classifier adalah algoritma klasifikasi yang menerapkan kemungkinan sederhana pada *Teorema Bayes* dan menerapkan aturan ketidaktergantungan (*independent*) pada setiap kondisi [20]. *Naïve Bayes Classifier* termasuk *Supervised Learning*, dikarenakan diperlukan data *training* dalam mengklasifikasikan data kedalam sebuah label [21]. *Naïve Bayes* sering diterapkan dalam memprediksi kemungkinan pada keanggotaan suatu kelas [24].

Rumus perhitungan algoritma *Naïve Bayes Classifier* dalam [25] dijabarkan dalam persamaan berikut :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan persamaan:

- X = Data dengan kelas yang belum diketahui
- H = Hipotesis data yang merupakan suatu kelas spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)
- P(H) = Probabilitas hipotesis H (probabilitas prior)
- P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) = Probabilitas

3.3 Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk menyatakan jumlah klasifikasi jumlah data uji [19]. *confusion matrix* yang dimaksud dalam [19] disajikan pada tabel 1.

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas sebenarnya	1	TP	FN
	0	FP	TN

Selanjutnya, berikut rumus *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, *f1 score* dan *recall* dalam [26] disajikan pada persamaan berikut :

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

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad F1\ Score = 2 \times \frac{precision \times recall}{precision+recall}$$

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Adapun cara yang digunakan dalam memperoleh data dilakukan dengan teknik *Crawling Youtube* menggunakan *secret_key* yang didapat pada layanan *Youtube Data API V3*. Dalam penelitian ini peneliti akan menganalisis data sentimen *youtube* terhadap tindakan pemerintah indonesia untuk menampung sementara pengungsi etnis rohingya. Dalam prosesnya digunakan bahasa pemrograman *python* dan *library googleapiclient* yang memungkinkan pengguna mencari informasi komentar di *platform youtube*. Hasil dari data yang dikumpulkan disimpan kedalam *file* berformat *.xls* dan data yang diambil pada *platform youtube* diperoleh sebanyak 7.547 data, hasil *crawling* dijabarkan dalam tabel 2.

Tabel 2. Hasil *Crawling* Data

No	Auth or	Updated _at	Like_coun t	Text	Video_id	Publi c
1	@y nien aro8 606	2024-02- 17T07:12 :58Z	0	mungkin masih mengkonsultasi dengan pemerintah Bangladesh akan dikembalikan ke coxs bazar di sana sudah dibangun megah penampungan	dgAwS5837N0	TRUE

2	@putukarme5228	2024-01-25T15:26:43Z	0	Kembalikan kemanusiaan bisa ditandang Rohingya minta ini minta tanah katanya	dgAwS5837N0	TRUE
...
7546	@user-tn7tn7wi1i	2023-12-11T09:41:04Z	0	Bangsa kurang ajar	K3BVsk-flk4	TRUE
7547	@wi7573	2023-12-11T09:39:09Z	0	Di kembalikan di laut	K3BVsk-flk4	TRUE

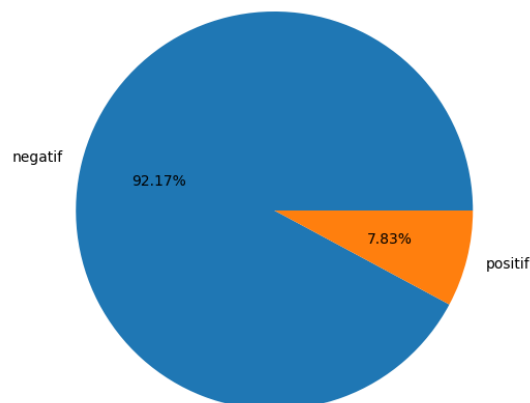
4.2. Pemberian Label

Pemberian label dilakukan secara manual, peneliti mulai membaca seluruh teks komentar. Hal ini dilakukan karena sebelumnya peneliti sudah mencoba memberikan label menggunakan *Lexicon Based* dan *Textblob*. Namun, cara tersebut tidak efektif karena sentimen yang diberikan tidak sesuai dengan isi dari komentarnya. Hasil dari pemberian label dengan cara manual Gambar 3.

#	author	updated_at	like_count	text	video_id	label
1	@yamenar0806	2024-02-17T07:12:5	0	0 mungkin masih mengkonsultasi dengan pemerintah Bangladesh akan dikembalikan ke cox bazar di sana sudah dibangun	dgAwS5837N0	TRUE positif
2	@putukarme5228	2024-01-25T15:26:4	0	0 Kembalikan kemanusiaan bisa ditandang Rohingya minta ini minta tanah katanya	dgAwS5837N0	TRUE negatif
3	@dianekaw3829	2024-01-23T09:45:1	0	0 ya pemerintah hanya menampung sementara tapi risik mereka pergi dari negara nya itu ingin menetap di indo pak	dgAwS5837N0	TRUE positif
4	@phonenor8525	2024-01-20T18:55:2	0	0 Pak Jokowi minta tolongRuhingya Banglades harusSebab mau pemilu rakyat Indonesia. Orang asing ruhingya banglades	dgAwS5837N0	TRUE positif
5	@Bellurina0961	2024-01-14T08:31:0	0	0 Kami bukan warga Aceh juga menolak keras pendatang gelap Rohingya.demi NKRI.	dgAwS5837N0	TRUE negatif
6	@user-nH9a7du3v	2024-01-22T09:16:2	0	0 Jangan percaya Jokowi dia biasa ngobrol yg sementara itu Jokowi km kurang dr 1 th dia sdh kehilangan kekuasaannya 1	dgAwS5837N0	TRUE negatif
7	@Hilrohmaniyah3247	2024-01-08T11:42:2	0	0 Tampung sementara, lalu negara mi selanjutnya yg mau menerimanya? sementara itu lama lg okkup 1, 2 th bs pulu	dgAwS5837N0	TRUE negatif
8	@User-uy1ss7bd8c	2024-01-07T07:50:0	0	0 Andai Jokowi suharto... Urusan pengungsi sudah selesai.	dgAwS5837N0	TRUE negatif
9	@SriMarti107	2024-01-04T02:50:4	0	0 Bpk Jokowi yang terhormat, silahkan bantu etnis Rohingya asal jang sampai tinggal di Indonesia karena Tanah air Ind	dgAwS5837N0	TRUE positif
10	@Usermyatyah4931	2024-01-03T17:08:0	0	0 Sementara bgnm yg sdh 9 th rohingya di Indonesia	dgAwS5837N0	TRUE negatif
11	@abbaq3443	2024-01-03T02:53:4	0	0 Teges...?ditampungnya berapa lama pak...?Kma faktanya, banyak yg di tampung sampai bertahun tahun disini.Indonesia	dgAwS5837N0	TRUE negatif
12	@maswaraja7829	2024-01-01T15:40:3	0	0 UHNCR itu penjajah...pk de Jokowi harus tegas...usir Rohingya.	dgAwS5837N0	TRUE negatif
13	@arifinharun0186	2023-12-31T18:21:2	0	0 Mana bukti ucapan le joko buktikan buktinya makin nambah	dgAwS5837N0	TRUE negatif
14	@User-ozak07pac2l	2023-12-31T04:45:0	0	0 Antar kesulo	dgAwS5837N0	TRUE positif
15	@riwayansugiana3843	2023-12-30T12:48:3	0	0 kenapa ditampung... thailand sj mampu mengusir Indonesia LEMBEX	dgAwS5837N0	TRUE negatif
16	@didikhariono8155	2023-12-28T00:14:0	0	0 Mnaaa tindakan tegasnya kok blm diangut2 ke negaranya	dgAwS5837N0	TRUE negatif
17	@jwanabahr4322	2023-12-27T11:14:3	0	0 Pak Jokowi kami meminta buatkan organisasi UHNCR	dgAwS5837N0	TRUE negatif
18	@Wahyudutrisno-hm3l	2023-12-27T01:15:5	0	0 Yh pak Jokowi usir pengungsi rohingya dari negara kita Indonesia	dgAwS5837N0	TRUE negatif
19	@User-ek8uBif9w	2023-12-26T15:53:1	0	0 Sementara gimana? Udah kelamaan dan kebanyakan	dgAwS5837N0	TRUE negatif
20	@Tubaguzrozk17470	2023-12-20T23:59:4	0	0 Sementara nya sampai kapan?1sthn/2sthn/3sthn....lebih cepat lebih baik usir imigran ilegal	dgAwS5837N0	TRUE negatif
21	@irekoki17417	2023-12-20T03:15:0	0	0 O TAMPUNG SEMENTERA MINGGA BERANAK PINAK DAN MUAJDI NEGARA KESATUAN ROHINGYA ROHINGYA Myanmar Myanmar	dgAwS5837N0	TRUE negatif
22	@rezenkoryuh42771	2023-12-24T00:47:5	0	0 Teges pak pulang aja...	dgAwS5837N0	TRUE negatif
23	@KirimBalikkeBangladesh-4	2023-12-23T10:44:4	0	0 Kirim Balik ke Bangladesh.	dgAwS5837N0	TRUE negatif
24	@User-vF2z9f4tj	2023-12-21T14:34:0	0	0 Secepatnya pak, di pulangkan kama warga Indonesia sudah muak dengan orang ruhingya.	dgAwS5837N0	TRUE negatif
25	@Pak-r2zn	2023-12-20T09:34:1	0	0 Di tampung sementara, kua " 10 th, seperti pengungsi yg tih dulu di Indonesia, jancok... jancok	dgAwS5837N0	TRUE negatif
26	@User-gedv8f01t	2023-12-19T22:52:5	0	0 Belajar dari sejarah, awalnya penjajahan.Adalah ketulusan, keluguan kesantunan bangsa kita menerima dan menghormi.	dgAwS5837N0	TRUE negatif
27	@Rayanaceh1210	2023-12-19T21:10:2	0	0 Secepat nya di pulangkan sebelum Indonesia hancur ,semoga Allah tunjukan jalan kebenaran ,kasihan rakyat sendiri ser	dgAwS5837N0	TRUE negatif
28	@User-rt8xw6gin	2023-12-19T17:04:3	0	0 Sementara nya sampai kapan ? ingat bapak tdi lama lagi sdh purna bakti jadi presiden. Pulangkan secepatnya lebih bagur	dgAwS5837N0	TRUE negatif

Gambar 3. Hasil Pemberian Labl

Dari pemberian label terhadap 7.547 data komentar baik diketahui 6956 (92.17%) sentimen Negatif dan 591 (7.83%) sentimen Positif, perbandingan label tersebut pada Gambar 4.



Gambar 4. Perbedaan Label Negatif dan Positif

4.3. Text Preprocessing

Setelah data diberikan label, tahap selanjutnya dilakukan pembersihan teks. Proses ini meliputi *Case Folding*, *Cleansing*, *Tokenizing*, *Word Normalization*, *Stopwords Removal*, dan *Stemming*. Tahap ini bertujuan agar data dapat digunakan tanpa adanya gangguan (*noise*) yang tidak diinginkan. Proses ini dapat mempengaruhi kinerja dari model klasifikasi yang akan dibuat. Adapun 6 tahapan pada *text preprocessing* tersebut dijabarkan sebagai berikut:

4.3.1 Case Folding

Komentar akan diubah menjadi huruf kecil (*lowercase*) dari yang sebelumnya berbentuk huruf kapital (*uppercase*). Tahap ini bertujuan untuk menyeragamkan teks agar tidak ada yang kapital ketika di proses. Hasilnya pada Tabel 3

Tabel 3. Hasil *Case Folding*

Text Asli	Case Folding
y sdh tampung semua demi kemanusiaan. indonesia kn sgt ² luas, kaya 😊😊😊	y sdh tampung semua demi kemanusiaan. indonesia kn sgt ² luas, kaya 😊😊😊

4.3.2 Cleansing

Komentar akan dicek apakah terdapat simbol seperti (:/\][!?!)(*&^%\$#@!), angka 1234567890, *link url* seperti *https* atau *www*, karakter khusus dan *mention* (@). Jika ada maka akan dihapus dan hanya menyisakan teks. Proses ini dilakukan agar teks yang dihasilkan bersih dari gangguan (*noise*). Hasilnya dalam Tabel 4.

Tabel 4. Hasil *Cleansing*

Hasil Case Folding	Cleansing
y sdh tampung semua demi kemanusiaan. indonesia kn sgt ² luas, kaya 😊😊😊	y sdh tampung semua demi kemanusiaan indonesia kn sgt luas kaya

4.3.3 Tokenizing

Komentar yang masih berbentuk kalimat, akan dipisahkan menjadi satuan kata, hal ini bertujuan untuk memperkecil pola pencarian kata pada tahap *stopwords* dan *stemming*. Hasilnya pada Tabel 5.

Tabel 5. Hasil *Tokenizing*

Hasil Cleansing	Tokenizing
y sdh tampung semua demi kemanusiaan indonesia kn sgt luas kaya	'y' 'sdh' 'tampung' 'semua' 'demi' 'kemanusiaan' 'indonesia' 'kn' 'sgt' 'luas' 'kaya'

4.3.4 Word Normalization

Komentar akan dinormalisasikan dari kata-kata yang memiliki singkatan, kata asing, dan kata gaul. kamus kata digunakan untuk membandingkan kata di *dataset* dengan 375 kamus kata singkatan dan kata normal. Hasilnya pada Tabel 6.

Tabel 6. Hasil *Word Normalization*

Hasil Tokenizing	Word Normalization
'y' 'sdh' 'tampung' 'semua' 'demi' 'kemanusiaan' 'indonesia' 'kn' 'sgt' 'luas' 'kaya'	'iya' 'sudah' 'tampung' 'semua' 'demi' 'kemanusiaan' 'indonesia' 'kan' 'sangat' 'luas' 'kaya'

4.3.5 Stopwords Remove

komentar akan dihilangkan dari kata yang tidak memiliki makna seperti kata penghubung dan keterangan, peneliti menggunakan *library nltk* untuk mengambil *stopwords* yang menyimpan 758 kata bahasa Indonesia. Hasilnya pada Tabel 7.

Tabel 7. Hasil *Stopwords Remove*

Hasil <i>Word Normalization</i>	<i>Stopwords Remove</i>
'iya' 'sudah' 'tampung' 'semua' 'demi'	'tampung' 'semua' 'demi'
'kemanusiaan' 'indonesia' 'kan 'sangat 'luas'	'kemanusiaan' 'indonesia' 'luas'
'kaya'	'kaya'

4.3.6 Stemming

Komentar akan dihilangkan dari imbuhan paa setiap kata, bertujuan agar kata yang akan dihasilkan nantinya merupakan kata yang baku yang tidak terindikasi imbuhan didalamnya. Peneliti menggunakan *library Sastrawi* untuk menghilangkan imbuhan secara otomatis. Hasilnya pada Tabel 8.

Tabel 8. Hasil *Stemming*

Hasil <i>Stopwords Remove</i>	<i>Stemming</i>
'tampung' 'semua' 'demi' 'kemanusiaan'	'tampung' 'semua' 'demi' 'manusia'
'indonesia' 'luas' 'kaya'	'indonesia' 'luas' 'kaya'

4.4. Klasifikasi Naïve Bayes Classifier

Dataset yang telah melewati tahap *text preprocessing* dan pembobotan TF-IDF dilanjutkan ke proses pembelajaran (*learning*) menggunakan *Naïve Bayes Classifier*. peneliti menggunakan *library sklearn*. Proses diawali dengan menginputkan data teks dan label yang di masukan kedalam x dan y. Selanjutnya, diberikan *text_size* 0.3 yang mengindikasikan data uji yang akan digunakan sebesar 30% dan dibuat *x_train* yang menyimpan teks data latih, *x_test* yang menyimpan teks data uji, *y_train* yang menyimpan label data latih, dan *y_test* yang menyimpan label data uji. Setelah itu, untuk data latih pada *x_train* dan *y_train* dimasukan kedalam *MultinomialNB* untuk dilakukan pembelajaran (*learning*) dan disimpan pada *model_naive_bayes*, terakhir dilakukan prediksi menggunakan data uji didalam *x_test* dan hasil pengujiannya disimpan didalam prediksi. Untuk proses ini Gambar 5.

```
x = seleksi
y = y
-----
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 1)
-----
model_naive_bayes = MultinomialNB().fit(x_train, y_train)
prediksi = model_naive_bayes.predict(x_test)
```

Gambar 5. Klasifikasi *Naïve Baye Clasifier*

4.5. Pengujian Model

Model yang telah dibuat akan diuji dengan tujuan agar peneliti dapat mengetahui persentase ketepatan dan keakuratan dalam mengklasifikasikan data. Berdasarkan 7.547 data, 70% data latih berjumlah 5.285 data, dan 30% data uji berjumlah 2.265 data. Nantinya, 2.265 data uji akan di uji apakah label yang di prediksi sesuai berdasarkan label aslinya.

```
TRAINING
Komentar Training      : 5282 data
Sentimen Training      : 5282 data
TESTING
Komentar Testing       : 2265 data
Sentimen Testing       : 2265 data
```

Gambar 6. Pembagian Data

Berdasarkan model analisis sentimen *Naïve Bayes Classifier*, diperoleh nilai akurasi sebesar 0.9395143487858719 atau 93% dari *confusion matrix*. Selain itu, dari total 2.265 data uji, 2101 dikategorikan sentimen Negatif dan 27 sentimen Positif. Namun, diketahui 137 yang seharusnya sentimen Positif justru diprediksi menjadi Negatif. Hasil prediksi pada Gambar 7.

```

Akurasi      : 0.9395143487858719
Precision    : 1.0
Recall       : 0.16463414634146342

Confusion Matrix
[[2101  0]
 [ 137  27]]

Classification Report
precision  recall  f1-score  support
0         0.94    1.00    0.97    2101
1         1.00    0.16    0.28     164

accuracy   0.94
macro avg  0.97  0.58  0.63  2265
weighted avg 0.94  0.94  0.92  2265
    
```

Gambar 7. Hasil *Confussion Matrix* klasifikasi

4.6. Evaluasi Model

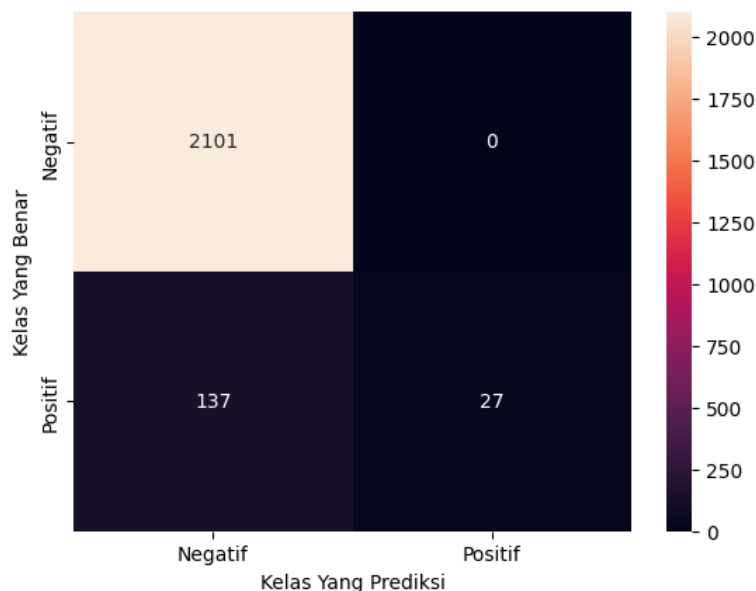
Peneliti ingin mengetahui performa dari model tersebut. Maka dilakukan perhitungan performa menggunakan *Confusion Matrix* 3x3 yang terdiri dari *Accuracy*, *Precision*, *Recall* dan *f1 Score*. Diketahui sebelumnya dari tahap pengujian model didapatkan hasil *confusion matrix* 3x3 pada Tabel 9.

Tabel 9. Hasil *Confusion Matrix*

		<i>False / Predicted Class</i>	
		Negatif	Positif
<i>True Class</i>	Negatif	2101	0
	Positif	137	27

Untuk mempermudah pembaca memahami hasil klasifikasi pada *Confusion Matrix* Peneliti juga memvisualisasikan nya kedalam Gambar 8.

Confusion Matrix - Analisis Sentimen Irwan Gunawan



Gambar 8. Visualisasi *Confusion Matrix*

Kemudian peneliti melakukan perhitungan untuk menghitung performa dari model yang telah dibuat mengacu dari tabel *confusion matrix* diatas. Sehingga, hasil perhitungan nilai akurasi secara manual disajikan pada persamaan berikut ini:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% & \text{Recall} &= \frac{27}{27+137} \times 100\% \\
 &= \frac{27+2101}{27+2102+0+137} \times 100\% & &= \frac{27}{164} \times 100\% \\
 &= \frac{2128}{2.265} \times 100\% & &= 0,164634146 \times 100\% \\
 &= 0,939514349 \times 100\% & &= 16\% \\
 &= 93\% & & \\
 \\
 \text{Precision} &= \frac{27}{27+0} \times 100\% & \text{f1 Score} &= 2 \times \frac{100 \times 16}{100+16} \\
 &= \frac{27}{27} \times 100\% & &= 2 \times \frac{1600}{116} \\
 &= 100\% & &= 27\%
 \end{aligned}$$

Berdasarkan perhitungan yang dilakukan, disimpulkan bahwasanya skor *accuracy* sebesar 93%, *precision* 100%, *recall* 16%, dan *f1 score* 27%. Dengan memperhatikan hal tersebut. Algoritma *Naïve Bayes Classifier* yang digunakan untuk mengembangkan model klasifikasi dalam analisis sentimen pada penelitian ini dapat dikategorikan baik dalam mengklasifikasikan.

5. KESIMPULAN

Berdasarkan penelitian ini diperoleh kesimpulan diantaranya *Naïve Bayes Classifier* dapat diterapkan dalam mengklasifikasikan data opini masyarakat terkait topik yang dibahas di *platform youtube*, diketahui sentimen negatif lebih unggul berjumlah 6956 data dibandingkan dengan sentimen negatif yang hanya berjumlah 591 data. Dengan begitu, masyarakat indonesia tidak setuju atau kurang baik memberikan pendapat terhadap tindakan pemerintah indonesia untuk menampung sementara pengungsi etnis rohingya. Selain itu, pada penelitian ini menunjukkan jika *Naïve Bayes Classifier* dapat dikategorikan baik dalam mengklasifikasikan karena mendapatkan skor *accuracy* sebesar 93%, *precision* 100%, *recall* 16%, dan *f1 score* 27%. dan dengan memperhatikan *Confusion Matrix* bahwasanya data uji yang berhasil diklasifikasikan kedalam sentimen negatif berjumlah 2101 data dan data uji yang berhasil diklasifikasikan kedalam sentimen positif berjumlah 27 data, serta data uji yang salah diklasifikasikan berjumlah 137 data.

Adapun saran dari peneliti diantaranya, penambahan dataset agar bisa lebih banyak seperti jutaan atau miliaran data serta diambil dari berbagai *platform* seperti instagram, tiktok, facebook dan sebagainya untuk memperkaya dataset. Selain itu, disarankan untuk menambahkan kamus kata pada tahapan *word normalization*, baik itu kamus bahasa asing, gaul, serta singkatan

UCAPAN TERIMA KASIH

Terima kasih kepada Allah Swt atas karunianya sehingga penulis mendapatkan berkah dalam menyelesaikan penelitian ini, kepada Bapak Dr. Mhd Furqan, S.Si, S.H M.Comp S.c selaku Dosen Pembimbing, kepada Bapak Ilka Zufria, M.Kom selaku Dosen Penguji, kepada Orang Tua Penulis Bapak Sriadi S.H dan Ibunda Sriani, dan kepada Raisa Azhari selaku sahabat yang terus kebersamai, memberikan support selama penulis melaksanakan penelitian ini.

REFERENSI

- [1] N. T. Pramita and M. S. Zahidi, "Analisis Peranan Indonesia Dalam Pemberian Bantuan Kemanusiaan Terhadap Etnis Rohingya Pada Era Jokowi," *Rev. Int. Relations*, vol. 5, no. 1, pp. 84–106, 2023.
- [2] D. R. Rameliah and U. W. Sagena, "Krisis Keamanan Manusia Di Myanmar Dan Penerapan Konsep Responsibility To Protect," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 27979–27986, 2023.
- [3] F. Adha, "Penanganan Pengungsi Di Indonesia Melalui Kerja Sama Kemenkumham Dengan International Organizatio Of Migration," *Popul. J. Sosia dan Hum.*, vol. 8, no. 2, pp. 123–133, 2023.
- [4] T. Listiarani, "Analisis Kebijakan Luar Negeri Indonesia Dalam Menerima Pengungsi Rohingya Di Indonesia," *J. PIR Power Int. Relations*, vol. 5, no. 1, pp. 19–32, 2020.
- [5] N. Meilani, M. Furqan, and Suhardi, "Analisis Sentimen Pengguna Aplikasi BSI Mobile Akibat Ransomware Menggunakan Algoritma Support Vector Machine," *Infotech J. Inform. Teknol.*, vol. 5, no. 1, pp. 42–51, 2024.
- [6] F. Hermawan, "Analisis Minat Masyarakat Pengguna Platform Youtube Sebagai Media Komunikasi Digital Masa Kini," *J. Manaj.*, vol. 14, no. 3, pp. 564–573, 2022.
- [7] M. Furqan, Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 52–61, 2022.
- [8] M. Furqan and A. F. A. Nasir, "Big Data Approach To Sentiment Analysi In Machine Learning-based Microblogs: Perspectives Of Religious Moderation Public Policy In Indonesia," *J. Appl. Eng. Technol. Sci.*, vol. 5, no. 2, 2024.
- [9] H. Setiawan and I. Zufria, "Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1003–1012, 2023.
- [10] S. Adelia, E. Milanda, J. Santari, D. T. Kesuma, E. Silvia, and F. Kurniawan, "Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes," *J. Inf. Technol. Ampera*, vol. 4, no. 3, 2023.
- [11] S. Mulyani and R. Novita, "Implementasi Algoritma Naive Bayes Clasifier Untuk Klasifikasi Sentimen Masyarakat Tentang Depresi Pada Youtube," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1355–1361, 2022.
- [12] Z. N. Aulia, G. K. Jati, and I. Santoso, "Analisis Sentimen Tanggapan Public Mengenai E-Tilang Melalui Media Sosial Youtube Menggunaka Algoritma Naive Bayes," *J. Ikraith Inform.*, vol. 7, no. 2, 2023.
- [13] W. Wahyuni, "Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes," *J. Inform. Ekon. Bisnis*, vol. 4, no. 4, pp. 148–153, 2022.
- [14] B. Laurensz and E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021.
- [15] R. K. Indradipradana and F. Haridha, "Kebijakan Luar Negeri Indonesia: Studi Kasus Penerimaan Pengungsi Rohingya Asal Myanmar Tahun 2020-2022," *Indones. Perspect.*, vol. 8, no. 2, pp. 211–236, 2023.
- [16] M. Hamdi, H. Maulidia, and H. Firlana, "Fenomena Pencari Suaka dan Pengungsi Etnis Rohingya Di Indonesia (Studi Kasus Penangan Rohingya Pada Provinsi Aceh)," *J. Ilm. Kaji*.

Keimigrasian, vol. 6, no. 1, pp. 55–71, 2023.

- [17] A. Mahadevi, F. Arindah, and J. M. Sandya, “Implementasi Hak Asasi Manusia Internasional Dalam Pemenuhan Asas Membership Oleh Myanmar Kepada Etnis Rohingya,” *J. Polit. Democr. Stud.*, vol. 3, no. 2, pp. 142–157, 2022.
- [18] J. R. Prabowo, Akim, and A. Sudirman, “Peran Tentara Nasional Indonesia Dalam Menanggulangi Pengungsi Dari Luar Negeri: Studi Kasus Pengungsi Rohingya Di Indonesia (2015-2020),” *Aliansi J. Polit. Keamanan dan Hubunga Int.*, vol. 1, no. 2, pp. 99–111, 2022.
- [19] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [20] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [21] A. A. A. Sumanjaya, Indriati, and A. Ridok, “Analisis Sentimen Data Tweets Terhadap Penanganan Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Dan Pemilihan Kata Bersentimen Menggunakan Lexicon Based,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1865–1872, 2022.
- [22] M. R. Fahlevvi, “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine,” *J. Teknol. dan Komun.*, vol. 4, no. 1, pp. 1–13, 2022.
- [23] H. Santoso, A. Armansyah, and D. Desliani, “Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naive Bayes Classifier,” *Techno.Com*, vol. 21, no. 3, pp. 644–654, 2022.
- [24] Supiyandi, A. Hussein, I. Gunawan, and W. L. R. Harjo, “Analisis Klasifikasi Broken Home pada Anak Menggunakan Metode Naïve Bayes Classifier,” *JISKA J. Inform. Sunan Kalijaga*, vol. 8, no. 2, pp. 90–101, 2023.
- [25] N. P. G. Naraswati, D. C. Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, and R. Nooraeni, “Analisis Sentimen Publik Dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia Dengan Naive Bayes Clasifiation,” *J. Sist. Inf.*, vol. 10, no. 1, pp. 222–238, 2021.
- [26] Hertati, E. Haerani, Novriyanto, and F. Syafira, “Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naive Bayes Clasifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1578–1584, 2023.