



Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes

Harry Setiawan*, Ilka Zufria

Fakultas Sains dan Teknologi, Ilmu Komputer, Universitas Islam negeri Sumatera Utara, Medan, Indonesia

Email: *harisetiawan.0712@gmail.com, ilkazufria@uinsu.ac.id

Email Penulis Korespondensi: harisetiawan.0712@gmail.com

Abstrak—Piala Dunia FIFA U-20 merupakan ajang bergengsi pesepakbola muda di seluruh dunia. Indonesia sedianya menjadi tuan rumah pada tahun 2023, namun FIFA akhirnya harus membatalkan penyelenggaraan Piala Dunia di Indonesia karena beberapa tokoh masyarakat Indonesia tidak terima akan kehadiran Timnas Israel di Indonesia pada ajang pertandingan sepak bola tersebut. Penolakan tersebut dilakukan dengan alasan keamanan, karena keikutsertaan Timnas Israel dinilai berpotensi mengancam keamanan Indonesia, khususnya mengingat konflik Palestina. Alasan lainnya, Indonesia tidak memiliki hubungan diplomatik dengan Israel, keputusan FIFA cukup menimbulkan berbagai macam pendapat baik bersifat positif maupun negatif. Tujuan penelitian ini berfungsi agar masyarakat dapat bersikap tabayyun dengan memberikan solusi dengan melakukan analisis sentimen yaitu mengumpulkan opini pengguna Twitter secara otomatis sehingga dapat berguna untuk masyarakat. Data yang diperoleh sebanyak 946 tweet setelah melalui tahap preprocessing kemudian data dapat digunakan pada tahap pelabelan dengan menggunakan metode Lexicon Based. Pengelompokan dibagi menjadi sentimen positif dan sentimen negatif yang diklasifikasi dengan algoritma Naïve Bayes yang diperkuat dengan pembobotan Lexicon Based menghasilkan sentimen positif sebanyak 150 tweet dengan persentase sebesar 15,86% dan sentimen negatif sebanyak 796 tweet dengan persentase sebesar 84,14%. Dari hasil confusion matrix diperoleh performa klasifikasi dengan algoritma Naïve Bayes yang diperkuat dengan pembobotan Lexicon Based kemudian menghasilkan persentase accuracy sebesar 84%, dengan precision sebesar 86%, recall sebesar 95% dan nilai f-measure 90%.

Kata Kunci: Analisis sentimen; Confusion Matrix; Lexicon Based; Naïve Bayes; Piala Dunia U-20; Twitter

Abstract—The FIFA U-20 World Cup is a prestigious event for young footballers around the world. Indonesia was originally to host the event in 2023, but FIFA eventually had to cancel the World Cup in Indonesia because some Indonesian public figures did not accept the presence of the Israeli national team in Indonesia at the football match. The refusal was made on security grounds, because the participation of the Israeli national team was considered a potential threat to Indonesia's security, especially considering the Palestinian conflict. Another reason, Indonesia does not have diplomatic relations with Israel, FIFA's decision is enough to cause a variety of opinions both positive and negative. The purpose of this research is to function so that people can be tabayyun by providing solutions by conducting sentiment analysis, namely collecting Twitter user opinions automatically so that they can be useful for the community. The data obtained is 946 tweets after going through the preprocessing stage then the data can be used in the labelling stage using the Lexicon Based method. Grouping is divided into positive sentiment and negative sentiment which is classified by the Naïve Bayes algorithm reinforced by Lexicon Based weighting resulting in positive sentiment as many as 150 tweets with a percentage of 15.86% and negative sentiment as many as 796 tweets with a percentage of 84.14%. From the confusion matrix results, the classification performance with the Naïve Bayes algorithm reinforced with Lexicon Based weighting then produces an accuracy percentage of 84%, with a precision of 86%, a recall of 95% and an f-measure value of 90%.

Keywords: Sentiment analysis; Confusion Matrix; Lexicon Based; Naïve Bayes; World Cup U-20; Twitter

1. PENDAHULUAN

Indonesia dipilih FIFA untuk menjadi tuan rumah Piala Dunia pesepakbola muda pada usia 20 tahun, hal tersebut memunculkan berbagai macam respon dari masyarakat mengenai kesiapan Indonesia sebagai tuan rumah pada ajang bergengsi tingkat dunia ini. Banyak masyarakat yang merasa bangga dan bahagia sebab ini adalah kesempatan bagi Indonesia untuk membuktikan diri sebagai Negara yang cukup mampu untuk menyelenggarakan acara olahraga internasional dan ini adalah waktu yang tepat untuk kampanye mengenai keanekaragaman budaya serta keindahan yang dimiliki oleh Indonesia. Namun tidak sedikit juga kalangan yang mengkhawatirkan fasilitas dan infrastruktur yang ada di Indonesia, terutama stadion-stadion yang digunakan untuk menggelar pertandingan belum sepenuhnya memadai dan memenuhi standar internasional. Selain itu, masalah keamanan menjadi perhatian, mengingat Indonesia dikenal sebagai Negara dengan tingkat keamanan yang masih perlu ditingkatkan [1].

Piala Dunia FIFA U-20 merupakan ajang bergengsi bagi pesepakbola muda di seluruh dunia. Indonesia sedianya menjadi tuan rumah pada tahun 2023, namun FIFA akhirnya harus membatalkan penyelenggaraan Piala Dunia di Indonesia karena alasan infrastruktur yang belum siap, keputusan FIFA terhadap pembatalan status tuan rumah ini menjadi topik yang hangat dibicarakan. Banyak persiapan untuk menjadi tuan rumah yang dilakukan semenjak 2019 untuk perhelatan sepak bola terbesar di dunia kemudian berakhir sia-sia. Pembatalan yang dilakukan FIFA disebabkan karena adanya beberapa tokoh masyarakat Indonesia tidak terima dengan hadirnya Timnas Israel untuk bertanding pada Piala Dunia FIFA U-20 di Indonesia. Penolakan tersebut dilakukan dengan alasan keamanan, karena keikutsertaan Timnas Israel dinilai berpotensi mengancam keamanan Indonesia, khususnya mengingat konflik Palestina. Alasan lainnya, karena Indonesia tidak saling hubungan dengan Israel secara diplomatik, sehingga penerimaan terhadap Timnas Israel tidak sejalan dengan makna UUD 1945.



Keputusan FIFA cukup menimbulkan berbagai macam pendapat baik bersifat positif maupun negatif seperti masyarakat yang kecewa dengan keputusan ini, sebaliknya masyarakat memberi respon positif sebab Indonesia terlepas dari inkonsistensi amanat UUD 1945. Oleh karena itu penulis melakukan penelitian yang diharapkan hasil dari penelitian ini berfungsi agar masyarakat bersikap tabayyun dengan memberikan solusi dengan melakukan analisis sentimen yaitu mengumpulkan opini pengguna Twitter secara otomatis sehingga dapat berguna untuk masyarakat.

Keputusan FIFA tentang pembatalan status kemudian menjadi topik yang menarik untuk dibahas, berbagai tanggapan dari opini yang bersifat positif maupun negatif selalu ada di media sosial terkhususnya Twitter. Twitter adalah alat komunikasi yang memungkinkan publik untuk mengekspresikan diri dan lebih diminati di komunitas global. Media sosial ini memungkinkan penggunaannya untuk membagi informasi dan mengungkapkan keinginan dan pendapat kepada pengguna lainnya secara real time terutama mengenai topik atau masalah yang sedang menjadi fokus pembahasan, informasi tersebut disebut sebagai kicauan (tweet) [2]. Twitter mengalami peningkatan pengguna secara pesat dan meraih popularitas di seluruh dunia sebab saat ini Twitter memiliki jumlah pengguna sekitar 78 juta di seluruh dunia [3].

Penelitian ini menerapkan metode Lexicon Based sebagai ekstraksi fitur dan Naïve Bayes sebagai metode untuk proses klasifikasi. Lexicon Based adalah kamus sekumpulan kata-kata yang biasa digunakan untuk mengungkapkan perasaan seperti positif atau negatif berdasarkan nilai dari setiap kata [4]. Dalam pengaplikasiannya, Lexicon Based memiliki dua kamus yang digunakan sebagai daftar kata. Kamus dengan kumpulan kata yang berhubungan dengan emosi positif dan kamus dengan kumpulan kata yang berhubungan dengan emosi negatif [5]. Naïve Bayes merupakan algoritma klasifikasi dengan menggunakan bentuk statistik dan probabilistik yang simpel berdasarkan teorema Bayes yang diberi asumsi yaitu setiap atributnya bersifat independen. Jadi, algoritma Naïve Bayes mengasumsikan atribut tertentu dari ada atau tidaknya kesamaan teks dengan ciri-ciri kelas lainnya [6].

Terkait analisis sentimen terdapat beberapa penelitian terdahulu seperti yang dilakukan oleh Dainamang dkk pada penelitian tersebut memakai metode Naïve Bayes mendapatkan hasil akurasi sebesar 85% pada pengujian dengan rentang split data 90% - 100% [7]. Pada penelitian Ilmawan dan Munde yang dimana membandingkan metode Naïve Bayes dan Support Vector Machine pada analisis sentimen komentar review pengguna aplikasi di Google Playstore yang hasilnya SVM Classifier memiliki nilai akurasi 81.46% sementara Naïve Bayes sebesar 75,41% [8]. Hal yang serupa ditemukan pada penelitian oleh Hasibuan dan Serdano pada analisis sentimen kebijakan pembelajaran tatap muka pasca pandemi Covid-19 dengan data indeks sebanyak 152 buah data dengan pembobotan TF-IDF memberikan hasil akurasi Support Vector Machine sebesar 88,09 sementara Naïve Bayes sebesar 75,92%. Adapun penelitian lainnya oleh Hudha dkk yang menghasilkan tingkat akurasi metode Naïve Bayes sebesar 90,36% pada analisis sentimen pengguna Youtube terhadap tayangan #MataNajwaMenantiTerawan [9].

Adapun beberapa penelitian terdahulu terkait penggunaan metode Lexicon Based seperti yang dilakukan oleh Dyo Fatra dkk pada topik perpindahan ibukota Indonesia, pada penelitiannya menggunakan metode Lexicon Based menghasilkan akurasi sebesar 77% dengan nilai rata-rata recall sebesar 75% dan f-measure sebesar 75% [10]. Kemudian penelitian oleh Najib dkk membandingkan metode Lexicon Based dan Support Vector Machine dengan topik kampanye pilpres 2019 di Twitter dengan dengan tingkat akurasi yang diperoleh berdasarkan metode metode SVM sebesar 83% dan Lexicon-based adalah 39% [4]. Selanjutnya penelitian oleh Mahendrajaya dkk menggunakan Lexicon Based sebagai metode untuk pelabelan opini dalam analisis sentimen pengguna Gopay yang berjumlah 1210 data dan metode klasifikasi menggunakan Support Vector Machine yang menghasilkan persentase accuracy sebesar 89,16% [11].

Pada penelitian ini penulis akan menganalisis sentimen respon pengguna Twitter terhadap pembatalan status tuan rumah piala dunia FIFA U-20 di Indonesia yang dimana penelitian ini akan mengelompokkan opini masyarakat kedalam 2 sentimen yaitu sentimen positif dan negatif. Penelitian ini akan menggunakan Naïve Bayes sebagai metode klasifikasi sentimen dan didukung dengan metode pendekatan Lexicon untuk mengoptimalkan hasil yang diperoleh berupa Confusion Matrix yaitu accuracy, recall dan f-measure dari analisis sentimen menggunakan Naïve Bayes tersebut.

2. METODOLOGI PENELITIAN

2.1 Penambangan Teks

Penambangan teks (text mining) merupakan metode yang efektif dalam memanfaatkan data teks yang tidak tersusun dengan cara menganalisisnya untuk menghasilkan suatu informasi terbaru kemudian mengidentifikasi model yang signifikan dan hubungan tersembunyi yang ada dalam data teks terkait [12]. Text mining adalah proses memeriksa dan menganalisis kumpulan big data dalam bentuk tekstual untuk memperoleh informasi yang berguna untuk suatu tujuan tertentu [13]. Cara kerja text mining pada dasarnya banyak meminjam dari proses data mining, perbedaan yang terdapat pada text mining menggunakan model dari bahasa normal tidak terstruktur, sedangkan data mining, diambil dari database terstruktur.



2.2 Analisis Sentimen

Analisis sentimen (sentiment analysis) adalah tahap untuk memperoleh informasi tentang suatu objek yang secara otomatis mengenali subjektivitas dari objek tersebut. Yang bertujuan untuk mengetahui apakah data teks yang dibuat oleh pengguna dapat digolongkan kepada opini positif, opini negatif atau opini yang cenderung netral. [14]. Analisis sentimen ialah studi komputasi untuk menganalisis perasaan dan opini orang terhadap suatu entitas. Bidang analisis sentimen telah menjadi topik penelitian yang banyak dilakukan dalam beberapa dekade terakhir [15]. Dari uraian sebelumnya dapat ditarik kesimpulan bahwa analisis sentimen merupakan metode untuk memperoleh informasi berupa pemikiran seseorang tentang suatu peristiwa. Analisis sentimen dapat digunakan untuk menentukan sentimen umum tentang isu, peristiwa, kepuasan layanan, harga saham, dan analisis persaingan berdasarkan data teks.

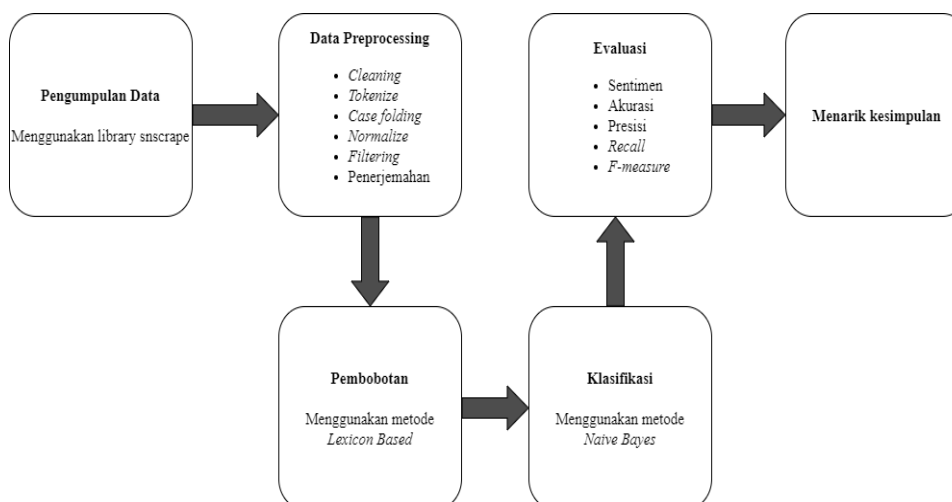
2.3 Piala Dunia U-20

Untuk yang pertama kali dalam sejarah perhelatan olah raga di Indonesia, FIFA memberi kesempatan untuk menjadi tuan rumah Piala Dunia U-20. Dengan predikat sebagai tuan rumah, Indonesia mendapat keistimewaan khusus dimana tim nasional Indonesia berhak menjadi peserta Piala Dunia U-20 2023. Bagi masyarakat Indonesia hal ini tentu menjadi kebanggaan tersendiri sebab terakhir kali Indonesia menjadi peserta Piala Dunia U-20 yaitu di Jepang pada tahun 1979. Untuk menjadi tuan rumah, Indonesia harus melalui beberapa tahapan seleksi diantaranya seleksi proposal perombakan stadion yang nantinya akan digunakan sebagai tempat penyelenggaraan pertandingan. Untuk Piala Dunia U20, Indonesia menyiapkan 10 stadion antara lain Stadion Utama Gelora Bung Karno yang terletak di Jakarta, Stadion Manahan Solo, Stadion Pakansari di Bogor, Stadion Patriot di Bekasi, Stadion Mandala Krida di Yogyakarta, Stadion I Wayan Dipta yang berada di Bali, Stadion Jalak Harupat di Bandung, Stadion Wibawa Mukti yang terletak di Cikarang, Stadion Jakabaring di kota Palembang dan Stadion Gelora Bung Tomo di Surabaya [16].

Setelah ditetapkan oleh FIFA sebagai tuan rumah, Indonesia harus berkomitmen dalam persiapan infrastruktur dan keamanan untuk menjadi tuan rumah Piala Dunia U-20 2023. Kewajiban tersebut tertulis dalam peraturan FIFA, misalnya kelayakan stadion dan kesesuaian dengan persyaratan dan sebagainya [17]. Piala Dunia U20 2023 dijadwalkan berlangsung mulai 29 Mei hingga 11 Juni dan akan diikuti 20 negara. Namun FIFA memutuskan untuk membatalkan status Indonesia sebagai tuan rumah karena infrastruktur yang belum siap dan konsistensi amanat UUD 1945. Pembatalan status tuan rumah Piala Dunia bisa disebabkan oleh faktor keamanan, kesehatan, kegagalan persiapan hingga alasan diluar kendali [18]. Indonesia bukan satu-satunya negara yang pernah mengalami pembatalan status sebagai tuan rumah, pada tahun 2003 FIFA membatalkan Irak sebagai tuan rumah Piala Dunia U20 sebab adanya konflik perang Irak dan pada tahun 1995 hal serupa dialami oleh Nigeria yang dibatalkan status tuan rumah karena wabah meningitis yang terjadi di negara tersebut ini adalah kali kedua negara tersebut digagalkan sebab pada tahun 1991 Nigeria dicabut status tuan rumahnya dikarenakan pemalsuan umur pemain.

2.4 Tahapan Penelitian

Penelitian ini akan menganalisis sentimen pengguna Twitter terhadap pembatalan Indonesia yang menjadi tuan rumah pada Piala Dunia U-20 yang dimana penelitian ini akan menjelaskan runtutan proses seperti pengumpulan data berupa teks tweet pengguna Twitter, kemudian mengolah data teks dan membaginya ke dalam sentimen bernilai positif dan negatif kemudian dalam penelitian ini juga akan membahas tentang penerapan algoritma Naïve Bayes dengan metode pendekatan Lexicon dalam pengklasifikasian. Tahapan penelitian diilustrasikan pada gambar 1 sebagai berikut:



Gambar 1. Ilustrasi Tahapan Penelitian



2.9 Confusion Matrix

Data yang telah diklasifikasi selanjutnya digunakan untuk pengujian model yaitu menghitung nilai accuracy, precision, recall dan f-measure berdasarkan Confusion Matrix. Confusion Matrix diperlukan untuk mengevaluasi seberapa baik kinerja sistem dalam proses klasifikasi data [11].

Tabel 1. Tabel Confusion Matrix

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) Hasil yang benar	FP (False Positive) Hasil yang tidak diharapkan
	FALSE	FN (False Negative)	TN (True Negative)

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Metode yang digunakan untuk proses pengumpulan data yaitu dengan cara Crawling data Twitter menggunakan kata kunci “FIFA batal U-20”. Crawling merupakan proses untuk mendata dan mengambil konten dari internet, lalu hasil dari proses tersebut disimpan kedalam database [19]. Dalam penelitian ini penulis akan menganalisis data sentimen Twitter terhadap pembatalan Indonesia sebagai tuan rumah pada kompetisi sepak bola dunia untuk usia 20 tahun dengan melakukan pengumpulan data. Proses pengumpulan data opini pengguna Twitter menggunakan metode crawling yaitu memanfaatkan fasilitas library snsrape dari bahasa pemrograman Python. Snsrape adalah library untuk mengambil data dari jejaring sosial dalam penelitian ini adalah Twitter dengan snsrape memungkinkan pengguna untuk mencari informasi spesifik dan terperinci di media sosial tanpa harus dibatasi dengan API resmi dari platform media sosial. Untuk dapat menggunakan library snsrape pengguna cukup menginstall library tersebut, hasil dari data yang dikumpulkan disimpan kedalam file dengan format .csv. Data yang akan diambil pada media sosial Twitter menggunakan kata kunci “fifa batal u-20” dengan rentang waktu data dari tanggal 25 Maret 2023 sampai dengan 30 Maret 2023 dan memperoleh hasil sebanyak 946 data. Hasil dari proses data crawling dapat dilihat pada tabel 2 berikut :

Tabel 2. Hasil Crawling Data

No.	Username	Tweet
1.	2karakterbisa	Batal jadi tuan rumah Piala Dunia U-20, apa yg bisa didapat Indonesia? Bahkan hikmah pun rasanya tak bisa didapat. Sanksi FIFA justru menanti. Israel tetep main, palestina ga otomatis merdeka, Indonesia kehilangan segalanya.
2.	stephvalentina_	Gue gak sebel pildun u-20 batal di Indo, gue sebel karena alasan batalnya adalah sentimen anti-Israel, sedangkan kemaren pas tragedi Kanjuruhan ratusan orang tewas, gak ada akuntabilitasnya tp masih aja dilanjutin. FIFA really need to fix their priority.
3.	Bsgoww	@Fajarilm @idextratime lahhh udah jelas dong batal jadi host karena ada penolakan Tim Israel main di Pildun u-20, paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir Jadi bukan batal karena tragedi Kanjuruhan

3.2 Preprocessing Data

Data preprocessing adalah tahap pertama dari pemrosesan text mining, yang tujuannya adalah untuk mengubah data menjadi sesuatu yang lebih bermakna [20]. Tahap preprocessing merupakan tahap awal dari peringkasan teks, tahap ini memiliki beberapa tahap derivasi yaitu pemecahan kalimat, pelipatan huruf, tokenisasi, dan penghilangan stopword [21]. Pada tahap penelitian ini data yang menggunakan data berupa data teks yang sebelumnya sudah diambil melalui proses crawling dan akan melakukan proses preprocessing data sebab data yang diambil masih mempunyai bentuk teks yang tidak terstruktur dengan banyak noise didalamnya, sehingga data terkait harus



dibersihkan terlebih dahulu. Berikut ini adalah langkah-langkah dalam melakukan preprocessing data yaitu cleaning, tokenize, case folding, normalize, filtering dan penerjemahan sebab pada penelitian ini penulis menggunakan library yang disediakan oleh bahasa pemrograman python seperti vader sentiment yang dimana library ini menggunakan kamus berbahasa Inggris.

- a. Cleaning, yaitu proses menghapus karakter yang tidak diperlukan pada data tweet seperti numeric, tanda baca, username, url, mention, hastag dan retweet [22]. Berikut adalah tahapan yang diterapkan dalam proses cleaning:

1. Menghapus simbol, angka, dan karakter selain huruf seperti `/?&%(/_+~*123"-`.
2. Menghilangkan nama pengguna (username) yang diawali dengan mention '@' seperti @Fajarilm.
3. Menghapus retweet dan url seperti `www`, dan `https://`. Retweet yang diberi tanda "RT".
4. Dan membuang bagian tidak penting lainnya yang akan menimbulkan noise pada data.

Pada tabel 3 berikut merupakan contoh data hasil tahap cleaning :

Tabel 3. Hasil Cleaning

Tweet	Hasil cleaning
@Fajarilm @idextratime lahhh udah jelas dong batal jadi host karena ada penolakan Tim Israel main di Pildun u-20, paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir. Jadi bukan batal karena tragedi Kanjuruhan	lahhh udah jelas dong batal jadi host karena ada penolakan Tim Israel main di Pildun u paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir Jadi bukan batal karena tragedi Kanjuruhan

- b. Tokenize, tahap ini dilakukan untuk memisahkan kalimat menjadi bentuk kata per kata agar memperoleh kata yang bernilai. Berikut adalah tahapan yang diterapkan dalam proses tokenizing:

1. Memindai setiap baris data teks dengan bentuk satuan tweet.
2. Pembagian isi dokumen menjadi kata demi kata satuan yang disebut token.
3. Menyimpan semua informasi kata tunggal.

Pada tabel 4 berikut merupakan contoh data hasil tahap tokenize :

Tabel 4. Hasil Tokenize

Hasil cleaning	Hasil tokenize
lahhh udah jelas dong batal jadi host karena ada penolakan Tim Israel main di Pildun u paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir Jadi bukan batal karena tragedi Kanjuruhan	['lahhh', 'udah', 'jelas', 'dong', 'batal', 'jadi', 'host', 'karena', 'ada', 'penolakan', 'Tim', 'Israel', 'main', 'di', 'Pildun', 'u', 'paragraf', 'selanjutnya', 'menuliskan', 'fifa', 'tidak', 'akan', 'berhenti', 'memberikan', 'asistensi', 'terkait', 'kejadian', 'oktober', 'lalu', 'dibawah', 'kepemimpinan', 'erick', 'thohir', 'Jadi', 'bukan', 'batal', 'karena', 'tragedi', 'Kanjuruhan']

- c. Case folding, yaitu langkah yang mengubah huruf besar ke dalam huruf kecil (lowercase) pada data. Berikut adalah tahapan yang diterapkan dalam proses Case folding :

1. Membaca setiap baris data tweet yang terdapat huruf kapital.
2. Mengubah seluruh huruf besar menjadi huruf kecil.

Pada tabel 5 berikut merupakan contoh data hasil tahap case folding :

Tabel 5. Hasil Case folding

Hasil tokenize	Hasil Case folding
['lahhh', 'udah', 'jelas', 'dong', 'batal', 'jadi', 'host', 'karena', 'ada', 'penolakan', 'Tim', 'Israel', 'main', 'di', 'Pildun', 'u', 'paragraf', 'selanjutnya', 'menuliskan', 'fifa', 'tidak', 'akan', 'berhenti', 'memberikan', 'asistensi', 'terkait', 'kejadian', 'oktober', 'lalu', 'dibawah', 'kepemimpinan', 'erick', 'thohir', 'Jadi', 'bukan', 'batal', 'karena', 'tragedi', 'Kanjuruhan']	lahhh udah jelas dong batal jadi host karena ada penolakan tim israel main di pildun u paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir jadi bukan batal karena tragedi kanjuruhan

- d. Normalize, langkah ini berfungsi untuk mengubah kata-kata yang disingkat menjadi kata-kata yang bermakna berdasarkan isi dari Kamus Besar Bahasa Indonesia (KBBI) sehingga informasi yang diperoleh dapat diproses dengan mudah [23]. Misalnya kata 'utk' menjadi 'untuk' dan 'udh' menjadi 'udah'. Berikut adalah tahapan yang diterapkan dalam proses Normalize :

1. Mencari kata yang disingkat pada data.
2. Data kata yang disingkat yang ditemukan maka akan diganti dengan kata normal yang memiliki makna berdasarkan isi KBBI.

Pada tabel 6 berikut merupakan contoh data hasil tahap normalize :



Tabel 6. Hasil Normalize

Hasil Case folding	Hasil Normalize
lahhh udah jelas dong batal jadi host karena ada penolakan tim israel main di pildun u paragraf selanjutnya menuliskan fifa tidak akan berhenti memberikan asistensi terkait kejadian oktober lalu dibawah kepemimpinan erick thohir jadi bukan batal karena tragedi kanjuruhan	['lahhh', 'sudah', 'jelas', 'dong', 'batal', 'jadi', 'host', 'karena', 'ada', 'penolakan', 'tim', 'israel', 'main', 'di', 'piala', 'dunia', 'kamu', 'paragraf', 'selanjutnya', 'menuliskan', 'fifa', 'tidak', 'akan', 'berhenti', 'memberikan', 'asistensi', 'terkait', 'kejadian', 'oktober', 'lalu', 'dibawah', 'kepemimpinan', 'erick', 'thohir', 'jadi', 'bukan', 'batal', 'karena', 'tragedi', 'kanjuruhan']

e. Filtering, atau stopword removal berfungsi untuk menghapus kata yang tidak memiliki nilai pada data seperti 'yang', 'dan', 'dari', 'di' dan 'untuk' [24]. Berikut adalah tahapan yang diterapkan dalam proses Filtering :

1. Membaca token kata-kata yang tersedia pada daftar stopwords.
2. Mencari setiap kata dan mencocokkan dengan kata yang terdapat pada daftar stopwords.
3. Jika kata yang cocok ditemukan dengan daftar stopwords maka kata tersebut akan dihapus.

Pada tabel 7 berikut merupakan contoh data hasil tahap filtering :

Tabel 7. Hasil Filtering

Hasil Normalize	Hasil Filtering
['lahhh', 'sudah', 'jelas', 'dong', 'batal', 'jadi', 'host', 'karena', 'ada', 'penolakan', 'tim', 'israel', 'main', 'di', 'piala', 'dunia', 'kamu', 'paragraf', 'selanjutnya', 'menuliskan', 'fifa', 'tidak', 'akan', 'berhenti', 'memberikan', 'asistensi', 'terkait', 'kejadian', 'oktober', 'lalu', 'dibawah', 'kepemimpinan', 'erick', 'thohir', 'jadi', 'bukan', 'batal', 'karena', 'tragedi', 'kanjuruhan']	lahhh batal host penolakan tim israel main paragraf menuliskan berhenti asistensi terkait kejadian oktober dibawah kepemimpinan erick thohir batal tragedi kanjuruhan

f. Penerjemahan, tahap ini berfungsi sebab pada penelitian ini penulis menggunakan library vader sentiment yang dimana library ini menggunakan kamus berbahasa Inggris. Berikut adalah tahapan yang diterapkan dalam proses penerjemahan :

1. Mencari kata berbahasa Indonesia.
2. Data kata yang ditemukan kemudian diterjemahkan kedalam bahasa Inggris menggunakan library googletrans.
3. Menyimpan data tweet berbahasa Inggris.

Pada tabel 8 berikut merupakan contoh data hasil tahap penerjemahan :

Tabel 8. Hasil Penerjemahan

Hasil Filtering	Hasil Penerjemahan
lahhh batal host penolakan tim israel main paragraf menuliskan berhenti asistensi terkait kejadian oktober dibawah kepemimpinan erick thohir batal tragedi kanjuruhan	lahhh cancel hosting the israel team's refusal to play paragraph writing stop assistance regarding the october incident under the leadership of erick thohir cancel kanjuruhan tragedy

3.3 Pembobotan dengan Lexicon Based

Penelitian ini juga menggunakan metode Lexicon Based sebagai media pelabelan atau pembobotan agar dapat memaksimalkan hasil yang diperoleh. Dalam pengaplikasiannya, Lexicon Based memiliki dua kamus yang digunakan sebagai daftar kata. Kamus dengan kumpulan kata yang berhubungan dengan emosi positif dan kamus dengan kumpulan kata yang berhubungan dengan emosi negatif [5]. Ketika proses sortir dilakukan pada satu bagian data, total bobot data menghasilkan nilai positif, negatif, atau netral . Setelah menyortir data, maka akan memberikan bobot dari data tersebut [10]. Perhitungan pada metode Lexicon dapat dilihat pada persamaan 5 berikut :

$$\text{Sentence}_{\text{sentiment}} = \begin{cases} \text{positive if } S_{\text{positive}} > S_{\text{negative}} \\ \text{neutral if } S_{\text{positive}} = S_{\text{negative}} \\ \text{negative if } S_{\text{positive}} < S_{\text{negative}} \end{cases} \quad (5)$$

Setelah melalui tahap pembobotan akan menghasilkan data berupa sentimen positif dan negatif dengan ketentuan jika nilai polaritas lebih besar atau sama dengan 0,1 maka akan diberi label sentiment positif dan jika nilai polaritas lebih kecil atau sama dengan 0,0 maka akan diberi label sentimen negatif. Hasil dari pelabelan dengan metode Lexicon akan digunakan sebagai data pendukung untuk klasifikasi dengan Naïve Bayes. Contoh hasil pembobotan dengan Lexicon Based dapat dilihat pada tabel 9 berikut :

**Tabel 9.** Hasil Pembobotan dengan metode Lexicon Based

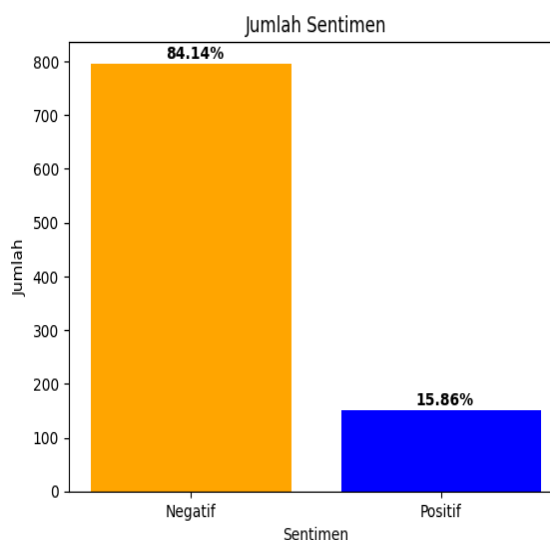
Tweet	Nilai polaritas	Sentimen
lahhh batal host penolakan tim israel main paragraf menuliskan berhenti asistensi terkait kejadian oktober dibawah kepemimpinan erick thohir batal tragedi kanjuruhan	-0.802	Negatif
sebel batal indo sebel alasan batalnya sentimen anti israel kemarin pas tragedi kanjuruhan ratusan tewas akuntabilitasnya dilanjutin really need to fix their priority	-0.9201	Negatif
ganjar indonesia bertanggungjawab oke let see pertanggungjawaban batal nolak doang solusi bangsa terancam rugi iya minimal lobi biar indonesia main amp gk diban	0.7167	Positif
indonesia batal tuan rumah ketum erick thohir berupaya maksimal melobi menerima keputusannya	0.7845	Positif

3.4 Klasifikasi dengan Naïve Bayes

Naïve Bayes adalah metode klasifikasi dengan model statistik dan probabilistik yang sederhana berdasarkan teorema Bayes yang diberi asumsi yaitu setiap atributnya bersifat independen. Jadi, algoritma Naïve Bayes mengasumsikan ada atau tidak suatu atribut tertentu yang tidak memiliki kesamaan dengan ciri-ciri kelas lainnya [6]. Naïve Bayes bermanfaat proses klasifikasi dan menentukan estimasi dengan hanya menggunakan sedikit data latih (training data) [25]. Naïve Bayes sering bekerja jauh lebih baik bahkan dalam situasi dunia nyata yang paling sulit sekalipun. Tujuan dari pengklasifikasi studi Naïve Bayes adalah untuk mengklasifikasikan hasil komentar online tentang penerapan teknologi yang telah melewati proses analisis opini [26]. Peneliti menggunakan metode Naïve Bayes untuk mengklasifikasi opini pengguna Twitter mengenai pembatalan status Indonesia sebagai tuan rumah Piala Dunia U-20 dan didukung dengan metode pendekatan Lexicon untuk mengoptimalkan hasil yang diperoleh berupa confusion matrix yaitu accuracy, precision, recall dan f-measure dari analisis Naïve Bayes tersebut.

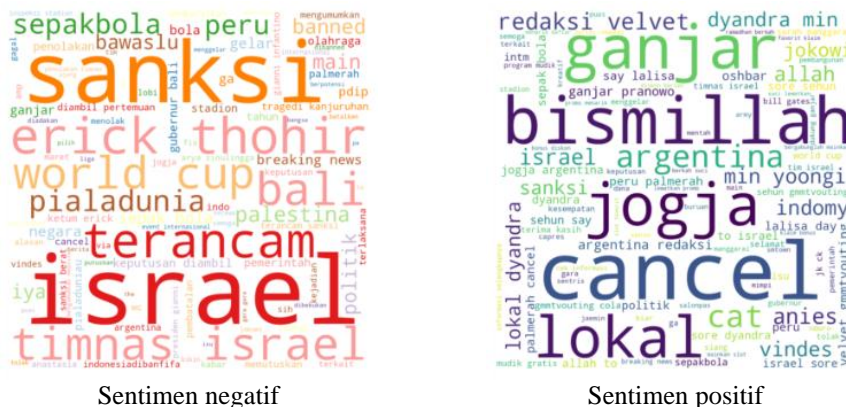
a. Sentimen

Hasil yang diperoleh dengan metode Naïve Bayes dan didukung dengan metode pendekatan Lexicon Based menghasilkan kelas sentimen negatif berjumlah 796 tweet atau 84,14% dan sentimen positif berjumlah 150 tweet atau 15,86% dari total data sebanyak 946 tweet, yang diilustrasikan pada gambar 2 berikut :

**Gambar 2.** Grafik jumlah Sentimen

b. Visualisasi wordcloud

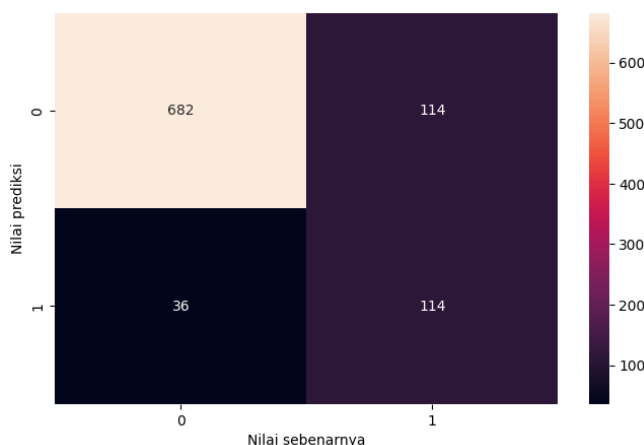
Hasil dari sentimen negatif dan positif tentang pembatalan Indonesia sebagai tuan rumah Piala Dunia U-20 belum diketahui kata yang terkandung dalam kelas sentiment negatif dan positif. Oleh karena itu diperlukan wordcloud yang berfungsi untuk membagikan kalimat opini yang termasuk ke dalam kelas sentimen positif atau negatif. Visualisasi data kata yang terdapat dalam kelas sentimen positif atau negatif. Visualisasi tersebut dapat dilihat pada gambar 3 berikut :



Gambar 3. Hasil Wordcloud

3.5 Evaluasi

Hasil pengujian akurasi menggunakan metode Naïve Bayes divisualisasikan dalam bentuk confusion matrix, yang dimana bentuk ini memberikan perbandingan hasil klasifikasi Naïve Bayes yang dibantu oleh metode pembobotan Lexicon Based. Model confusion matrix dalam bentuk matrix ukuran 2x2 yang menunjukkan 682 tweet yang bernilai True Positive (TP) dan 114 tweet yang bernilai True Negative (TN). Visualisasi model confusion matrix digambarkan pada gambar 4 berikut:



Gambar 4. Visualisasi Confusion Matrix

Hasil dari confusion matrix, True Positive (TP) disebut sebagai prediksi kelas positif yang benar sedangkan False Positive (FP) merupakan nilai yang salah. Untuk True Negative (TN) merupakan kelas negatif yang diprediksi bernilai benar sementara False Negative (FN) untuk kelas yang salah. Hasil dari kelas tersebut dapat dilihat pada tabel 10 berikut :

Tabel 10. Hasil Confusion Matrix

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	682	114
	FALSE	36	114

Nilai kelas pada confusion matrix kemudian dihitung secara manual untuk memperoleh persentase nilai accuracy, precision, recall dan f-measure.

a. Perhitungan Accuracy :

$$\text{Accuracy} = \frac{(682 + 114)}{(682 + 114 + 36 + 114)} \times 100\% = \frac{796}{946} \times \frac{100}{100} = 84 \%$$

b. Perhitungan Precision :

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% = \frac{682}{(682 + 114)} \times 100\% = \frac{682}{796} \times \frac{100}{100} = 86 \%$$

c. Perhitungan Recall



$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{682}{(682 + 36)} \times 100\% = \frac{682}{718} \times \frac{100}{100} = 95 \%$$

d. Perhitungan F-measure

$$F - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% = \frac{2 \times 86 \times 95}{86 + 95} = \frac{16.340}{181} \times \frac{100}{100} = 90 \%$$

Berdasarkan hasil perhitungan confusion matrix dapat dilihat performa klasifikasi dengan metode Naïve Bayes yang diperkuat dengan pembobotan Lexicon Based menghasilkan tingkat akurasi sebesar 84% dari total 946 tweet, dengan persentase sebesar 86% untuk untuk precision, 95% untuk nilai recall, dan 90% untuk nilai f-measure. Hasil perhitungan manual sesuai dengan output perhitungan pada bahasa pemrograman Python yang dapat dilihat pada gambar 5 berikut :

	precision	recall	f1-score	support
Negatif	0.86	0.95	0.90	718
Positif	0.76	0.50	0.60	228
accuracy			0.84	946

Gambar 5. Output oleh Python

4. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen respon pengguna Twitter terhadap pembatalan Indonesia sebagai tuan rumah Piala Dunia U-20 menggunakan data hasil crawling dengan kata kunci 'fifa batal u-20' diperoleh data sebanyak 946 tweet kemudian data tersebut melalui tahap preprocessing yang dimana tahapan didalamnya meliputi cleaning, tokenize, case folding, normalize, filtering dan penerjemahan agar data yang diperoleh dapat digunakan pada tahap pelabelan dengan metode Lexicon Based. Selanjutnya data yang telah dilabeli akan diklasifikasi dengan memakai metode Naïve Bayes. Hasil sentimen diperoleh berupa 150 tweet positif dengan persentase sentimen sebesar 15,86%, dan 796 tweet negatif dengan persentase sentimen sebesar 84,14%. Persepsi negatif memiliki jumlah lebih banyak daripada positif, hal ini disebabkan karena sebagian besar tweet menunjukkan kekecewaan terhadap keputusan FIFA tentang pembatalan Indonesia sebagai tuan rumah Piala Dunia U-20, ketakutan masyarakat terhadap sanksi FIFA dan beberapa ungkapan kekecewaan kepada tokoh masyarakat. Dari hasil confusion matrix diperoleh performa klasifikasi dengan metode Naïve Bayes yang diperkuat dengan pembobotan Lexicon Based yang menghasilkan persentase akurasi sebesar 84% dari total 946 tweet, dengan persentase sebesar 86% untuk untuk precision, 95% untuk nilai recall, dan 90% untuk nilai f-measure. Untuk penelitian selanjutnya diharapkan pengembangan penelitian seperti menerapkan algoritma yang berbeda atau memaksimalkan kinerja pembobotan dengan metode pembobotan lainnya agar hasil yang diperoleh lebih akurat dalam menganalisis sentimen dan diharapkan agar menggunakan media sosial selain Twitter seperti Facebook, Instagram dan Youtube.

REFERENCES

- [1] M. S. Alif and A. R. Pratama, "Analisis Kesadaran Keamanan di Kalangan Pengguna E-Wallet di Indonesia," J. Inf., vol. 2, no. 1, pp. 1–7, 2021, [Online]. Available: <https://journal.uin.ac.id/AUTOMATA/article/view/17279>
- [2] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," J. Pilar Nusa Mandiri, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [3] C. Darmawan, H. Silvana, H. N. Zaenudin, and R. Effendi, "Pengembangan hubungan interpersonal remaja dalam penggunaan media sosial di Kota Bandung," J. Kaji. Komun., vol. 7, no. 2, p. 159, 2019, doi: 10.24198/jkk.v7i2.21163.
- [4] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter," Fountain Informatics J., vol. 4, no. 2, p. 41, 2019, doi: 10.21111/fij.v4i2.3573.
- [5] P. A. Sumitro, Rasiban, D. I. Mulyana, and W. Saputro, "Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based," J-ICOM - J. Inform. dan Teknol. Komput., vol. 2, no. 2, pp. 50–56, 2021, doi: 10.33059/j-icom.v2i2.4009.
- [6] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," Smatika J., vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [7] S. A. Dainamang, N. Hayatin, and D. R. Chandranegara, "Analisis Sentimen Media Sosial Twitter terhadap RUU Omnibus Law dengan Metode Naïve Bayes dan Particle Swarm Optimization," Komputika J. Sist. Komput., vol. 11, no. 2, pp. 211–218, 2022, doi: 10.34010/komputika.v11i2.6037.
- [8] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," Ilk. J. Ilm., vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [9] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," JIKO (Jurnal Inform. dan Komputer), vol. 5, no.



- 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
- [10] A. H. Dyo fatra, N. H. Hayatin, and C. S. K. Aditya, “Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia,” *J. Repos.*, vol. 2, no. 7, p. 977, 2020, doi: 10.22219/repositor.v2i7.937.
- [11] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, “Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine,” *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [12] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, “Text mining in big data analytics,” *Big Data Cogn. Comput.*, vol. 4, no. 1, pp. 1–34, 2020, doi: 10.3390/bdcc4010001.
- [13] R. R. Kurniawan and I. Zufria, “Penerapan Text Mining Pada Sistem Penyeleksian Judul Skripsi Menggunakan Algoritma Latent Dirichlet Allocation (LDA),” *Indones. J. Comput. Sci.*, vol. 11, no. 3, pp. 1036–1052, 2022, [Online]. Available: <http://3.8.6.95/ijcs/index.php/ijcs/article/view/3120/84>
- [14] A. Chamekh, M. Mahfoudh, and G. Forestier, “Sentiment Analysis Based on Deep Learning in E-Commerce,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13369 LNAI, pp. 498–507, 2022, doi: 10.1007/978-3-031-10986-7_40.
- [15] A. Ligthart, C. Catal, and B. Tekinerdogan, *Systematic reviews in sentiment analysis: a tertiary study*, vol. 54, no. 7. Springer Netherlands, 2021. doi: 10.1007/s10462-021-09973-3.
- [16] Pemerintah Pusat, “Instruksi Presiden (INPRES) tentang Dukungan Penyelenggaraan FIFA U-20 World Cup Tahun 2021,” SK No 031271 A, 2020 [Online]. Available: <https://peraturan.bpk.go.id/Home/Details/145868/inpres-no-8-tahun-2020>
- [17] FIFA, “FIFA Football Stadiums Guidelines,” Fédération Internationale de Football Association, 2022. <https://www.fifa.com/technical/stadium-guidelines> (accessed May 01, 2023).
- [18] FIFA, “FIFA removes Indonesia as host of FIFA U-20 World Cup 2023™,” Fédération Internationale de Football Association, 2023. <https://www.fifa.com/about-fifa/organisation/media-releases/fifa-removes-indonesia-as-host-of-fifa-u-20-world-cup-2023-tm> (accessed May 01, 2023).
- [19] Samsir, Ambiyar, U. Verawadina, F. Edi, and R. Watrianthos, “Analisis Sentimen Pembelajaran Daring pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [20] M. Arifiansyah Ayub, “Analisis Topik Ekonomi Dengan Algoritma K-Means Pada Media Online Era Pandemi Covid-19 Di Sulawesi Tenggara,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 133–138, 2021, doi: 10.33387/jiko.v4i2.3235.
- [21] H. Jayadianti, R. Damayanti, and ..., “Latent Semantic Analysis (Lsa) Dan Automatic Text Summarization (Ats) Dalam Optimasi Pencarian Artikel Covid 19,” *Semin. Nas. ...*, vol. 2020, no. Semnasif, pp. 52–59, 2020, [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/4085>
- [22] M. R. Nurhusen, J. Indra, and K. A. Baihaqi, “Analisis Sentimen Pengguna Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak (BBM) Menggunakan Metode Logistic Regression,” vol. 7, pp. 276–282, 2023, doi: 10.30865/mib.v7i1.5491.
- [23] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter,” *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [24] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 5, no. 2, pp. 697–711, 2021.
- [25] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022, doi: 10.30865/mib.v6i1.3530.
- [26] 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.