



Sentimen Analisis Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naïve Bayes

Rozaq Sulastiyono, Agus Setiawan, Setiya Nugroho*

Teknik, Teknik Informatika, Universitas Muhammdiyah Magelang, Magelang

Jl. Mayjen Bambang Soengeng, Glagak, Sumberrejo, Kec. Mertoyudan, Kabupaten Magelang, Jawa Tengah, Indonesia

Email: ¹rozaq082000@gmail.com, ²setiawan@unimma.ac.id, ^{3*}setiya@ummgl.ac.id

Email Penulis Korespondensi: setiya@ummgl.ac.id

Submitted: 27/06/2023; Accepted: 31/07/2023; Published: 31/07/2023

Abstrak—Sepak bola adalah olahraga yang paling populer di dunia termasuk Indonesia. Beberapa turnamen sepak bola mengalami pengunduran jadwal akibat Covid-19 termasuk Piala Dunia U-20 yang akan dilaksanakan pada tahun ini di Indonesia. Namun turnamen tersebut mendapatkan penolakan dari beberapa pihak yang tidak menginginkan salah satu negara peserta mengikuti turnamen. Akibatnya federasi FIFA membatalkan Indonesia sebagai tuan rumah Piala Dunia U-20 2023. Dengan adanya polemik yang terjadi dalam beberapa minggu terakhir ini banyak menimbulkan opini maupun pendapat dari para pengguna media sosial salah satunya adalah twitter. Pada penelitian ini berfokus untuk mengklasifikasi terhadap opini dari cuitan di twitter mengenai pembatalan Indonesia menjadi tuan rumah piala dunia U-20. Klasifikasi sentiment ini dengan metode Lexicon Based untuk menentukan sentimen positif, negatif dan netral, klasifikasi dengan metode Multinomial Naïve Bayes untuk perhitungan akurasi dengan Confusion Matrix. Berdasarkan sistem yang dibangun, didapatkan akurasi sebesar 85% dengan precision sebesar 85%, recall sebesar 85%, dan nilai f1-score sebesar 83%.

Kata Kunci: Analisis Sentimen; Naive Bayes; Piala Dunia U-20; Lexicon Based; Twitter

Abstract—Football is the most popular sport in the world, including Indonesia. Several football tournaments have been postponed due to Covid-19, including the U-20 World Cup which will be held this year in Indonesia. However, the tournament was rejected by several parties who did not want one of the participating countries to take part in the tournament. As a result, the FIFA federation canceled Indonesia as the host for the 2023 U-20 World Cup. The polemic that has occurred in the last few weeks has generated many opinions and opinions from social media users, one of which is Twitter. This study focuses on classifying opinions from tweets on Twitter regarding the cancellation of Indonesia from hosting the U-20 World Cup. This sentiment classification uses the Lexicon Based method to determine positive, negative and neutral sentiments, classification uses the Multinomial Naïve Bayes method for calculating accuracy with the Confusion Matrix. Based on the system built, an accuracy of 85% is obtained with a precision of 85%, a recall of 85%, and an f1-score value of 83%.

Keywords: Sentiment Analysis; Naive Bayes; U-20 World Cup; Lexicon Based; Twitter

1. PENDAHULUAN

Twitter merupakan salah satu media sosial terpopuler yang berperan sebagai wadah komunikasi di masyarakat. Twitter memungkinkan orang di seluruh dunia untuk tetap berhubungan dengan keluarga, teman, dan orang yang mereka cintai menggunakan komputer atau ponsel mereka. Layanan yang disediakan Twitter untuk penggunaannya termasuk pembuatan pesan status (disebut "tweet"), yang dapat dibaca oleh pengguna Twitter lainnya dan biasanya pendapat pengguna tentang berbagai topik dengan batas 140 karakter, sehingga twitter menjadi salah satu situs yang menyediakan kumpulan data opini dari masyarakat di seluruh dunia[1].

Salah satu olahraga yang paling diminati masyarakat dunia adalah olahraga sepak bola[2]. Indonesia sendiri termasuk negara penggemar sepak bola paling banyak di dunia. Indonesia sendiri merupakan salah satu negara dengan penggemar sepak bola terbanyak di dunia dengan 77% penduduk Indonesia memiliki ketertarikan dengan olahraga sepak bola[3]. Beberapa turnamen sepak bola ditunda akibat Covid-19, termasuk Piala Dunia U-20 yang akan digelar di Indonesia tahun ini. Namun, turnamen tersebut ditolak oleh beberapa pihak yang tidak menginginkan negara peserta untuk mengikuti turnamen tersebut. Akibatnya, FIFA membatalkan pencalonan Indonesia sebagai tuan rumah Piala Dunia U-20 2023. Turnamen ini menjadi salah satu trending topik di Twitter yang memunculkan berbagai tweet perspektif warganet yang memicu timbulnya pro-kontra antara masyarakat dan pihak tertentu, karena mendapatkan penolakan dari beberapa pihak yang tidak menginginkan salah satu negara peserta mengikuti turnamen, yang membuat Timnas Indonesia U-20 resmi dibubarkan Shin Tae Yong pada Tanggal 1 April 2023 di Stadion Utama Gelora Bung Karno[4].

Analisis sentimen masih menjadi bagian dari riset opini. Artinya, proses memahami, mengekstrak, dan memproses data teks secara otomatis untuk mendapatkan informasi sentimen didalam kalimat opini[5]. Terdapat beberapa metode yang digunakan dalam sentimen analisis, diantaranya dengan metode Support Vector Machine dalam analisis opini masyarakat terhadap program Kartu Prakerja[6], metode K-Nearest Neighbor dalam analisis sentiment pengguna twitter terhadap penggunaan aplikasi Ovo, Gopay, LinkAja[7], dan metode Naïve bayes dalam analisis sentiment terhadap tokoh publik[8].

Beberapa penelitian terkait sentimen analisis dalam sepak bola dilakukan oleh beberapa peneliti, diantaranya dalam analisis opini tentang tragedi kanjuruhan menggunakan metode naïve bayes. Penelitian ini bertujuan untuk mengklasifikasi sentimen serta mengetahui tingkat akurasi pada algoritma Multinomial Naïve Bayes dalam tragedi Kanjuruhan. Hasil dari penelitian didapatkan akurasi sebesar 75% dari 2.402 data dengan

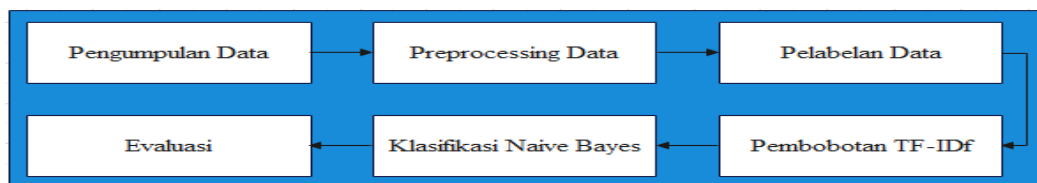
data training sebesar 90% dan data uji sebesar 10% [9]. Berdasarkan penelitian sebelumnya mengenai “Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022”. Tujuan penelitian ini untuk menganalisis opini publik tentang analisis sentimen masyarakat dalam penggunaan Twitter mengenai Piala Dunia FIFA 2022. Hasil dari penelitian ini menghasilkan 85% dengan metode SVM[10]. Penelitian yang sejenis juga dengan judul “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor”. Tujuan dari penelitian ini adalah untuk menganalisis sentimen pada setiap kalimat dari pengguna twitter terhadap persepakbolaan Indonesia apakah memiliki sentimen negatif atau positif menggunakan K-Nearest Neighbor. Hasil dari penelitian didapatkan akurasi 79% dengan metode KNN dari 2000 data tweet berbahasa indonesia didapatkan[11].

Penelitian lain tentang penggunaan metode Naïve bayes dalam analisis sentimen yaitu opini masyarakat tentang performa Timnas Sepakbola Indonesia. Salah satu caranya adalah dengan melakukan analisis opini tentang Timnas Indonesia di jejaring sosial Twitter, dengan menggunakan metode klasifikasi dan algoritma Naive Bayes untuk mengklasifikasikan tweet positif atau negatif yang ditulis orang tentang Timnas Indonesia. Hasil penelitian disimpulkan bahwa analisis sentimen menggunakan algoritma Naïve Bayes dengan akurasi sebesar 83%, kemudian skor akurasi positif sebesar 86%, skor akurasi negatif sebesar 81% [12]. Penelitian lain tentang penggunaan metode Naïve bayes dalam analisis sentiment komentar netizen Instagram terhadap racism di sepak bola Indonesia. Percobaan dilakukan pada 650 data komentar, data tersebut dibagi menjadi 2 kategori yaitu kombinasi banyaknya dokumen latih dan dokumen uji. Hasil pengujian yang telah dilakukan, accuracy tertinggi diperoleh pada pembagian 90:10 dengan nilai accuracy 82% [13]. Informasi yang diperoleh untuk penelitian seringkali mengandung banyak noise, yang membuat informasi tersebut tidak terstruktur. Dengan bantuan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diolah menjadi informasi yang lebih terstruktur

Berdasarkan permasalahan di atas penulis melakukan analisis sentimen media sosial di Twitter terkait dengan pembatalan Indonesia menjadi tuan rumah Piala Dunia U-20. Tujuan penelitian ini adalah untuk menilai opini masyarakat di Twitter dengan menggunakan pendekatan metode Naïve Bayes pada pembatalan Indonesia yang akan menjadi tuan rumah Piala Dunia. Kontribusi pada penelitian ini adalah untuk dapat memberikan gambaran umum terkait dengan adanya polemik yang terjadi pada persepakbola Indonesia khususnya pada pembatalan piala dunia u-20 serta memberikan representasi dari masyarakat Indonesia dalam menghadapi masalah ini dan memberikan masukan kepada para pihak tentang peningkatan kualitas sepak bola sedang diterapkan di Indonesia. Selanjutnya, penelitian ini dapat membantu kemajuan penelitian terhadap analisis sentimen. Pada penelitian ini berfokus untuk mengklasifikasi terhadap opini dari cuitan di twitter yang mengandung nilai positif, negative mengenai pembatalan Indonesia menjadi tuan rumah piala dunia U-20 menggunakan Naïve bayes dengan model multinomial. Model multinomial pada Naïve bayes dipilih karena dapat memperhatikan jumlah kemunculan kata pada dokumen dan dapat mengklasifikasikan permasalahan yang bersifat multiple-class [14]. Metode ini dipilih karena teknik yang digunakan untuk menentukan setiap kelas masalah dibagi per masalah. Perhitungan numerik berdasarkan pendekatan kelompok. Naïve Bayes memiliki beberapa keunggulan seperti sederhana dan akurasi yang tinggi [15]. Penelitian ini menggunakan algoritma Naïve Bayes pada label otomatis menggunakan lexicon based.

2. METODOLOGI PENELITIAN

Tahapan dalam penelitian ini yang dilakukan langkah demi langkah seperti yang ditunjukkan pada gambar 1.



Gambar 1. Alur penelitian

Pada gambar 1 terdapat 6 tahap dalam penelitian terhadap pembatalan Indonesia menjadi tuan rumah piala dunia U-20 yang meliputi Pengumpulan data, preprocessing data, pelabelan data, pembobotan tf-idf, klasifikasi naïve bayes, dan evaluasi. Data twitter dikumpulkan di tahap pengumpulan data yang kemudian diproses pada tahap preprocessing data. Data yang telah diproses kemudian dilabeli pada tahap pelabelan data yang kemudian diekstraksi di tahap pembobotan TF-IDF. Setelah itu dilakukan klasifikasi dengan Naïve bayes dan hasil tersebut akan dievaluasi. Penjelasan mengenai tahap-tahap tersebut akan dijelaskan dibawah ini.

2.1 Pengumpulan Data

Data yang diperoleh adalah data yang berbahasa Indonesia. Pengambilan data atau crawling kemudian digabungkan menjadi opini atau daftar opini kemudian diolah pada tahap pre-processing. Pengumpulan data berlangsung dari tanggal 23 Maret 2023 sampai 14 April 2023, dengan data yang diperoleh 2397 data Tweet. Data yang berhasil diambil disimpan di Excel dalam format csv. [16].



2.2 Pre-Processing Data

Pada tahap text preprocessing, teks dimodifikasi dan dibersihkan sehingga dapat dikelola untuk klasifikasi. Tahapan yang dilakukan dalam text preprocessing adalah Case Folding, Data Cleaning, Tokenization, Stopword Removal, dan Stemming.

1. Case Folding bertujuan untuk mengubah setiap huruf kapital menjadi huruf kecil, sehingga tidak terjadi kesalahan saat menggabungkan karakter atau huruf dalam sebuah kata[17].
2. Data Cleaning adalah proses membersihkan data dari kolom komentar duplikat dan atribut yang tidak penting seperti menghapus data dari karakter yang tidak perlu seperti tanda baca dan berbagai karakter lainnya[18].
3. Tokenization bertujuan memisahkan kata menjadi potongan kata sebelum dianalisis lebih lanjut[19].
4. Stopword Removal dilakukan untuk menghilangkan kata-kata yang tidak berpengaruh pada proses klasifikasi sentiment seperti waktu, penghubung dan lainnya[20].
5. Stemming adalah langkah mengubah setiap kata dari akhiran menjadi kata dasar. Sumber sebenarnya adalah algoritma Nazief dan Adriani yang terdapat library sastrawi[21].

2.3 Pelabelan Data

Proses pelabelan data dilakukan secara otomatis. Penulis melakukan pelabelan secara otomatis dengan menggunakan Lexicon Based. Klasifikasi positif dapat dilihat pada isi tweet yang meliputi kata-kata positif, dukungan dan pernyataan setuju. Kelas negatif adalah kelas yang datanya berisi kata-kata yang bermakna negatif, ejekan, dan kontra [22].

2.4 Pembobotan TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode algoritma yang memberikan bobot pada teks. TF adalah berapa kali kata tersebut muncul dalam dokumen, sedangkan IDF adalah kebalikan dari dokumen yang mengandung kata tersebut. TF dan IDF dikalikan untuk mendapatkan bobot kata[23].

2.5 Klasifikasi Naïve Bayes

Naive Bayes adalah metode klasifikasi yang berguna menurut aturan Bayes yang ada menggunakan perhitungan probabilistik. Dalam proses klasifikasi, klasifikasi dilakukan dengan menggunakan nilai probabilitas maksimal. Peneliti sering menggunakan metode Naive Bayes karena dianggap efektif untuk mendapatkan hasil yang akurat dengan akurasi yang tinggi. Rumus untuk menghitung Naive Bayes ada di persamaan[24].

$$P(c|d) = \frac{P(c) \cdot P(d|c)}{P(d)} \tag{1}$$

P(c/d) : posterior yaitu peluang kelas c diberikan dokumen d

P(c) : prior atau peluang awal munculnya kategori c

P(d/c) : Nilai Likelihood

P(d) : Evidence atau peluang munculnya dokumen d

2.5.1 Multinomial Naïve Bayes

Multinomial Naive Bayes adalah teknik pembelajaran mesin yang digunakan untuk aplikasi klasifikasi, khususnya dalam natural language processing[25]. Ini adalah variasi dari algoritma Naive Bayes, yang didasarkan pada teorema Bayes dan premis independensi fitur. Fitur pada varian multinomial diasumsikan memiliki distribusi multinomial, yang umumnya terjadi pada data teks yang fiturnya adalah jumlah kata. Algoritma beroperasi dengan menghitung kemungkinan setiap kelas yang diberi fitur input dan memilih kelas dengan probabilitas tertinggi sebagai kelas yang diprediksi. Multinomial Naive Bayes telah digunakan dalam berbagai aplikasi, termasuk penyaringan spam, analisis sentimen, dan deteksi penyakit.

2.6 Evaluasi

Tujuan evaluasi dalam penelitian ini menentukan performansi atau kinerja dari model yang diusulkan. Metode analisis yang digunakan dalam penelitian ini adalah Naive Bayes. Dua set utama yang diperlukan dalam klasifikasi Naive Bayes, yaitu set pelatihan dan set tes. Training set digunakan sebagai data training untuk mengidentifikasi objek yang akan diklasifikasikan, sedangkan test set digunakan sebagai pengukur akurasi dan presisi klasifikasi Naive Bayes[26]. Berdasarkan hasil klasifikasi keseluruhan yang diperoleh, proses ini dilakukan untuk menguji hasil klasifikasi menggunakan metode Naive Bayes. Pada tahap evaluasi ini penulis menggunakan confusion matrix.

Tabel 1. Confusion Matrix

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | TP | FN |
| | Negative | FP | TN |

Nilai True Positive (TP) dan True Negative (TN) adalah hasil klasifikasi yang mempunyai nilai benar. Nilai False Positive (FP) adalah nilai yang diprediksi bernilai positive namun bernilai negatif, dan berbanding terbalik dengan False Negative (FN) dimana terprediksi sebagai negative namun memiliki nilai negative[1]. Dari hasil perhitungan confusion matrix didapatkan nilai Accuracy, Presicion, Recall dan F1-Score.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{FP+FN} \tag{4}$$

$$\text{F1 - Score} = 2 * \frac{\text{Precision*Recall}}{\text{Precision+Recall}} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Tahapan ini menggunakan data sekunder. Dataset berisi teks dalam bahasa Indonesia yang diambil dari media sosial Twitter dengan menggunakan bahasa pemrograman Python. Untuk membuat Steam Twitter API, perlu memerlukan kunci dan token akses sebagai bukti autentikasi pengembang Twitter. Pengumpulan data berlangsung dari tanggal 23 Maret 2023 sampai 14 April 2023, dengan data yang diperoleh 2397 data Tweet. Data yang berhasil diambil disimpan di Excel dalam format csv. Hasil pengumpulan data menggunakan twitter API[27] disajikan seperti pada gambar 2.

| | Uatetime | Text |
|------|------------------------------|---|
| 0 | 2023-03-30 23:59:24+00:00 | https://t.co/wvkhKE1lum Argentina kemungkinan besar menggantikan Indonesia sebagai tuan rumah Piala Dunia U-20 2023. Presiden FIFA Gianni Infantino sudah memberi sinyal positif. |
| 1 | 2023-03-30 23:57:35+00:00 | Jokowi Kecewa dan Sedih Indonesia Batal Gelar Piala Dunia U-20 2023 https://t.co/IsokV24W00 |
| 2 | 2023-03-30 23:57:28+00:00 | Sebagai penggemar sepakbola Indonesia, sudah saatnya move on dari berita batalnya Indonesia menjadi tuan rumah Piala Dunia U-20 2023. Saatnya berpikir kedepan, ada banyak hal yang harus dipikirkan, misale jawab pertanyaan 'kapan nikah?' |
| 3 | 2023-03-30 23:56:59+00:00 | Jika kalian bilang FIFA standar ganda, iya sih. Rusia bisa kena hukuman dan Israel tidak. Tapi tolong diingat, yg menolak Rusia bertanding adalah negara-negara barat seperti Inggris dan Prancis. Nah, di Piala Dunia U-20 2023, hanya publik Indonesia yang menolak. Lemah. |
| 4 | 2023-03-30 23:56:33+00:00 | PKB: Pecinta Bola Tandai Parpol & Tokoh yang Gagalkan Piala Dunia U-20 Para partai dan tokoh oportunist mulai memanfaatkan kekisruhan PD-U20 utk cari simpati masyarakat. Partai yg tadinya mnolak sdang tmbakan pluru ke PSSI seakan PSSIyg salah. CUCI TANGAN https://t.co/LR3t2J1gTR |
| ... | ... | ... |
| 2392 | 2023-04-08 00:27:05+00:00 | Pemerintah merenovasi puluhan stadion di Indonesia sebagai tindak lanjut Tragedi Kanjuruhan dan juga Piala Dunia U-20 yang akhirnya batal Lantas, seberapa besar anggaran yang digelontorkan dan seperti apa sih desainnya? Selengkapny di Focus kumparan. https://t.co/m9P0Mf7I0I |
| 2393 | 2023-04-08 00:20:57+00:00 | Ketum PSSI Erick Thohir terus menjalin komunikasi dengan FIFA, Alhamdulillah berbuah hasil maksimal Kegigihan Lobi yang dilakukan @erickthohir adalah Indonesia berhasil terhindar dari sanksi berat pasca batal menjadi tuan rumah Piala Dunia U-20 2023. https://t.co/RNrxnbgblF |
| 2394 | 2023-04-08 00:14:10+00:00 | Fenomena Piala Dunia U-20 Menarik untuk Elektabilitas Ganjar https://t.co/9V4attu8uV |
| 2395 | 2023-04-08 00:11:55+00:00 | Iran Puji Ketegasan Indonesia Menolak Keikutsertaan Timnas Israel di Piala Dunia U-20 https://t.co/fuxA90WNLO https://t.co/FdQxfMirBe |
| 2396 | 2023-04-08 00:00:06+00:00 | FIFA pada akhirnya memberikan kartu kuning untuk sepak bola Indonesia, yang artinya kita hanya diberikan sanksi ringan berupa sanksi administrasi, pasca keputusan pembatalan posisi Indonesia sebagai tuan rumah Piala Dunia U-20 2023. https://t.co/m9P0Mf7I0I |

2397 rows x 2 columns

Gambar 2. Hasil Pengumpulan Data

3.2 Pre-processing Data

Preprocessing data merupakan tweet yang telah dikumpulkan dari twitter, banyak memiliki noise sehingga perlu adanya tahapan proses menghilangkan noise agar proses analisis sentimen menjadi lebih akurat dan dapat digunakan secara general.

3.2.1 Case Folding

Tahapan ini merubah huruf kapital diubah menjadi huruf kecil, agar nantinya memudahkan dalam identifikasi kata. Hasil Case Folding disajikan seperti pada tabel 2.

Tabel 2. Case Folding

| Sebelum | Sesudah |
|---|---|
| @RasSaaaa Yang seharusnya bertanggung jawab memang PSSI atas gagalnya piala dunia U-20, mereka seakan-akan mengupayakan namun nihil sampai sekarang | @rassaaaa yang seharusnya bertanggung jawab memang pssi atas gagalnya piala dunia u-20, mereka seakan-akan mengupayakan namun nihil sampai sekarang |

3.2.2 Data Cleaning

Pada proses ini untuk menghapus seperti karakter “@”, link URL, titik, koma, tanda seru, tanda tanya, menghapus emoji, hastag, spasi ganda, mention, menghapus retweet, menghapus tanda baca serta angka yang tidak digunakan pada tahap selanjutnya. Hasil Data Cleaning disajikan seperti pada tabel 3.

Tabel 3. Data Cleaning

| Sebelum | Sesudah |
|---|--|
| @rassaaaa yang seharusnya bertanggung jawab memang pssi atas memang pssi atasagalnya piala dunia u-20, mereka seakanakan mengupayakan namun nihil sampai sekarang | yang seharusnya bertanggung jawab memang pssi atasagalnya piala dunia mereka seakanakan mengupayakan namun nihil sampai sekarang |

3.2.3 Tokenization

Setelah proses cleaning data selanjutnya tahap tokenisasi yaitu proses memecah teks menjadi kata, simbol, karakter, atau tanda baca untuk membuat token yang dapat dianalisa. Hasil Tokenization disajikan seperti pada tabel 4.

Tabel 4. Tokenization

| Sebelum | Sesudah |
|--|---|
| yang seharusnya bertanggung jawab memang pssi atasagalnya piala dunia mereka seakanakan mengupayakan namun nihil sampai sekarang | ['yang', 'seharusnya', 'bertanggung', 'jawab', 'memang', 'pssi', 'atas', 'agalnya', 'piala', 'dunia', 'mereka', 'seakanakan', 'mengupayakan', 'namun', 'nihil', 'sampai', 'sekarang'] |

3.2.4 Stopword Removal

Menggunakan data yang telah dilakukan tokenize dengan jumlah yang sama dapat dilanjutkan ke tahap stopword removal. Proses ini menghilangkan kata-kata yang tidak perlu atau yang tidak mengandung opini. Hasil Stopword Removal disajikan seperti pada tabel 5.

Tabel 5. Stopword Removal

| Sebelum | Sesudah |
|---|---|
| ['yang', 'seharusnya', 'bertanggung', 'jawab', 'memang', 'pssi', 'atas', 'agalnya', 'piala', 'dunia', 'mereka', 'seakanakan', 'mengupayakan', 'namun', 'nihil', 'sampai', 'sekarang'] | ['bertanggung', 'pssi', 'agalnya', 'piala', 'dunia', 'seakanakan', 'mengupayakan', 'nihil'] |

3.2.5 Stemming

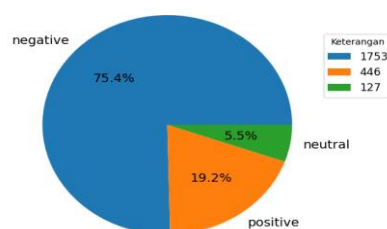
Pada tahap akhir preproceesing, dilakukan proses untuk mengubah semua kata dalam dokumen menjadi kata dasar dengan menghapus semua imbuhan. Penelitian ini menggunakan library Python sastrawi dalam proses stemming. Hasil Stemming disajikan seperti pada tabel 6.

Tabel 6. Stemming

| Sebelum | Sesudah |
|---|--|
| ['bertanggung', 'pssi', 'agalnya', 'piala', 'dunia', 'seakanakan', 'mengupayakan', 'nihil'] | ['tanggung', 'pssi', 'agal', 'piala', 'dunia', 'seakanakan', 'upaya', 'nihil'] |

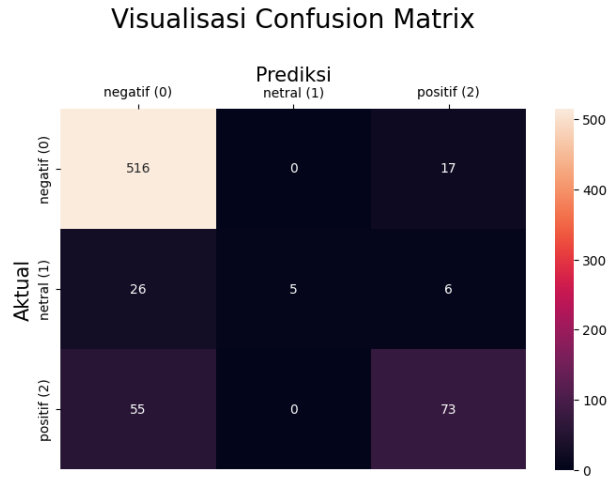
3.3 Pelabelan Data

Proses pelabelan dilakukan dengan menggunakan metode berbasis lexicon. Sentimen yang digunakan terdiri dari tiga kategori: positif, negatif dan netral. Kemudian, data yang telah diterjemahkan sebelumnya dibawa ke tahap label untuk diproses, yang dibagi menjadi tiga kategori yaitu positif, negatif, dan netral. Hasil pelabelan data disajikan seperti pada gambar 3.



Gambar 3. Hasil Pelabelan Data

digunakan untuk mengevaluasi akurasi model naïve bayes dan hasil menggunakan confusion matrix dapat dilihat pada gambar 7.



Gambar 7. Hasil Confusion Matrix

Berdasarkan gambar diatas mendapatkan hasil evaluasi sebagai uji accuracy Naïve Bayes sebesar 85% yang terbilang cukup baik dengan precision sebesar 85%, recall sebesar 85%, dan nilai f1-score sebesar 83%. Hasil evaluasi disajikan pada gambar 8.

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| -1 | 0.86 | 0.97 | 0.91 | 533 |
| 0 | 1.00 | 0.14 | 0.24 | 37 |
| 1 | 0.76 | 0.57 | 0.65 | 128 |
| accuracy | | | 0.85 | 698 |
| macro avg | 0.87 | 0.56 | 0.60 | 698 |
| weighted avg | 0.85 | 0.85 | 0.83 | 698 |

Gambar 8. Hasil Evaluasi

4. KESIMPULAN

Berdasarkan hasil yang telah dilakukan pada penelitian analisis sentimen masyarakat pada Twitter terhadap pembatalan Indonesia menjadi tuan rumah Piala Dunia U-20, menggunakan metode Naïve Bayes model Multinomial hasil sentiment cenderung negatif dengan persentase 75%, positif tweet memiliki nilai persentase 19%, dan netral 5%. Data tweet yang telah diproses dan dilabeli dibagi menjadi dua jenis yaitu data latih mendapatkan sebanyak 80% dan data uji sebanyak 20%. Dalam melakukan klasifikasi Naïve Bayes memiliki akurasi sebesar 85% dengan precision sebesar 85%, recall sebesar 85%, dan nilai f1-score sebesar 83%. Dari hasil dataset yang digunakan dalam penelitian ini berupa tweet media sosial Twitter, dapat disimpulkan bahwa Multinomial Naive Bayes memiliki akurasi klasifikasi sentimen yang cukup baik dalam penelitian ini. Berdasarkan hasil penelitian diharapkan pada penelitian selanjutnya dapat menggunakan model pada naïve bayes yang lain seperti model Bernoulli atau model klasifikasi lain dan juga menggunakan cross validation dalam memaksimalkan hasil akurasi.

REFERENCES

- [1] 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.
- [2] Y. Yudhanto and A. Putra, “Perancangan Dan Pembuatan Sistem Kompetisi Sepak Bola Berbasis Web,” *Indones. J. Appl. Informatics*, vol. 1, no. 2, p. 23, 2017, doi: 10.20961/ijai.v1i2.14330.
- [3] I. Indriati, M. Marji, and S. Pakpahan, “Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7259–7267, 2019.
- [4] M. H. Sidik, S. Widiyanesti, and ..., “Analisis Sentimen dan Topic Modelling Terhadap Tim Nasional Indonesia di Kejuaraan AFF Suzuki Cup 2020 Berdasarkan Opini Pengguna Twitter,” *eProceedings ...*, vol. 9, no. 5, pp. 2783–2796, 2022, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/view/18369%0Ahttps://openlibrar>



- ypublications.telkomuniversity.ac.id/index.php/management/article/download/18369/17980
- [5] R. Lasepa, S. Riyadi, S. Ramadhan, and D. D. Saputra, "Analisis Sentimen Terhadap Perspektif Warganet Atas Tragedi Kanjuruhan Malang di Twitter Menggunakan Naïve Bayes Classifier," vol. 8, no. 1, pp. 1–8, 2021.
 - [6] N. Hendrastuty et al., "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, [Online]. Available: <http://situs.com>
 - [7] S. S. Salim and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.
 - [8] D. Rusdianan and D. Rosiyadi, "Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 230–235, 2019.
 - [9] I. S. Thalib, S. K. Gusti, F. Yanto, and M. Affandes, "Klasifikasi Sentimen Tragedi Kanjuruhan Pada Twitter Menggunakan Algoritma Naïve Bayes," vol. 4, pp. 467–473, 2023, doi: 10.30865/json.v4i3.5852.
 - [10] C. Of, N. Bayes, S. Vector, M. For, and A. Twitter, "Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022," vol. 13, no. 01, 2023.
 - [11] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Perseppakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
 - [12] M. M. S. Reino Prajamukti, Jayanta, "Klasifikasi Dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma Naïve Bayes (Studi kasus :Timnas Indonesia Senior,U-23, dan u-19)," *Seinasi-Kesi*, pp. 1–8, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/seinasikesi/article/view/1909>
 - [13] M. I. Yusuf, "Analisis Sentimen Komentar Netizen Instagram Terhadap Racism Di Sepak Bola Indonesia Dengan Metode Naive Bayes," vol. 2, pp. 1–8, 2022.
 - [14] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial," *J. Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, doi: 10.14710/j.gauss.v9i3.27963.
 - [15] D. Ramadhan and E. B. Setiawan, "Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," ... *Telkomuniversity.Ac.Id*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>
 - [16] M. Undap, V. P. Rantung, and P. T. D. Rompas, "Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based," *Jointer - J. Informatics Eng.*, vol. 2, no. 02, pp. 39–46, 2021, doi: 10.53682/jointer.v2i02.44.
 - [17] A. Rosadi et al., "Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classifier," *J. Ilm. Komputasi*, vol. 20, no. 1, pp. 39–52, 2021, doi: 10.32409/jikstik.20.1.401.
 - [18] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
 - [19] 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.
 - [20] D. Winarso, Yanda Noor Yudha, and Syahril, "Analisis Sentimen Masyarakat Pada Twiter Terhadap Isu Covid-19 Menggunakan Metode Lexicon Based," *J. Fasilkom*, vol. 11, no. 2, pp. 97–103, 2021, doi: 10.37859/jf.v11i2.2772.
 - [21] A. Syakur, "Implementasi Metode Lexicon Base Untuk Analisis Sentimen Kebijakan Pemerintah Dalam Pencegahan Penyebaran Virus Corona Covid-19 Pada Twitter," *J. Ilm. Inform. Komput.*, vol. 26, no. 3, pp. 247–260, 2021, doi: 10.35760/ik.2021.v26i3.4720.
 - [22] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
 - [23] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
 - [24] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, "Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - [25] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.
 - [26] M. M. Mala Olhang, S. Achmadi, and F. . A. Wibisono, "Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier (Nbc)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 214–221, 2020, doi: 10.36040/jati.v4i2.2695.
 - [27] A. Z. Malik, E. Utami, and S. Raharjo, "Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 dengan Metode K-NN," *J. Inf. Politek. Indones. Surakarta*, vol. 5, no. 2, pp. 1–7, 2019.
 - [28] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.