

Implementasi Algoritma *Naive Bayes* Terhadap Analisis Sentimen Perubahan Piala Dunia U-20

Epliani Limbong Rara^{1*}, Evangs Mailoa²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail *Corresponding Author*: 672019319@student.uksw.edu

Abstract

The change of host in the FIFA U-20 World Cup 2023 has sparked controversy among various groups, as seen on the social media platform Twitter. This controversy arose due to Indonesia's rejection of hosting the U-20 World Cup, allegedly because Indonesia refused the participation of the Israel U-20 national team to play in Indonesia. This research aims to implement the Naive Bayes method to classify public sentiment on Twitter regarding this change, distinguishing between positive and negative sentiments. The testing was conducted using a confusion matrix, which resulted in accuracy, precision, and recall values of 96%, 94.83%, and 99.56% respectively. Based on these results, it can be concluded that the use of the Naive Bayes method in this research is quite good, and the performance of the system improves with an increased amount of training data.

Keywords: *World Cup; Naive Bayes; Sentiment Classification; Confusion Matrix*

Abstrak

Adanya perubahan tuan rumah pada turnamen Piala dunia U-20 FIFA 2023 menimbulkan kontroversi dari berbagai golongan yang dapat dilihat dari media sosial *Twitter*, lantaran adanya penolakan Indonesia menjadi tuan rumah pada Piala Dunia U-20 yang diduga karena Indonesia menolak keikutsertaan Timnas Israel U-20 untuk bermain di Indonesia. Penelitian ini bertujuan untuk mengimplementasikan metode *Naive Bayes* dalam melakukan klasifikasi sentimen masyarakat pada *Twitter* terkait adanya perubahan tersebut yang berupa sentimen positif dan negatif. Pengujian dilakukan menggunakan *confusion matrix* yang diperoleh hasil nilai *accuracy*, *precision*, dan *recall* masing-masing adalah 96%, 94,83%, dan 99,56%. Berdasarkan hasil tersebut dapat disimpulkan bahwa penggunaan metode *Naive Bayes* pada penelitian ini cukup baik, semakin banyak data *training* yang digunakan semakin baik hasil kinerja dari sistem.

Keywords: *Piala dunia; Naive Bayes; Klasifikasi Sentimen; Confusion Matrix*

1. Pendahuluan

Piala dunia U-20 FIFA (*Federation Internationale de Football Association*) 2023 merupakan salah satu topik perbincangan hangat di kalangan masyarakat saat ini dikarenakan gagalnya Indonesia menjadi tuan rumah dalam penyelenggaraan turnamen tersebut. Dilansir dari situs resmi FIFA fifa.com pada Rabu 29 Maret 2023 bahwa FIFA resmi menghapus Indonesia sebagai tuan rumah pada piala dunia 2023 tanpa mengubah jadwal turnamen yang ada [1]. Gagalnya Indonesia menjadi tuan rumah ini diduga karena Indonesia menolak keikutsertaan Timnas dari sebuah negara tertentu pada U-20 untuk bermain di Indonesia karena negara tersebut dianggap tidak memiliki hubungan diplomatik dengan Indonesia sampai hingga sebuah negara tertentu di Timur Tengah merdeka, namun ada juga yang mengatakan bahwa bukan karena keikutsertaan negara tertentu tersebut karena turnamen ini merupakan acara resmi FIFA [2].

Peristiwa gagalnya Indonesia menjadi tuan rumah ini menuai kontroversi sehingga menimbulkan pro dan kontra di kalangan Masyarakat, ada yang setuju dengan pembatalan tersebut karena disangkutkan dengan tragedi Kanjuruhan yang menilai bahwa Indonesia belum siap menjadi tuan rumah, dan ada juga yang tidak setuju terutama bagi pecinta sepak bola yang menantikan turnamen ini berlangsung di Indonesia. Peristiwa tersebut terutama sangat berdampak pada sektor ekonomi Indonesia yang diperkirakan mengalami kerugian hingga triliunan. Selain kehilangan potensi keuntungan dari penyelenggaraan turnamen

tersebut, kerugian Indonesia paling besar dirasakan oleh berbagai sektor seperti sektor pariwisata, akomodasi, transportasi, pakaian, kuliner, bahkan aksesori [3]. Selain pada sektor ekonomi, peristiwa tersebut juga berdampak pada sektor politik. Saat ini, isu adanya campur aduk antara olahraga dan politik masih diperbincangkan dikalangan masyarakat yang dikaitkan dengan penolakan keikutsertaan sebuah negara tertentu dalam turnamen karena isu agama sangat berpengaruh terhadap pilihan masyarakat menjelang Pemilu 2024 [4].

Bentuk opini masyarakat ini dapat dilihat di berbagai media sosial salah satunya melalui *Twitter*. *Twitter* merupakan salah satu media sosial yang sangat populer di kalangan banyak orang dan penggunaannya terus meningkat. Dari awal peluncurannya, *Twitter* sudah menjadi situs yang paling banyak dikunjungi oleh orang di internet [5] dengan jumlah pengguna di tahun 2022 mencapai 18,45 juta [6]. Melalui *Twitter*, pengguna dapat memberikan dan mengakses informasi atau topik yang sedang *trending* di dunia secara cepat dan luas [7].

Oleh karena itu, penting untuk mengetahui bagaimana pandangan masyarakat mengenai batalnya Indonesia menjadi tuan rumah Piala Dunia U-20 tersebut agar pembaca, penyelenggara maupun pemerintah dapat mengetahui seperti apa tanggapan (sentimen) masyarakat mengenai peristiwa tersebut atas keputusan yang telah diambil.

Dalam melakukan analisis sentimen pada penelitian ini menggunakan metode *Naive Bayes* dengan bantuan *tools Rapidminer*. Metode *Naive Bayes* merupakan salah satu metode *machine learning* yang sering digunakan untuk melakukan pengklasifikasian pada data [8]. Oleh karena itu, metode ini cocok untuk diimplementasikan pada penelitian ini dalam menentukan klasifikasi sentimen pengguna *Twitter* terkait perubahan Piala Dunia U-20 ke dalam kategori positif dan negatif. Penelitian ini difokuskan pada perubahan Piala Dunia U-20, gagalnya Indonesia menjadi tuan rumah. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi bagi pemerintah dalam setiap pengambilan keputusan sehingga membawahkan ke arah yang lebih baik.

2. Tinjauan Pustaka

Beberapa penelitian sebelumnya telah dilakukan mengenai analisis sentimen menggunakan metode *Naive Bayes*, diantaranya Klasifikasi Sentiment Transformasi dan Reformasi Sepak Bola Indonesia Pada *Twitter*, dimana dalam mengklasifikasikan opini masyarakat pada *Twitter* yang dibagi ke dalam kategori Positif dan Negatif. Diperoleh hasil dengan nilai *accuracy* = 88%, nilai *precision* = 87%, dan nilai *recall* = 88% [9]. Penelitian ini membahas tentang sentimen terhadap terpilihnya Indonesia menjadi tuan rumah Piala Dunia U-20 2023 yang dilakukan sebelum adanya perubahan batalnya Indonesia tuan rumah Piala Dunia U-20 2023.

Pada penelitian Analisis Sentimen Terhadap PSSI Atas Tragedi Kanjuruhan Menggunakan Multinomial *Naive Bayes* yang membahas mengenai analisis opini Masyarakat pada *Twitter* terhadap PSSI sebelum dan sesudah terjadinya peristiwa Kanjuruhan ke dalam sentimen positif dan negative. Peristiwa ini seringkali dikaitkan dengan batalnya Indonesia menjadi tuan rumah yang dinilai belum siap sepenuhnya untuk menjadi tuan rumah. Hasil dari penelitian ini diperoleh nilai *accuracy* sebelum peristiwa Kanjuruhan sebesar 73% dan sesudah peristiwa Kanjuruhan sebesar 68% dimana keduanya menghasilkan sentimen negatif lebih banyak daripada sentimen positif [10].

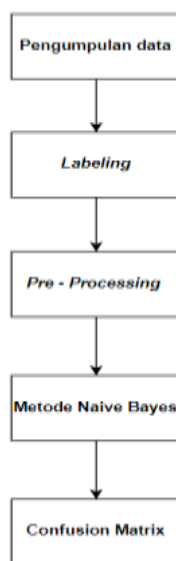
Penelitian lain dilakukan oleh Ahmad Faldhi Yudianto, Statiswaty, dan Natalis Ransi dengan judul *Naive Bayes Classifier Terhadap Sentiment Analysis Mengenai PSSI Setelah Tragedi Kanjuruhan Malang* yang berfokus pada kinerja PSSI. Penelitian ini menggunakan total keseluruhan data sebanyak 1500 data *tweet* dengan data *training* dan data *testing* masing – masing adalah 1200 data dan 300 data yang menghasilkan sentimen positif sebanyak 32 *tweet* dan sentimen negatif sebanyak 268 *tweet* dari data *testing* [11].

Beberapa penelitian yang sudah dilakukan sebelumnya terkait analisis sentimen mengenai topik Piala Dunia menggunakan metode *Naive Bayes* khususnya yang sudah dipaparkan di atas terkait terpilihnya Indonesia menjadi tuan rumah yang dilakukan untuk mengklasifikasikan sentimen positif dan sentimen negatif. Namun pada penelitian ini lebih difokuskan pada pengimplementasian algoritma *Naive Bayes* untuk mengetahui nilai akurasi yang dihasilkan dari pengimplementasian algoritma. Oleh karena itu, penelitian ini dilakukan untuk mengimplementasikan Algoritma *Naive Bayes* dalam melakukan analisis sentimen terhadap batalnya Indonesia menjadi tuan rumah Piala Dunia U-20 2023 tersebut.

3. Metodologi

Objek yang digunakan dalam penelitian ini adalah sentimen pengguna *Twitter* terkait perubahan Piala Dunia U-20 FIFA 2023 menggunakan metode *Naive Bayes* dalam menentukan klasifikasi sentimen. *Naive Bayes Classifier* merupakan salah satu algoritma yang sering digunakan dalam melakukan klasifikasi data dalam jumlah yang besar. Klasifikasi yang dilakukan pada penelitian ini berdasar dari teorema *Bayes* [5] dengan mencari terlebih dahulu nilai probabilitas tertinggi pada proses klasifikasi data *testing* sehingga dihasilkan kategori yang paling tepat berdasarkan nilai probabilitas tertinggi tersebut [12].

Tahapan penelitian yang dilakukan mulai dari pengumpulan data, *labeling*, *preprocessing*, metode *Naive Bayes*, dan *confusion matrix* seperti gambar berikut.



Gambar 1 Tahapan Penelitian

1) Pengumpulan Data

Tahap pengumpulan data pada penelitian dilakukan melalui *crawling* data pada media sosial *Twitter* menggunakan *Netlytic*. Proses ini membutuhkan akun *Twitter* yang akan disambungkan dengan *Netlytic* menggunakan API (*Application Programming Interface*) sebagai penghubung antar keduanya. Proses pengambilan data dilakukan dengan memasukkan kata kunci pada kolom pencarian *Netlytic* untuk menampilkan *dataset*. Kata kunci yang digunakan dalam mengumpulkan *dataset* pada penelitian ini diantaranya piala dunia u-20 2023, pildun u-20 2023, sanksi FIFA, pembatalan FIFA, Indonesia gagal jadi tuan rumah, dan tuan rumah piala dunia u-20. Data yang berhasil dikumpulkan sebanyak 9000 data *tweet* yang terdiri dari *retweet*, *original*, *mention*, *reply*, dan *quote* dalam bahasa Indonesia pada media sosial *Twitter*. *Database* yang telah diperoleh dapat disimpan dengan format *excel* dan *csv*. Selanjutnya *dataset* akan dibersihkan, diproses, dan dianalisa menggunakan *tools Rapidminer*. Total keseluruhan data yang digunakan setelah melalui proses penghapusan data yang duplikat yaitu 2800 data *tweet* yang siap diproses ke tahap selanjutnya.

2) Labeling

Setelah data berhasil dikumpulkan, kemudian data dibagi menjadi dua yaitu data *training* dan data *testing* yang masing – masing 1000 data dan 1800 data. Pada penelitian ini data *training* akan diberi label dengan sentimen positif dan sentimen negatif yang telah dilakukan secara manual. Data *training* yang sudah diberi label tersebut akan digunakan *Naive Bayes* sebagai data *training* untuk membangun model klasifikasi yang nantinya akan digunakan untuk melakukan *labeling* secara otomatis pada data *testing*.

3) Preprocessing

Tahapan *preprocessing* pada tahap ini sangat penting karena sangat berpengaruh dalam menentukan kualitas model *Naive Bayes* yang dihasilkan saat mengimplementasikan algoritma *Naive Bayes* [13]. *Preprocessing* data pada penelitian ini dilakukan untuk menghilangkan *noise* sehingga menghasilkan data yang bersih untuk memudahkan proses pada sistem [14]. Untuk menghasilkan data yang bersih, maka harus dilakukan *cleaning* data seperti *replace duplicate*, *replace emoticon*, *replace URL*, *replace username (tag)*, *replace hashtag* dan *replace retweet*. Pada penelitian ini, proses *replace duplicate* dan *replace emoticon* dilakukan menggunakan *excel*, sedangkan proses *replace URL*, *replace username (tag)*, *replace hashtag* dan *replace retweet* dilakukan menggunakan *Rapidminer* dengan metode TF-IDF. Metode ini digunakan untuk menghitung frekuensi kemunculan dan prioritas dari suatu kata pada data yang digunakan [15].

Tahapan *preprocessing* dokumen sebagai berikut.

a). *Tokenize*

Tahap ini dilakukan untuk memisahkan satu kata dengan kata lainnya menjadi satu atribut tunggal/tersendiri sehingga dapat dibedakan [16].

b). *Transform Cases*

Tahap ini merupakan tahapan *preprocessing* untuk menyamakan semua kata ke dalam huruf kecil (*lowercase*).

c). *Filter Stopwords*

Tahap ini dilakukan untuk menghilangkan kosa kata yang tidak penting serta tidak digunakan yang hanya sebagai kata penghubung pada data yang digunakan untuk memudahkan proses klasifikasi [16]. Dalam menghilangkan kosa kata tersebut, dibutuhkan metode *stopwords* dengan bantuan *stoplist* atau *wordslit* yang sudah disiapkan.

d). *Filter Tokens*

Pada tahap ini juga menghilangkan kosa kata yang tidak penting dengan menggunakan parameter untuk memfilter token berdasarkan panjang karakter yang dimasukkan.

4) Metode *Naive Bayes*

Setelah mendapatkan data yang bersih melalui *preprocessing* data, selanjutnya penerapan algoritma *Naive Bayes* untuk membangun model klasifikasi *Naive Bayes* berdasarkan data *training* yang telah diberi label yang akan digunakan untuk menentukan sentimen *tweet* ke dalam kategori positif dan negatif pada data *testing*. Klasifikasi *Naive Bayes* merupakan klasifikasi statistik yang mampu melakukan prediksi peluang/probabilitas terhadap keanggotaan kelas suatu data ke dalam kelas tertentu [13]. Klasifikasi *Naive Bayes* dihitung dengan persamaan 1 [5].

$$P(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (1)$$

Keterangan:

B = Data sampel yang telah diketahui kelasnya

A = Hipotesa yang menentukan bahwa B masuk ke dalam kelas C

$p(A|B)$ = Probabilitas yang menunjukkan bahwa hipotesa B terbukti diperoleh ke dalam C

$p(B|A)$ = Probabilitas data sample B bila A dikondisikan dikondisikan pada B

$p(A)$ = Probabilitas dari hipotesa A

$p(B)$ = Probabilitas data sample B

5) *Confusion Matrix*

Setelah memperoleh model *Naive Bayes* berdasarkan data *training* yang telah diberi label pada tahap sebelumnya, selanjutnya akan dilakukan evaluasi dari model klasifikasi *Naive Bayes*. Evaluasi tersebut dilakukan dengan pengujian pada data *testing* dalam menentukan klasifikasi sentimen positif atau negatif secara otomatis berdasarkan model Klasifikasi *Naive Bayes* dari data *training* untuk menentukan keakuratan hasil. Selanjutnya mengukur *performance* dari klasifikasi model *Naive Bayes* yang telah dibangun menggunakan *confusion matrix* [15] untuk mengetahui *accuracy*, *precision*, dan *recall*. *Accuracy* merupakan pengukuran tingkat kedekatan

hasil pengukuran dengan dengan nilai yang sebenarnya (fakta), pada persamaan 2 [17].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{2}$$

Precision merupakan pengukuran tingkat ketetapan dari informasi yang diberikan oleh user dengan informasi yang diminta oleh pengguna, pada persamaan 3 [17].

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{3}$$

Recall merupakan tingkat keberhasilan dalam pengambilan data yang relevan dari sistem, pada persamaan 4 [17].

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{4}$$

Penelitian ini menggunakan *dataset* yang telah diperoleh dari *Netlytic* sebagai objek pengujian dengan menggunakan data *testing* sebanyak 1800 data *tweet* untuk menentukan sentimen positif dan negatif menggunakan metode *Naive Bayes* berdasarkan data *training* yang telah dipelajari sebelumnya (sebagai data latihan) yang kemudian hasilnya akan dievaluasi menggunakan *confusion matrix* untuk menghasilkan nilai akurasi dari algoritma yang digunakan.

4. Hasil dan Pembahasan

Berdasarkan *tweet* hasil *crawling* data yang diperoleh menunjukkan bahwa masyarakat berfokus pada gagalnya Indonesia menjadi tuan rumah pada Piala dunia U-20 dan tak sedikit juga masyarakat yang berfokus pada tindakan ketua PSSI yang mampu bergerak cepat mengambil tindakan mengenai peristiwa tersebut.

DATE	USER	POSTS, N = 31
2023-05-04	aduanfemad	@diedalame Mungkin sebagai ganti buat Indonesia yang gagal jadi tuan rumah dunia keren.
2023-05-04	Dan930	@diedalame Ini agenda FIFA cuna buat menghibur fans sepak bola Indonesia yang gagal jadi Tuan rumah pola dunia.
2023-05-05	iam_rnans	@gopran10 @HelaLuhro @dianpfitro @garagarabola. Beda level dan beda situasi lah bro. Anal lu lau, sejak 2013. Lum IS lebih banyak gagal juara tho dibandingkan jadi juara untuk di level AFF. Dan sejak 2013, IS cuma bisa di level AFF dsngn ilan di AFC (jika Indonesia bukan tuan rumah). Indonesia selalu gagal di AFC U-23.
2023-05-05	Itulter40551058	@gatoreshul Mungkin ini rasa leraa kasih Argentina buat Indonesia karena Indonesia gagal jadi tuan rumah piala U-20 akhirnya tuan rumah jatuh ke Argentina 🤔
2023-05-05	valenagi	@dteobalme Salas Budi Indonesia gagal jadi tuan rumah 🤔
2023-05-05	ahmashulaa	@dteobalme Konopras Indonesia gagal jadi tuan rumah dem leras error argentina ke Indonesia sebagai gantinya tuan rumah piala di kasih ke argentina 🤔
2023-05-05	SINDOnews	Glorious Karya Weird Genius Tetap Jadi Lagu Resmi Piala Dunia U-20 di Argentina meski Indonesia gagal Jadi tuan rumah #Indonesia #News https://t.co/XU5a4f0BjR
2023-05-05	SINDOnews	Indonesia Gagal Jadi Tuan Rumah, Glorious Karya Weird Genius Tetap Jadi Lagu Resmi Piala Dunia U-20 di Argentina #BeritaRasa #KabarBeritaRasa #Indonesia #News https://t.co/XU5a4f0BjR
2023-05-05	wan@koptar09	@Nipateh@alabola Ini Argentina mau lawan Indonesia sebagai ucapan fermakasih lera gara? Indonesia gagal jadi tuan piala u20, gantinya Argentina yg jadi tuan rumah.
2023-05-05	achmadrofaid	@Nipateh@alabola Mungkin ini sebagai penghibur masyarakat Indonesia karena gagal jadi tuan rumah piala dunia U20 yang pindah ke Argentina.

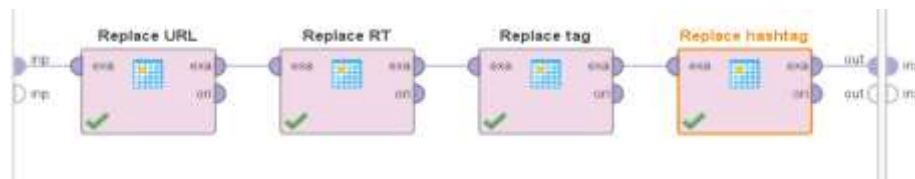
Gambar 2 Hasil *Crawling* data menggunakan *Netlytic*

Dataset yang berhasil dikumpulkan sebelumnya akan dilakukan penghapusan kolom yang tidak dibutuhkan dengan menyisakan *description* yang berisi *tweet* sebelum dilakukan proses *labeling* dan *preprocessing* dokumen.

Negatif	RT @Box2BoxBola: Inilah keluh kesah kami soal Indonesia gagal jadi tuan rumah Piala Dunia U20. Mulai dari Parpol sampai ke hal melobi FIFA...
Positif	RT @DM_N4BPL: Alhamdulillah, keberhasilan mas Ketum PSSI Erick Thohir, Indonesia terhindar dari sanksi berat FIFA namun justru malah berhas...

Gambar 3 Contoh Data *Labeling*

Setelah proses *labeling*, selanjutnya data tersebut dimasukkan ke *Rapidminer* dimana pada filter *examples* menggunakan parameter “*is not missing*” untuk data *training* yang sudah diberi label. Tahapan *cleaning* data yang dilakukan *cleaning* data seperti *replace URL*, *replace username (tag)*, *replace hashtag* dan *replace retweet (RT)* seperti pada gambar berikut.



Gambar 4 *Cleaning* data

Setelah dilakukan *cleaning* data, selanjutnya akan dilakukan *pre – processing* dokumen dengan metode TF-IDF berikut.



Gambar 5 *Preprocessing Document*

Preprocessing dokumen tersebut menghasilkan data yang berhasil dipecah kata per kata menjadi (token) dengan jumlah 2.696 atribut yang berhasil di ekstrak atau dipecah.

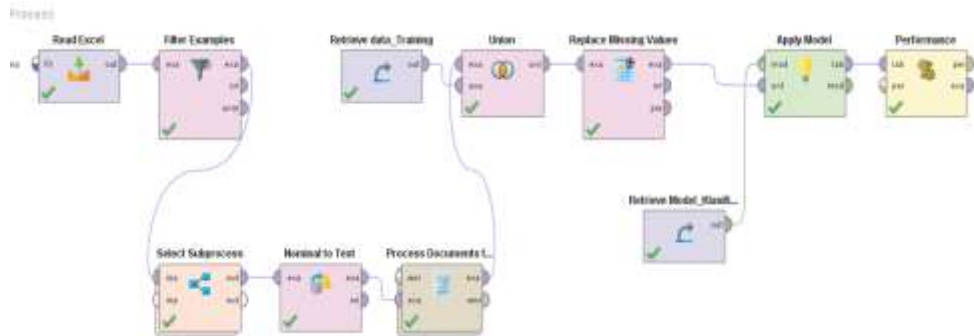
Row N...	text	absikan	acaranya	acungan	acungi	adain	adem	adik
1	kalo sebelum ti...	0	0	0	0	0	0	0
2	Indonesia ga...	0	0	0	0	0	0	0
3	berta indone...	0	0	0	0	0	0	0
4	Indonesia ga...	0	0	0	0	0	0	0
5	Indonesia ga...	0	0	0	0	0	0	0
6	Indonesia ga...	0	0	0	0	0	0	0
7	wapres indon...	0	0	0	0	0	0	0
8	Indonesia ga...	0	0	0	0	0	0	0
9	partisi pssi...	0	0	0	0	0	0	0
10	Indonesia ga...	0	0	0	0	0	0	0
11	wapres indon...	0	0	0	0	0	0	0
12	Indonesia ga...	0	0	0	0	0	0	0
13	wapres indon...	0	0	0	0	0	0	0

ExampleSet (1.800 examples, 1 special attribute, 2.696 regular attributes)

Gambar 6 Hasil *Preprocessing* dokumen

Tahap selanjutnya adalah mengimplementasikan *Naive Bayes* untuk membangun model klasifikasi berdasarkan data *training* yang sudah diberi label yang nantinya akan digunakan untuk melakukan *labeling* secara otomatis pada pengujian data *testing*.

Model klasifikasi *Naive Bayes* yang telah diperoleh akan evaluasi seberapa akurat model tersebut dengan menghitung *accuracy*, *precision* dan *recall*. Tahap ini dilakukan dengan menggunakan data *testing*, pada filter *example* parameter “*is missing*” untuk data *testing* yang belum diberi label. Data *testing* yang dimasukkan sudah sudah bersih dan siap untuk di *preprocessing*.



Gambar 7 Evaluasi *Confusion Matrix*

Hasil evaluasi disajikan dengan tabel *confusion matrix* pada gambar berikut.

accuracy: 96.00%

	true Positif	true Negatif	class precision
pred. Positif	678	37	94.83%
pred. Negatif	3	282	98.95%
class recall	99.56%	88.40%	

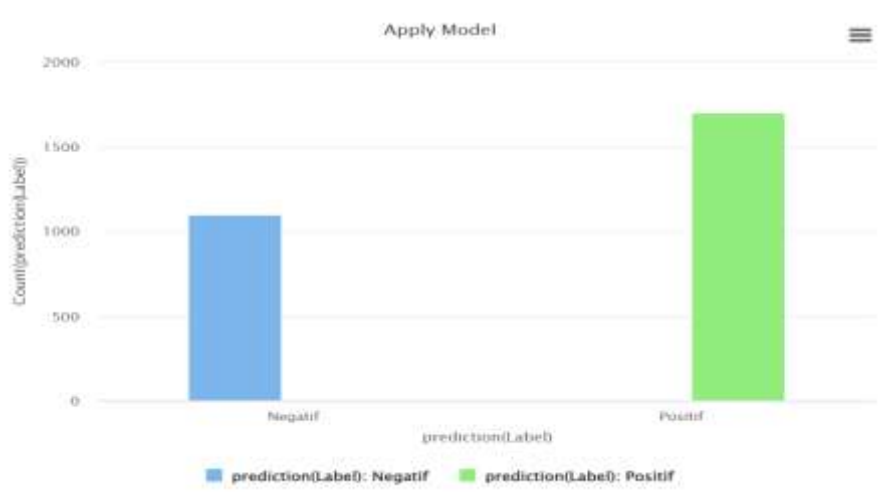
Gambar 8 Hasil *Confusion Matrix*

Confusion Matrix menampilkan seberapa banyak nilai dari *true positive*, *true negative*, *false positive* dan *false negative* dari data *testing* yang telah diuji. Pada hasil pengujian *confusion matrix* akan menghitung tingkat *accuracy*, *precision*, dan *recall*. Pada tabel *confusion matrix* tersebut terbagi menjadi beberapa bagian, kolom *true positive* (TP) yaitu sampel yang diprediksi bernilai positif dan hasilnya memang benar bernilai positif sebanyak 678, *true negative* (TN) yaitu sampel yang diprediksi bernilai negatif dan memang benar bernilai negatif sebanyak 282, *false positive* (FP) yaitu sampel diprediksi bernilai positif tetapi hasilnya bernilai negatif (*type I error*) sebanyak 37, dan *false negative* (FN) merupakan sampel yang diprediksi bernilai negatif namun hasilnya bernilai positif (*type II error*) sebanyak 3.

Berdasarkan hasil pengujian yang telah dilakukan menggunakan *confusion matrix*, diperoleh nilai *accuracy* sebesar 96%, nilai sebesar *precision* = 94.83%, dan nilai *recall* sebesar 99.56%. Jika dilakukan perhitungan secara manual, maka hasilnya akan sama dengan hasil yang diberikan oleh *Rapidminer*. Perhitungan *accuracy*, *precision*, dan *recall* sebagai berikut.

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{678 + 282}{678 + 282 + 37 + 3} \times 100\% \\
 &= \frac{960}{1000} \times 100\%
 \end{aligned}$$

$$\begin{aligned}
 &= 96\% \\
 \text{Precision} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{678}{678+37} \times 100\% \\
 &= \frac{678}{717} \times 100\% \\
 &= 94,83\% \\
 \text{Recall} &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{678}{678+3} \times 100\% \\
 &= \frac{678}{681} \times 100\% \\
 &= 99,56\%
 \end{aligned}$$



Gambar 9 Visualisasi sentimen positif dan negatif

Visualisasi menampilkan perbedaan sentimen yang bernilai positif dan sentimen yang bernilai negatif. Dari grafik yang diperoleh pada *Rapidminer* menunjukkan bahwa terdapat 1.702 *tweet* yang bernilai positif dan 1.098 *tweet* yang bernilai negatif. Hasil visualisasi tersebut menunjukkan perbedaan sentimen yang cukup signifikan yang didominasi oleh sentimen yang bernilai positif, sehingga dapat dikatakan bahwa sebagian besar respon dari pengguna *Twitter* tersebut setuju terhadap kebijakan yang telah diambil oleh pemerintah dengan berbagai pertimbangan. Dari hasil sentimen yang telah diperoleh tak sedikit juga yang memberikan tanggapan negatif terhadap keputusan tersebut. Oleh karena itu, sebagai saran bagi pengambil kebijakan/pemerintah kedepannya untuk lebih bijak lagi dengan mempertimbangkan segala konsekuensi yang akan ditimbulkan serta evaluasi berkelanjutan guna menilai dampak dari kebijakan yang telah diambil berdasarkan hasil sentimen. Hal ini akan membantu upaya perbaikan berkelanjutan untuk acara serupa yang akan diadakan di Indonesia kedepannya.



Gambar 10 Tampilan *Worldcloud*

Gambar diatas merupakan tampilan visualisasi teks yang menunjukkan frekuensi kemunculan setiap kata, dimana semakin besar teks berarti kata tersebut sering muncul dan begitupun sebaliknya, semakin kecil teks berarti kata tersebut tidak sering muncul dalam *tweet* [17].

Dilihat dari hasil nilai akurasi yang diperoleh dari evaluasi *confusion matrix* pada Gambar 8 menunjukkan bahwa penggunaan algoritma *Naive Bayes* dalam penelitian ini sangat bagus sehingga dapat direkomendasikan untuk penelitian selanjutnya terutama pada analisis sentimen.

5. Simpulan

Berdasarkan hasil penelitian yang telah diperoleh pada penelitian terhadap analisis sentimen perubahan piala dunia 2023 melalui *Twitter* dengan keseluruhan data yang digunakan 2800 data yang dibagi menjadi data *training* 1000 data dan data *testing* 1800 data, dapat disimpulkan bahwa penggunaan metode *Naive Bayes* menghasilkan tingkat akurasi yang cukup baik dilihat dari hasil evaluasi *performance confusion matrix* yang cukup tinggi. Dimana diperoleh nilai *accuracy*, *precision*, dan *recall* masing – masing adalah 96%, 94,83%, dan 99,56%. Jumlah data *training* yang digunakan akan berpengaruh terhadap hasil kinerja dari sistem, semakin banyak data *training* yang digunakan dengan pelabelan yang baik maka semakin baik hasil kinerja dari sistem. Hasil visualisasi oleh sistem menunjukkan bahwa sebagian besar respon dari pengguna *Twitter* memberikan tanggapan yang positif (setuju) terhadap gagalnya Indonesia menjadi tuan rumah Piala Dunia U-20 serta menerima hal tersebut dengan berbagai pertimbangan. Dilihat dari nilai *accuracy* yang diperoleh cukup tinggi menandakan bahwa semakin baik model klasifikasi yang dihasilkan. Berdasarkan penelitian yang telah dilakukan tersebut, maka disarankan untuk penelitian selanjutnya bahwa perlu diperhatikan dalam melakukan pelabelan pada data *training* harus dilakukan sebaik mungkin sehingga menghasilkan model klasifikasi *Naive Bayes* yang baik karena berpengaruh pada proses *labeling* otomatis pada data *testing* dan *performance* model yang dihasilkan oleh sistem.

Daftar Pustaka

- [1] FIFA, “*FIFA Removes Indonesia as Host of FIFA U-20 World Cup 2023™*”, Fifa, 29 Maret 2023. [Online]. Tersedia: <https://www.fifa.com/about-fifa/organisation/media-releases/fifa-removes-indonesia-as-host-of-fifa-u-20-world-cup-2023-tm> [Diakses: 23 April 2023].
- [2] CNN Indonesia, “*FIFA Removes Indonesia as Host of FIFA U-20 World Cup 2023™*”, CNN Indonesia, 29 Maret 2023. [Online]. Tersedia: <https://www.cnnindonesia.com/olahraga/20230329170303-142-930948/fifa-indonesia-batal-jadi-tuan-rumah-piala-dunia-u-20-2023> [Diakses: 23 April 2023].

- [3] H. Gunwan, "Dampak Besar Batalnya Indonesia Jadi Tuan Rumah Piala Dunia U-20 Secara Ekonomi", *Tribun News*, 3 April 2023. [Online]. Tersedia: <https://www.tribunnews.com/bisnis/2023/04/03/dampak-besar-batalnya-indonesia-jadi-tuan-rumah-piala-dunia-u-20-secara-ekonomi> [Diakses: 11 September 2023].
- [4] Y. Assidiq, "Dampak Pembatalan Piala Dunia U-20 Terhadap Politik Dalam Negeri dan Luar Negeri", *ReJogja Republika*, 4 April 2023, [Online]. Tersedia: <https://rejogja.republika.co.id/berita/rsku63399/dampak-pembatalan-piala-dunia-u20-terhadap-politik-dalam-dan-luar-negeri> [Diakses: 11 September 2023].
- [5] F. Ratnawati, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Opini Film Pada *Twitter*," *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 3, no. 1, pp.50-59, 2018.
- [6] M. A. Rizatya, "Pengguna *Twitter* di Indonesia Capai 18,45 Juta pada 2022", *DataIndonesia* Tersedia: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022> [Diakses: April. 11, 2023].
- [7] D. Darwis, N. Siskawati, & Z. Abidin, "Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen Review Data *Twitter* BMKG Nasional," *Jurnal TEKNO KOMPAK*, vol. 15, no. 1, pp.131–145, 2021.
- [8] T. Taufiq, & Y. Yudihartanti, "Penerapan Theorema Bayes Pada Penilaian Kelayakan Angkutan Kota". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 111-122, 2021.
- [9] D. Putri Yani, S. Kurnia Gusti, F. Yanto, & M. Affandes, "Klasifikasi Sentimen Transformasi dan Reformasi Sepak Bola Indonesia Pada *Twitter* Menggunakan Algoritma *Bernoulli Naïve Bayes*," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol.4, no. 3, pp.451–58, 2023. DOI: 10.30865/json.v4i3.5829.
- [10] S. Lestari, S. Saepudin, "Analisis Sentimen Vaksin *Sinovac* Pada *Twitter* Menggunakan Algoritma *Naive Bayes*," *SISTEMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, pp.163-170, 2021.
- [11] A. F. Yudianto, Statistwaty, & N. Ransi, "*Naive Bayes Classifier* Terhadap *Sentiment Analysis* Mengenai PSSI Setelah Tragedi Kanjuruhan Malang," *Jurnal ANIMTOR*, vol. 1, no. 2, pp.12-19, 2023.
- [12] R. Feldan, J. Sanger, *The Text Mining Handbook*, New York: Cambridge University Press.
- [13] F. Handayani, F. Setio Pribadi, "Implementasi Algoritma *Naive Bayes Classifier* dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *Jurnal Teknik Elektro*, vol. 7, no.1, pp.19–24, 2015.
- [14] L. Adiani, H. Sujaini, & Tursina, "Implementasi *Sentiment Analysis* Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, no. 2, pp.183–190, 2020. DOI: 10.26418/justin.v8i2.36776.
- [15] N. Ikhbar Wibowo, T. Andika Maulana, H. Muhammad, & N. Aini Rakhmawati, "Perbandingan Algoritma Klasifikasi Sentimen *Twitter* Terhadap Insiden Kebocoran Data *Tokopedia*," *JISKa*, vol. 6, no.2, pp.120–129, 2021.
- [16] R. Sholehurrohman, I. Sabda Ilman, "Analisis Sentimen *Tweet* Kasus Kebocoran Data Penggunaan *Facebook* oleh *Cambridge Analytica*," *Jurnal*, vol. 3, no. 1, pp.140-147, 2022.
- [17] A. Turmudi Zy, A. Nugroho, A. Rivaldi, & I. Afriantoro, "Analisis Sentimen Terhadap Pembobolan Data pada *Twitter* dengan Algoritma *Naive Bayes*," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 2, pp.202–213,2022, doi: 10.37012/jtik.v8i2.1240.