

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)

ANALISIS SENTIMEN TERHADAP PRO DAN KONTRA BATALNYA INDONESIA MENJADI TUAN RUMAH PIALA DUNIA U20 PADA MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER (NBC)

Septian Bayu Adi

Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Hasyim Asy'ari Tebuireng Jombang
Alamat e-mail: adibay1910@gmail.com

Aries Dwi Indriyanti

Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Hasyim Asy'ari Tebuireng Jombang
Alamat e-mail: ariesdwi@unesa.ac.id

Abstrak

Piala Dunia U20 FIFA, yang sebelumnya dikenal sebagai Piala Dunia Pemuda FIFA, adalah kejuaraan sepak bola resmi yang diadakan oleh FIFA setiap dua tahun sekali. Kejuaraan ini diikuti oleh setiap negara yang mengirimkan tim nasional sepak bola berusia di bawah 20 tahun. Banyak negara berlomba untuk menjadi tuan rumah Piala Dunia U20, meskipun biaya penyelenggaraannya sangat mahal karena keuntungan yang didapatkan oleh tuan rumah, terlebih lagi dalam bidang ekonomi. Indonesia ditunjuk sebagai tuan rumah piala dunia u20 pada tanggal 24 Oktober 2019 lalu dibatalkan pada 29 maret 2023 karena itu dengan banyaknya opini dari masyarakat tentang dibatalkannya Indonesia menjadi tuan rumah sehingga dapat dilakukannya penelitian terkait hal tersebut. Penelitian ini bertujuan untuk mengetahui apakah masyarakat lebih pro atau kontra terhadap batalnya piala dunia u20 di indonesia menggunakan algoritma Naive Bayes. Data yang digunakan sejumlah 1150 data komentar twitter berdasarkan dari hasil perhitungan Algoritma *Naive Bayes* Studi ini menemukan bahwa tweet dengan sentimen kontra berjumlah 62,67% dan untuk tweet sentimen pro berjumlah 37,33%. Dengan nilai akurasi rata-rata 73,7%, nilai presisi 73,7%, dan nilai recall 82,5%, dapat disimpulkan bahwa mayoritas orang Indonesia yang menggunakan media sosial Twitter memiliki kecenderungan untuk menentang batalnya menjadi tuan rumah piala dunia u20.

Keywords: Piala Dunia, Classification, Naive Bayes

Abstract

The FIFA U20 World Cup, formerly known as the FIFA Youth World Cup, is an official football championship held by FIFA every two years. This championship is participated by every country that sends a national soccer team under the age of 20. Many countries are vying to host the U20 World Cup, even though it is very expensive to host due to the advantage that the hosts get. especially in the economic field. Indonesia was appointed as the host of the u20 world cup on October 24 2019 and then canceled on March 29 2023 because of that, with the abundance of public opinion about Indonesia being canceled as host, it can be used as research material to find information. This study aims to find out whether the public is more pro or con against the cancelation of the U20 World Cup in Indonesia using the Naive Bayes algorithm. The data used is 1150 Twitter comment data based on the calculation results of the Naive Bayes Algorithm. This study found that tweets with contra sentiment amounted to 62.67% and for pro sentiment tweets amounted to 37.33%. With an average accuracy value of 73.7%, a precision value of 73.7%, and a recall value of 82.5%, it can be won by Indonesians using social media Twitter who have a tendency to cancel the fight to host the u20 world cup.

Keywords: World Cup, Classification, Naive Bayes

PENDAHULUAN

Salah satu aktivitas paling penting untuk dilakukan secara rutin dalam kehidupan manusia adalah olahraga, yang dapat mengharumkan nama negara dan daerah melalui kompetisi nasional, regional, dan internasional. Prestasi olahraga khususnya olahraga sepakbola dapat meningkatkan citra negara di mata dunia,, semua negara berlomba-lomba untuk menjadi yang terbaik. (Irwanto & Muslimah Zahro Romas, 2017).

Piala Dunia U20 FIFA, sebelumnya dikenal sebagai Piala Dunia Pemuda FIFA, adalah turnamen sepak bola resmi yang diselenggarakan oleh FIFA setiap dua tahun sekali. FIFA diikuti berbagai negara dengan mendegalisasikan tim sepakbola nasionalnya yang berusia dibawah 20 tahun untuk mewakili negaranya dengan syarat melalui kualifikasi. Piala Dunia U20 FIFA pertama kali diperkenalkan dan masih Bernama Kejuaraan khusus Dunia Remaja yang diikuti 76 Negara. Diantara 76 negara yang mengikuti, beberapa negara bergantian menjadi tuan rumah piala dunia U20 Meskipun penyelenggaraan

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)

Piala Dunia sangat mahal, banyak negara berlomba untuk menjadi tuan rumah. dikarenakan banyak sekali manfaat atau kelebihan yang didapatkan oleh negara tersebut ketika menjadi tuan rumah terlebih lagi dalam bidang ekonomi. (Aryani, dkk, 2023). Salah satu ajang olahraga yang ramai diperbincangkan masyarakat dunia tahun ini adalah tentang batalnya piala dunia U20 di indonesia.

Adanya urgensi terkait batalnya piala dunia di indonesia yang mengundang banyak kontroversi di twitter baik pro maupun kontra menjadi suatu hal yang harus diketahui oleh publik. Dengan adanya penelitian ini dapat mengetahui pendapat publik di media sosial twitter mengenai batal nya Indonesia menjadi tuan rumah piala dunia U20 apakah cenderung pro atau kontra. Oleh karena itu, penulis mengangkat tema penelitian dalam bentuk skripsi dengan judul “Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)”.

METODE

Tahapan yang dilakukan dalam penelitian ini digambarkan dalam flowchart di bawah ini:



Gambar 1. Tahapan Alur Penelitian (flowchart)

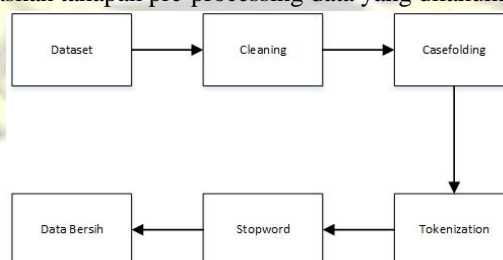
a. Pengumpulan Data

Pengumpulan data adalah langkah pertama dalam analisis sentimen. Twitter memungkinkan pengembang menggunakan API (Application Programming Interface) untuk mengambil data dari media sosialnya.

Setelah melakukan pengumpulan data, kemudian data tersebut diberi label (pelabelan). Pelabelan di sini bertujuan untuk memudahkan pemrosesan program sistem selama proses klasifikasi. Label sentimen yang digunakan adalah sentimen pro dan sentimen kontra. Setelah melakukan proses pelabelan data dibagi menjadi 2 bagian, yaitu data training atau dikenal dengan data latih dan data testing atau yang dikenal dengan data uji. Data tersebut memiliki kegunaannya masing-masing, data latih berfungsi untuk melatih sistem untuk mengenali kata-kata sesuai sentimen sedangkan data uji berfungsi sebagai data yang akan diprediksi oleh sistem untuk mengetahui tingkat keberhasilan sistem dengan menguji hasil pelatihan

b. Pre-Processing

Setelah melakukan pengumpulan dan pelabelan data, langkah yang harus dilakukan selanjutnya adalah preprocessing. Tahapan ini yaitu menyiapkan data agar menjadi data yang siap untuk dilakukan analisis.. Ada beberapa tahap dalam preprocessing pada Gambar 2 menjelaskan tahapan pre-processing data yang dilakukan pada penelitian ini.



Gambar 2. Tahapan Pre-Processing Data

1. Dataset

Setelah melakukan pencarian data, semua data dikumpulkan dan berfungsi untuk dilakukan nya *pre-processing* data seperti *cleaning*, *casefolding*, *stopword*, *tokenization*.

2. Cleaning

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode *Naive Bayes Classifier* (NBC)

Cleaning adalah tahap pertama untuk melakukan preprocessing data. Data yang didapat dari proses crawling merupakan data mentah sehingga memerlukan proses cleaning untuk membuang ataupun menghilangkan data yang tidak lengkap, relevan dan juga akurat. Data yang tidak lengkap dan relevan lebih baik dibuang. cleaning juga akan mempengaruhi performa dari teknik *data mining* karena data karena data akan berkurang secara jumlah dan kompleksitasnya. Adapun perbedaan dari menggunakan data yang dibersihkan dan tidak yaitu, data yang dibersihkan memiliki akurasi yang lebih tinggi dibandingkan dengan data asli yang tidak dibersihkan.

3. *Case Folding*

Case folding merupakan tahapan untuk mengubah huruf kapital atau huruf besar pada semua dokumen Tweet menjadi huruf kecil. Case Folding bertujuan untuk menghapus data yang memiliki perbedaan pada besar kecilnya huruf. Gambar 3 menampilkan flowchart pada tahapan case folding.



Gambar 3 Diagram Alir *Case Folding*

4. *Tokenization*

Tahap *tokenizing* terdiri dari pemotongan kata berdasarkan setiap kata dari satu kalimat. Proses tokenisasi menghasilkan kata tunggal yang nantinya akan masuk ke dalam database untuk memudahkan proses pembobotan kata karena kata-kata dalam dokumen tersebut dipisahkan oleh spasi. Gambar 4 menunjukkan flowchart atau alur proses *tokenization*.



Gambar 4 Diagram Alir *Tokenizing*

5. *Stopword*

Stopword berfungsi untuk menghapus kata-kata yang kurang atau tidak memiliki makna pada dokumen tweet. Kata-kata tersebut tidak ikut andil dalam keakuratan klasifikasi pada data sentiment. Tahap stopword removal dilakukan. Database kata stopword berisi kata yang akan dihilangkan. Jika ada kata yang sesuai dalam dokumen Tweet, kata tersebut akan dibuang dan akan dikosongkan dengan kata lain akan digantikan dengan spasi. Gambar 5 menunjukkan flowchart untuk penghapusan stopword.

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)



Gambar 5 Diagram Alir Stopword Removal

c. Klasifikasi Naive Bayes

Metode yang digunakan untuk mengklasifikasikan data twitter untuk mendapatkan hasil klasifikasi pro atau kontra ialah Naive Bayes. Setelah melakukan pre-processing data, data melakukan training dengan jumlah data yang lebih besar dibandingkan dengan data uji atau data testing. Pada pembagian dataset tergantung kebutuhan. (Id, 2021:106).

Jumlah data training yang dipakai pada penelitian ini adalah 80% dari jumlah data yang ada. Pada proses ini menggunakan bantuan Python dan melakukan klasifikasi data dengan algoritma Naive Bayes.

Persamaan (1) Teorema Naive Bayes sebagai berikut:

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)} \quad (1)$$

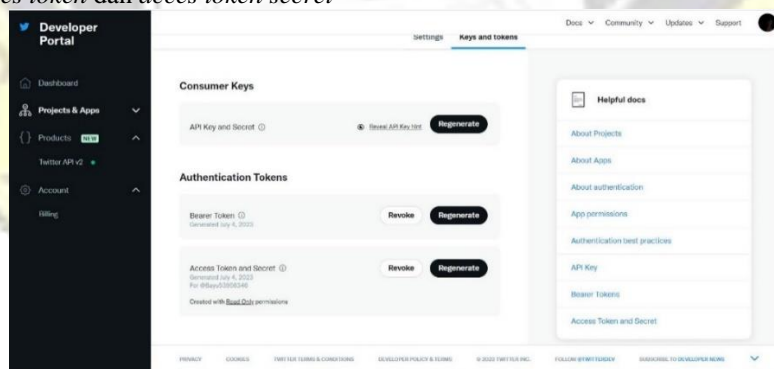
d. Evaluasi performa klasifikasi

Selain itu, klasifikasi teks membutuhkan penilaian metode yang tepat. Penelitian ini menggunakan klasifikasi multikelas, yang mencakup klasifikasi positif dan negative atau dalam penelitian ini dibagi dengan sentiment pro dan sentiment kontra. Akurasi klasifikasi multikelas hanya diukur pada kinerja pengklasifikasi data dari seluruh kelas. Metode diuji dengan menghitung akurasi, *Precision*, dan *Recall*.

HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Untuk dapat mengambil sebuah data dari media sosial Twitter, langkah pertama yaitu mendaftar sebagai Developer di sosial media Twitter dan mengisi semua form yang ditunjukkan. Jika semua proses pendaftaran selesai, anda akan diberikan *consumer key*, *consumer secret*, *access token* dan *access token secret* akan diberikan. Ini akan digunakan untuk mengakses data di Twitter. Gambar 6 menunjukkan halaman yang menunjukkan kunci konsumen, diberikan *consumer key*, *consumer secret*, *access token* dan *access token secret*



Gambar 6 Akses Api Key

Data yang berhasil dikumpulkan dianggap sebagai data yang relevan dengan topik. “Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20” pada media sosial Twitter. Data yang dikumpulkan menggunakan Bahasa Indonesia dan berjumlah 1150 data, data diberikan pelabelan secara manual untuk menjadi data training di penelitian ini, data di labelkan menjadi kategori “pro” dan “kontra”. Berikut hasil dari data yang telah dikumpulkan dalam format xls.

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)

1	text
2	Exco PSSI, Arya Sinulingga membantah kabar Piala Dunia U-20 2023 akan diselenggarakan di Peru. Dirinya mengatakan belum ada surat resmi dari FIFA
3	Indonesia Batal Tuan Rumah Piala Dunia U-20?, Pelatih Shin Tae-Yong Ikut Resah https://t.co/ftm8NxyE6
4	@Palliwest Lebih baik dipindah Piala Dunia U 20 dari pada menimbulkan keributan di Negeri ini,walau sanksi FIFA tdk mengikutkan PSSI ke ajang Intern
5	FIFA membatalkan drawing pembagian grup Piala Dunia U-20 yang seharusnya digelar di Bali pada (31/3) buntut sejumlah penolakan atas tim nasional
6	@Dennysregar7 Berarti tidak percaya komitmen TNI-Polri dalam menjaga keamanan dan ketertiban Piala Dunia U-20 di Indonesia ?
7	Caper, Baper, Tantrum dimana-mana, dan Cari Gara-gara.... Wali Kota Bengkulu Tolak Israel Bermain di Piala Dunia U-20 https://t.co/AWkd1T06N0 via
8	@Ekagud1 @SakitPe14248653 @Adi_8002 Yg standar ganda itu FIFA, aturan nya bisa melakukan pencoret Rusia ikut serta piala dunia di Qatar, tetapi i
9	Jusuf Kalla: Keikutsertaan Israel Piala Dunia U 20 di Indonesia Bisa Jadi Momentum Perdamaian https://t.co/bYPCz2dHAI
10	I Wayan Sukadana berharap Presiden Jokowi turun gunung" agar masalah Piala Dunia U-20 2023 terselesaikan. #U20WorldCupEMTEK #U20WorldCupK
11	Reaksi Shin Tae-yong, Drawing Piala Dunia U-20 Batal dan Nasib Timnas Indonesia U-20 https://t.co/Q8RIGOAik tidak perlu gelisah Pak, itulah negeri k
12	Gus Wagub : Saya sepakat, bahwa penyelenggaraan (Piala Dunia U-20) ini tetap bisa berjalan tanpa kita mengorbankan komitmen mewujudkan kemerc
13	Meski Drawing Batal, FIFA Tetap Lanjutkan Verifikasi Stadion Piala Dunia U-20 2023 di Indonesia https://t.co/dakLhXntlc
14	la sepakat dan mendukung penuh Gubernur Jateng, Ganjar Pranowo yang meminta panitia bersama pihak terkait agar mengupayakan langkah terobosa
15	Wakil Gubernur Jawa Tengah, mendukung penuh sikap Gubernur Jawa Tengah Ganjar Pranowo menolak kehadiran Timnas Israel dalam Piala Dunia U-2
16	Dukung Ganjar, Gus Yasin Juga Ingin Piala Dunia U-20 Tanpa Israel #GusYasin #PanglimaSGN #SantriGayengNusantara #TajYasin #WagubJateng @TajY
17	Shin Tae-yong Gelasah soal Nasib Piala Dunia U-20 2023, Merasa Sia-sia jika Batal https://t.co/353wGL1nH
18	Pelatih Timnas Indonesia Shin Tae-yong ikut was-was mendengar kabar tuan rumah Piala Dunia U-20 dipindahkan dari Indonesia. Acara drawing yang di
19	#TimnasIndonesia_U20 sekalian ya mau main SBG tuan rumah utk gelaran piala dunia u_20_(2023), eh musim tahun politik.
20	Saya berharap Piala Dunia U-20 2023 bisa berjalan lancar di Indonesia, Apalagi, Piala Dunia U-20 juga untuk kemajuan sepak bola Indonesia. Itu penting
21	Drawing Piala Dunia U-20 2023 Dibatalkan Timnas Indonesia U-20 Tetap Lanjutkan TC di Jakarta https://t.co/KgRu4QEOH
22	Respons Shin Tae Yong Soal Pembatalan Drawing Piala Dunia U-20 2023, Gelasah #ShinTaeYong https://t.co/CVcLJNfsh #PakvsAfghanistan
23	Bukan soal Piala Dunia U-20, Israel Kini Memanas Unjuk Rasa dimana-mana karena ini https://t.co/CVcLJNfsh #PakvsAfghanistan
24	1. Membunuh 135, sunofter.hola.2. Bembunohwa.hobac.3. 135, nuawaa.dilomakan.4. Berharan.masih.iadi.tuan.rumah.Piala.Dunia.U.20.1_HIBOKRIT.EASIK

Gambar 7. Dataset yang terkumpul

3.2 Pre-Processing Data

Tahapan ini bertujuan agar sistem komputer lebih mengenali bentuk data set. Selain itu, tahapan ini bertujuan untuk mengubah data yang tidak tersusun rapi menjadi data yang tersusun rapi. Dalam text preprocessing terdapat 4 proses yaitu cleaning, casefolding, tokenize, stopword.

a. Cleaning

Pada proses ini menghapus atribut yang tidak dibutuhkan dan tidak memiliki makna, seperti tanda baca, hastag, mention, reTweet, whitespace, dll. Berikut Gambar 8 adalah contoh cleaning pada salah satu data set.

text	sentimen	clean_tweet
Exco PSSI, Arya Sinulingga membantah kabar Piala Dunia U-20 2023 akan diselenggarakan di Peru. Dirinya mengatakan belum ada surat resmi dari FIFA mengenai pembatalan drawing Pildun U-20 2023 , apalagi pengurangan niat menggelar ajang tersebut di Indonesia #MostUpdate1058 https://t.co/GPHfMuQ5g	kontra	Exco PSSI Arya Sinulingga membantah kabar Piala Dunia U akan diselenggarakan di Peru Dirinya mengatakan belum ada surat resmi dari FIFA mengenai pembatalan drawing Pildun U apalagi pengurangan niat menggelar ajang tersebut di Indonesia MostUpdate
Indonesia Batal Tuan Rumah Piala Dunia U-20?, Pelatih Shin Tae-Yong Ikut Resah https://t.co/ftm8NxyE6	kontra	Indonesia Batal Tuan Rumah Piala Dunia U Pelatih Shin Tae Yong Ikut Resah
@Palliwest Lebih baik dipindah Piala Dunia U 20 dari pada menimbulkan keributan di Negeri ini,walau sanksi FIFA tdk mengikutkan PSSI ke ajang Internasional pasa berikutnya,kita harus siap	pro	Lebih baik dipindah Piala Dunia U dari pada menimbulkan keributan di Negeri ini walau sanksi FIFA tdk mengikutkan PSSI ke ajang Internasional pasa berikutnya kita harus siap
FIFA membatalkan drawing pembagian grup Piala Dunia U-20 yang seharusnya digelar di Bali pada (31/3) buntut sejumlah penolakan atas tim nasional Israel. Lebih lanjut, wartawan senior olahraga Indonesia Yesayas Oktavianus menyebut FIFA sudah membatalkan ... https://t.co/Q38rSG0WB	kontra	FIFA membatalkan drawing pembagian grup Piala Dunia U yang seharusnya digelar di Bali pada buntut sejumlah penolakan atas tim nasional Israel Lebih lanjut wartawan senior olahraga Indonesia Yesayas Oktavianus menyebut FIFA sudah membatalkan
@Dennysregar7 Berarti tidak percaya komitmen TNI-Polri dalam menjaga keamanan dan ketertiban Piala Dunia U-20 di Indonesia ?	kontra	Berarti tidak percaya komitmen TNI Polri dalam menjaga keamanan dan ketertiban Piala Dunia U di Indonesia

Gambar 8. Hasil Cleaning

b. Case Folding

Pada tahap ini bertujuan untuk mengubah semua kata menjadi huruf kecil. Ditunjukkan pada Gambar 9 untuk hasil tahapan casefolding

clean_tweet	hasil_casefolding
Exco PSSI Arya Sinulingga membantah kabar Piala Dunia U akan diselenggarakan di Peru Dirinya mengatakan belum ada surat resmi dari FIFA mengenai pembatalan drawing pildun U apalagi pengurangan niat menggelar ajang tersebut di Indonesia Mostupdate	exco pssi aya sinulingga membantah kabar piala dunia u akan diselenggarakan di peru dirinya mengatakan belum ada surat resmi dari fifa mengenai pembatalan drawing pildun u apalagi pengurangan niat menggelar ajang tersebut di indonesia mostupdate
Indonesia Batal Tuan Rumah Piala Dunia U Pelatih Shin Tae Yong Ikut Resah	indonesia batal tuan rumah piala dunia u pelatih shin tae yong ikut resah
Lebih baik dipindah Piala Dunia U dari pada menimbulkan keributan di Negeri ini walau sanksi FIFA tdk mengikutkan PSSI ke ajang Internasional pasa berikutnya kita harus siap	lebih baik dipindah piala dunia u dari pada menimbulkan keributan di negeri ini walau sanksi fifa tdk mengikutkan pssi ke ajang internasional pasa berikutnya kita harus siap
FIFA membatalkan drawing pembagian grup Piala Dunia U yang seharusnya digelar di Bali pada buntut sejumlah penolakan atas tim nasional Israel Lebih lanjut wartawan senior olahraga Indonesia Yesayas Oktavianus menyebut FIFA sudah membatalkan	fifa membatalkan drawing pembagian grup piala dunia u yang seharusnya digelar di bali pada buntut sejumlah penolakan atas tim nasional israel lebih lanjut wartawan senior olahraga indonesia yesayas oktavianus menyebut fifa sudah membatalkan
Berarti tidak percaya komitmen TNI Polri dalam menjaga keamanan dan ketertiban Piala Dunia U di Indonesia	berarti tidak percaya komitmen tni polri dalam menjaga keamanan dan ketertiban piala dunia u di indonesia
Caper Baper Tantrum dimana mana dan Cari Gara gara Wali Kota Bengkulu Tolak Israel Bermain di Piala Dunia U via	caper baper lantrum dimana mana dan cari gara gara wali kota bengkulu tolak israel bermain di piala dunia u via
Yg standar ganda itu FIFA aturan nya bisa melakukan pencoret Rusia ikut serta piala dunia di Qatar tetapi kenapa pada zionis Israel FIFA tidak mencoret zionis israel ikut piala dunia u via	yg standar ganda itu fifa aturan nya bisa melakukan pencoret rusia ikut serta piala dunia di qatar tetapi kenapa pada zionis israel fifa tidak mencoret zionis israel ikut piala dunia u via

Gambar 9. Hasil Case Folding

c. Tokenization

Tokenization adalah proses memecahkan kalimat menjadi potongan kata atau token untuk mengetahui asal munculnya kata. Ditunjukkan Gambar 10 hasil tokenisasi pada data set.

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode *Naive Bayes Classifier* (NBC)

	hasil_casefolding	hasil_token
0	exco pssi arya sinulingga membantah kabar piala...	[exco, pssi, arya, sinulingga, membantah, kaba...
1	indonesia batal tuan rumah piala dunia u pela...	[indonesia, batal, tuan, rumah, piala, dunia, ...
2	lebih baik dipindah piala dunia u dari pada m...	[lebih, baik, dipindah, piala, dunia, u, dari,...
3	fifa membatalan drawing pembagian grup piala ...	[fifa, membatalan, drawing, pembagian, grup, ...
4	berarti tidak percaya komitmen tni polri dalam...	[berarti, tidak, percaya, komitmen, tni, polri,...
...
1145	hari ini mei bertepatan pada hari kebangkita...	[hari, ini, mei, bertepatan, pada, hari, keban...
1146	revolusi mental hasilnya batal jadi tuan rumah...	[revolusi, mental, hasilnya, batal, jadi, tuan...
1147	pelipur lara dikala piala dunia u batal federa...	[pelipur, lara, dikala, piala, dunia, u, batal,...
1148	lg berandai andai nih andai november nanti ad...	[lg, berandai, andai, nih, andai, november, na...
1149	oh gitu ya habis piala dunia u di indonesia b...	[oh, gitu, ya, habis, piala, dunia, u, di, ind...

Gambar 10. Hasil Tokenization Data

d. Remove Stopword

Pada tahap ini bertujuan untuk menghilangkan kata-kata yang tidak mempengaruhi sentiment suatu kalimat ditunjukkan pada Gambar 11

	hasil_token	hasil_stopword
0	[exco, pssi, arya, sinulingga, membantah, kaba...	[exco, pssi, arya, sinulingga, membantah, kaba...
1	[indonesia, batal, tuan, rumah, piala, dunia, ...	[indonesia, batal, tuan, rumah, piala, dunia, ...
2	[lebih, baik, dipindah, piala, dunia, u, dari,...	[dipindah, piala, dunia, u, menimbulkan, kerib...
3	[fifa, membatalan, drawing, pembagian, grup, ...	[fifa, membatalan, drawing, pembagian, grup, ...
4	[berarti, tidak, percaya, komitmen, tni, polri,...	[percaya, komitmen, tni, polri, menjaga, keama...
...
1145	[hari, ini, mei, bertepatan, pada, hari, keban...	[mei, bertepatan, kebangkitan, nasional, indon...
1146	[revolusi, mental, hasilnya, batal, jadi, tuan...	[revolusi, mental, hasilnya, batal, tuan, ruma...
1147	[pelipur, lara, dikala, piala, dunia, u, batal,...	[pelipur, lara, dikala, piala, dunia, u, batal,...
1148	[lg, berandai, andai, nih, andai, november, na...	[lg, berandai, andai, nih, andai, november, yg...
1149	[oh, gitu, ya, habis, piala, dunia, u, di, ind...	[oh, gitu, ya, habis, piala, dunia, u, indones...

Gambar 11. Hasil Remove Stopword

d. Wordcloud

Wordcloud merupakan visualisasi kata yang paling sering muncul didalam data set yang diambil. Ditunjukkan pada Gambar 12 adalah contoh visualisasi data dengan wordcloud



Gambar 12. Hasil Wordcloud Kata Paling Sering Keluar

3.3 Klasifikasi Naive Bayes

Setelah melakukan pre processing dan membagi data training dan juga data testing, selanjutnya adalah proses perhitungan klasifikasi sistem dengan menggunakan metode *Naive Bayes classifier*.

Perhitungan manual menggunakan metode *Naive Bayes classifier*.

Menghitung data uji. "Sepakbola indonesia keren" Pada data uji berikut kata yang masuk ke dalam training adalah "sepakbola" dan "indonesia"

$$P(C_i|X) = P(X|C_i)P(C_i)$$

$$\begin{aligned}
 &P(\text{Uji} | \text{Positif}) \\
 &= P(\text{sepakbola} | \text{Pro}) \cdot P(\text{indonesia} | \text{Pro}) \cdot P(\text{Pro}) \\
 &= 0,6666666666666667 \cdot 0,6666666666666667 \cdot 0,3 \\
 &= 0,133333333333
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{Uji} | \text{Kontra}) \\
 &= P(\text{sepakbola} | \text{Kontra}) \cdot P(\text{indonesia} | \text{Kontra}) \cdot P(\text{Kontra}) \\
 &= 0,1666666666666667 \cdot 0,1666666666666667 \cdot 0,6 \\
 &= 0,016666666666
 \end{aligned}$$

Kesimpulan.

Nilai probabilitas tertinggi yaitu 0.1333333333 pada uji Pro, sehingga komentar tersebut diklasifikasikan ke dalam kelas Pro.

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)

Lalu kita melakukan perhitungan secara otomatis menggunakan aplikasi python.

```
#prediksi data uji
predicted=clf.predict(vec_uji)

data_uji.insert(2, column='label_bayes', value=predicted)
data_uji.head(148)
```

Gambar 13. Source code Sistem Klasifikasi Naive Bayes

Berikut ditunjukkan hasil dari prediksi metode *naive bayes*

	hasil_stemming	Label	label_bayes
781	['gubernur', 'wayan', 'koster', 'tolak', 'Isra...	pro	pro
38	['upaya', 'selenggara', 'piala', 'dunia', 'u,...	kontra	pro
678	['udah', 'excited', 'bgt', 'jd', 'gin', 'ya']	kontra	kontra
497	['klo', 'pribadi', 'sihi', 'batal', 'indonesia'...	pro	kontra
1022	['kmrn', 'piala', 'dunia', 'batal', 'gelar', 'l...	kontra	kontra
...
432	['udah', 'banned', 'aja', 'gua', 'mah']	pro	kontra
542	['rugi', 'tiket', 'pesawat', 'kesana', 'udah',...	pro	kontra
10	['gus', 'wagub', 'sepakat', 'selenggara', 'pia...	kontra	pro
1064	['gak', 'heran', 'ajang', 'piala', 'dunia', 'b...	kontra	kontra
399	['putus', 'hormat', 'upaya', 'global', 'rang',...	pro	pro

Gambar 14. Hasil Prediksi Naive Bayes

3.4 Analisis Validitas Data

Analisis validitas data merupakan tahap untuk mengevaluasi metode. Hasil evaluasi yang digunakan pada metode ini berupa nilai confusion matrix berisi nilai akurasi, presisi, dan recall yang diambil dari data test. Dalam menentukan nilai confusion matrix.

a. Akurasi

akurasi atau dalam Bahasa inggrisnya disebut dengan *Accuracy* yaitu nilai yang dihasilkan dari hasil perhitungan dengan banyaknya data yang berhasil diprediksi dengan total data keseluruhan. *Accuracy* yang di hasilkan dari hasil perhitungan yang ditunjukkan pada persamaan di bawah ini.

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

Dari 225 data uji yang telah dilakukan klasifikasi maka didapatkan hasil TP 104, TN: 62, FP: 22, FN 37. Setelah diketahui semuanya maka dapat melakukan perhitungan akurasi sesuai dengan rumus Perhitungan *accuracy* nya adalah sebagai berikut:

$$Accuracy = \frac{62 + 104}{62 + 104 + 22 + 37} = \frac{166}{225} = 0.73$$

Disini penulis melakukan perhitungan manual dan didapatkan hasil dari *accuracy* antara data uji manual dan data uji pada sistem adalah 73.7%. Memiliki presentase angka yang cukup tinggi diklasifikasikan dengan tepat oleh sistem.

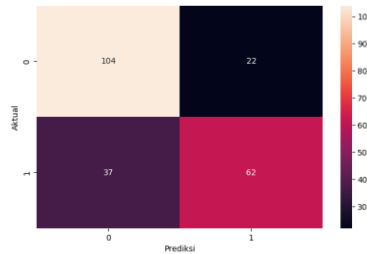
Dibawah ini adalah code dan hasil dari perhitungan otomatis menggunakan Python memiliki hasil yang sama yaitu 73.7% menandakan tidak ada kekeliruan antara perhitungan manual dan otomatis.

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix
f, ax = plt.subplots(figsize=(8,5))
sns.heatmap(confusion_matrix(y_test,predicted), annot=True, fmt=".0f", ax=ax)
plt.ylabel("Aktual")
plt.xlabel("Prediksi")
plt.show()
```

Gambar 15. Source Code Confusion Matrix

Gambar 15 menampilkan visualisasi dari source code confusion matrix diatas. Setelah diketahui *confusion matrix* dari model *naive bayes* yang telah dibuat, selanjutnya melakukan perhitungan nilai akurasi, presisi dan recall dari model klasifikasi Naive Bayes dengan mengimport *accuracy_score*, *precision_score*, *recall_score* yang disediakan oleh library *sklearn*.

Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20 Pada Media Sosial Twitter Dengan Menggunakan Metode Naive Bayes Classifier (NBC)



Gambar 16. Visualisasi Confusion Matrix

```

Nilai Accuracy adalah : 0.7377777777777778
Nilai Precision adalah : 0.7375886524822695
Nilai Recall adalah : 0.8253968253968254
=====
              precision    recall  f1-score   support

   kontra      0.74      0.83      0.78      126
    pro       0.74      0.63      0.68       99

 accuracy              0.74      225
 macro avg           0.74      0.73      0.73      225
 weighted avg       0.74      0.74      0.73      225
    
```

Gambar 17. Hasil Confusion Matrix

Nilai akurasi yang didapat algoritma Naive Bayes adalah 73.7% artinya sejumlah 73.7% model Naive Bayes dapat mengklasifikasikan data yang benar. Dengan mengetahui nilai akurasi model, maka ditentukan bahwa sistem memiliki kemampuan untuk menentukan akurasi antara informasi yang diinginkan pengguna dengan hasil respon yang diberikan oleh sistem. Akurasi atau tingkat keberhasilan sistem pada penelitian ini adalah 73,7%.

Hasil pengujian ini menunjukkan bahwa perhitungan dan hasil manual sama-sama akurat dan hasilnya sesuai dengan yang diharapkan. Ini terjadi karena data pengujian dan data pelatihan memiliki kosa kata yang sama, yang memungkinkan sistem untuk mempelajari dan memproses data pengujian dengan baik.

PENUTUP

Simpulan

Berdasarkan hasil dan pembahasan dari penelitian analisis sentimen terhadap pro dan kontra batalnya indonesia menjadi tuan rumah piala dunia U20 pada media sosial twitter dengan menggunakan metode Naive Bayes classifier, maka penulis menarik kesimpulan sebagai berikut:

- Hasil Implementasi metode naive bayes ditampilkan dengan Bahasa pemrograman Python untuk menganalisa komentar publik terkait batalnya indonesia menjadi tuan rumah piala dunia U20 dari data yang sudah diambil dari twitter sebanyak 1150 data komentar, didapatkan hasil bahwa masyarakat Indonesia merespon batalnya indonesia menjadi tuan rumah piala dunia U20 dengan cukup kontra dimana terdapat 689 sentimen kontra yang didalamnya terdapat komentar yang menyayangkan gagalnya indonesia menjadi tuan rumah. tapi ternyata tidak sedikit juga sentimen pro yaitu sebanyak 461 komentar yang menunjukkan bahwa masih banyak juga orang yang setuju bahwa piala dunia U20 ini layak batal karena israel dan tragedi kanjuruhan. lebih banyaknya sentimen kontra memperlihatkan bahwa banyak masyarakat kecewa dengan tindakan tindakan masyarakat bahkan pemerintah yang menolak piala dunia U20 di Indonesia.
- Penelitian ini berhasil menerapkan metode Naive Bayes untuk analisis sentimen terhadap pro dan kontra batalnya indonesia menjadi tuan rumah piala dunia U20 pada media sosial twitter dengan tingkat akurasi yang cukup baik sebesar 73.7%

Saran

Adapun saran yang dapat diberikan untuk penelitian lanjutan terkait dengan penelitian ini ialah:

- Menggunakan algoritma klasifikasi lain untuk membandingkan hasil uji model untuk menemukan algoritma yang memiliki klasifikasi terbaik. Untuk mengurangi risiko data tidak seimbang,
- Penelitian tentang klasifikasi teks membutuhkan data sentimen yang cukup besar. Pengklasifikasian sentimen awal harus dilakukan dengan banyak orang, karena ini dapat mengurangi obyektifitas hasil sentimen.
- Pada penelitian ini, komentar yang memiliki emoticon dan bahasa asing diabaikan, pada penelitian selanjutnya diharapkan memasukkan emoticon dan bahasa asing ke dalam perhitungan klasifikasi agar mendapatkan hasil yang lebih akurat.
- Pada penelitian ini data hanya dari social media twitter menjadikan peneliti kesulitan dalam hal mencari data, pada penelitian selanjutnya diharapkan memakai lebih dari satu sumber social media agar memudahkan peneliti.

DAFTAR PUSTAKA

Aryani, Dwi Kartika, Dkk. (2023) Dampak Dibatalkannya Indonesia Sebagai Tuan Rumah Piala Dunia U-20 Tahun 2023 Terhadap Perekonomian Indonesia. Jurnal Kreasi Rakyat. 1(2). 122-128

**Analisis Sentimen Terhadap Pro Dan Kontra Batalnya Indonesia Menjadi Tuan Rumah Piala Dunia U20
Pada Media Sosial *Twitter* Dengan Menggunakan Metode *Naive Bayes Classifier* (NBC)**

- Cahyani, T., Gata, W., Saputra, D., Novitasari, H., & Hernawati, H. (2023). Analisis Sentimen Terhadap Telkomsel Dan Xl Berbasis Machine Learning Pada Data *Twitter*. *Intecom: Journal Of Information Technology And Computer Science*, 6(1), 265-274.
- Cahyo Ryan Dwi, Et Al. 2014. Deteksi Dan Validasi Informasi Gempa Secara Real-Time Berbasis Social Sensor Dengan *Twitter*. *Jurnal Teknik Pomits Vol. 2, No. 1*.
- Darmawati, D., Munjin, R. A., & Seran, G. G. (2017). Pengaruh Supervisi Kepala Sekolah Terhadap Kinerja Guru Di Smp Negeri 1 Parung Kecamatan Parung Kabupaten Bogor. *Jurnal Governansi*, 1(1), 13. <https://doi.org/10.30997/jgs.v1i1.294>
- Faiz, M. N., Basori, A. H., & Harjoko, A. (2020). Aplikasi Metode Term Frequency-Inverse Document Frequency (Tf-Iidf) Dalam Analisis Sentimen Pada Ulasan Produk Online. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(12), 5602-5610.
- Giovani, Angelina Puput, Dkk. (2020). Analisis Sentimen Aplikasi Ruang Guru Di *Twitter* Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 116-124
- Heridiansyah, J. (2012). Pengaruh Advertising Terhadap Pembentukan Brand Awareness Serta Dampaknya Pada Keputusan Pembelian Produk Kecap Pedas Abc. *Jurnal Stie Semarang*, 4(2), 53-73.
- Indarwati, Khofifah Diah, & Hery Februariyanti (2023) Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Go-Jek Menggunakan Metode Naive Bayes Classifier. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*. 10(1)
- I. Taufik dan S.A.Pamungkas. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma *Support Vector Machine (SVM)*. *Jurnal "LOG!K@,"* 8(1), 71-79.
- Irwanto, Muslimah Zahro Romas. (2019) Profil Peran Psikologi Olahraga Dalam Meningkatkan Prestasi Atlet Di Serang-Banten Menuju Juara. *Prosiding Seminar Nasional Iptek Olahraga*. 1-14
- Id, Ibnu. Daqiqil. (2021). *Machine Learning: Teori, Studi Kasus Dan Implementasi Menggunakan Python*. Unri Press.
- Katresna, Dena Nurani & Faisal Muhammad Dzikry. (2020) Implementasi Algoritma Naive Bayes Pada Klasifikasi *Tweet* Untuk Mengetahui Tingkat Kemalasan Siswa. *Jurnal Siliwangi*. 6(2) 66-70
- Listiowarni, I. (2019). Implementasi *Naive Bayes* sian Dengan Laplacian Smoothing Untuk Peminatan Dan Lintas Minat Siswa Sman 5 Pamekasan. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(2), 124-129.
- Maulana, Muhammad Ari, Muhammad Lukmanul Hakim & Sukma. (2023) Politik, Olahraga, Dan Islam Studi Kasus Pembatalan Ri Menjadi Tuan Rumah Piala Dunia U20 2023. *Journal Islamic Education*, 1(2), 16-24
- Muljono., Artanti, D. P., & Syukur, A. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma *Naive Bayes* . *Jurnal Konferensi Nasional Sistem Informasi*.
- Murtiningsih, S., & Riptanti, E. (2019). Pemanfaatan Media Sosial *Twitter* Dalam Kampanye Sosial Politik. *Jurnal The Messenger*, 11(1), 78-90.
- Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016, September). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode *Naive Bayes Classifier* . In *Prosiding Seminar Sains Nasional Dan Teknologi (Vol. 1, No. 1)*.
- Pamungkas, F., & Kharisudin, I. (2021). Analisis Sentimen Dengan Svm, Naive Bayes Dan Knn Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial *Twitter*. *Prisma, Prosiding Seminar Nasional Matematika*, 4, 628-634.
- Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM. *Jurnal Media Informatika Budidarma*, 5(1), 149.
- Prasetio, D., Hermawan, D., & Sucahyo, Y. G. (2018). Pemanfaatan Api Instagram Dalam Pengembangan Aplikasi Android. *Jurnal Ilmiah Teknologi Informasi Asia*, 12(2), 146-153.
- Pratama, A. E., Atik Ariesta, & Gata, G. (2022). Analisis Sentimen Masyarakat Terhadap Tim Nasional Indonesia Pada Piala Aff 2020 Menggunakan Algoritma K-Nearest Neighbors. *Jurnal Ticom: Technology Of Information And Communication*, 10(3), 187-196.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. 5(April), 75-82.
- Sari, Retno. (2020) Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn), *Jurnal Sains Dan Manajemen*, 8(1)