

PENERAPAN ALGORITMA KNN UNTUK ANALISIS SENTIMEN KOMENTAR YOUTUBE INDONESIA TUAN RUMAH PIALA DUNIA U-20

Hadi Rahmah Esa Putra^{1*}, Utomo Budiyanto²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}1911500302@student.budiluhur.ac.id, ²utomo.budiyanto@budiluhur.ac.id

(* : corresponding author)

Abstrak-Youtube merupakan sebuah situs web berbagi video yang sangat begitu populer di kalangan pengguna internet dunia. Sebelumnya, Indonesia akan dijadikan tuan rumah Piala Dunia U-20 Tahun 2023 oleh Federasi Sepak Bola Internasional (FIFA), seperti yang ditayangkan di sosial media Youtube. Namun, karena adanya pihak-pihak yang melakukan penolakan kepada Timnas Israel U-20 untuk bermain di Indonesia membuat Indonesia batal menjadi Tuan Rumah Piala Dunia U-20 Tahun 2023. Penolakan beberapa pihak tersebut terhadap Timnas Israel U-20 terjadi karena upaya mendukung kemerdekaan Palestina sesuai amanat Presiden Pertama Indonesia yaitu Ir. Sukarno. Seperti yang diketahui, pembatalan piala dunia U-20 menyebabkan pro dan kontra yang sering diperdebatkan akhir-akhir ini. Banyak masyarakat yang berkomentar di Youtube terhadap pembatalan ini. Maka dari itu perlu dilakukannya analisis sentimen dengan mengolah data komentar. Namun dengan banyaknya jumlah data komentar dan tidak memungkinkannya untuk menghitung satu persatu maka digunakanlah pendekatan Text Mining untuk mengolah banyaknya komentar tersebut apakah dikategorikan sebagai komentar positif, negatif atau netral dengan algoritma K-Nearest Neighbor. Dalam penelitian ini menggunakan nilai K yang berbeda, penetapan nilai K yang berbeda lantaran untuk mendapatkan nilai akurasi yang tinggi, lalu hasil prediksi Sebelum Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 didapat 105 komentar positif (42.0%), 66 komentar negatif (26.4%) dan 79 komentar netral (31.6%) dengan total 250 komentar dan memprediksi Sesudah Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 didapat 27 komentar positif (10.8%), 174 komentar negatif (69.6%) dan 49 komentar netral (19.6%) dengan total 250 komentar sedangkan hasil pengujian didapatkan akurasi Sebelum Pembatalan mendapatkan akurasi sebesar 91%, presisi 91% dan recall 100% dengan nilai K=3. Dan hasil akurasi Sesudah Pembatalan mendapatkan akurasi sebesar 80%, presisi 0% dan recall 0% dengan nilai K=5.

Kata Kunci: piala dunia u-20, analisis sentimen, *k-nearest neighbor*, Youtube

IMPLEMENTATION OF KNN ALGORITHM FOR SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON THE INDONESIAN HOST OF FIFA U-20 WORLD CUP

Abstract-YouTube is a highly popular video-sharing website among internet users worldwide. Previously, Indonesia was designated as the host for the U-20 World Cup in 2023 by the International Federation of Association Football (FIFA), as shown on the social media platform YouTube. However, due to objections from certain parties against the participation of the Israel U-20 national team in Indonesia, Indonesia's role as the host for the 2023 U-20 World Cup was canceled. The objections were based on efforts to support Palestinian independence, following the mandate of Indonesia's first President, Ir. Sukarno. As known, the cancellation of the U-20 World Cup has sparked pro and contra debates recently. Many people have commented on this issue on YouTube. Therefore, sentiment analysis is needed by processing the comment data. However, due to the large amount of comment data, it is not feasible to manually analyze each one. Hence, Text Mining approach is utilized to process the extensive comments, categorizing them as positive, negative, or neutral using the K-Nearest Neighbor algorithm. In this research, different values of K were used, with the selection of different K values aimed at achieving high accuracy. The sentiment analysis results before the cancellation of Indonesia as the host of the U-20 World Cup showed 105 positive comments (42.0%), 66 negative comments (26.4%), and 79 neutral comments (31.6%) out of a total of 250 comments. After the cancellation, the sentiment analysis predicted 27 positive comments (10.8%), 174 negative comments (69.6%), and 49 neutral comments (19.6%) out of a total of 250 comments. The accuracy of the sentiment analysis before the cancellation was 91%, with a precision of 91% and a recall of 100% with K=3. After the cancellation, the accuracy of the sentiment analysis was 80%, with 0% precision and 0% recall with K=5.

Keywords: U-20 World Cup, sentiment analysis, *k-nearest neighbor*, Youtube.

1. PENDAHULUAN

Youtube adalah platform media sosial yang menampilkan video dalam durasi tertentu yang dibuat oleh para pembuat video (*creator youtube*) dengan genre yang berbeda-beda. Tingginya antusiasme masyarakat dalam menggunakan *Youtube* membuat media sosial ini semakin dimanfaatkan oleh berbagai kalangan. *Youtube* menempati urutan media sosial terpopuler yang ada di Indonesia [1]. *Youtube* Indonesia sendiri pada saat tahun 2022 sempat ramai dengan berita akan diselenggarakannya Piala Dunia U-20 Tahun 2023. Piala Dunia (*World Cup*) merupakan turnamen sepak bola terbesar yang dilaksanakan empat tahun sekali oleh FIFA. Yang diikuti oleh tim nasional senior putra anggota Federasi Sepak Bola Internasional (*Fédération Internationale de Football Association, FIFA*), badan pengatur sepak bola dunia [2].

Penyelenggaraan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 yang ditayangkan di sosial media *Youtube*, seperti unggahan channel PSSI TV [3] pada 17 Agustus 2022 yang menayangkan diselenggarakannya Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Tahun 2023. Namun, karena adanya pihak-pihak yang melakukan penolakan kepada Timnas Israel U-20 untuk bermain di Indonesia membuat Indonesia batal menjadi Tuan Rumah Piala Dunia U-20 Tahun 2023, seperti unggahan dari channel KompasTV [4] pada 29 Maret 2023. Seperti yang diketahui, pembatalan piala dunia U-20 menyebabkan pro dan kontra yang sering diperdebatkan akhir-akhir ini. Banyak masyarakat yang berkomentar di *Youtube* terhadap pembatalan ini.

Maka dari itu perlu dilakukannya analisis sentimen dengan mengolah data komentar. Namun dengan banyaknya jumlah data komentar dan tidak memungkinkannya untuk menghitung satu persatu maka digunakanlah pendekatan *Text Mining* untuk mengolah banyaknya komentar tersebut apakah dikategorikan sebagai komentar positif, negatif atau netral. *Text Mining* merupakan salah satu bidang yang luar biasa dalam penggalian informasi yang memiliki arti menggali informasi sebagai pesan dari mana sumber informasi umumnya diperoleh dari arsip dan tujuannya untuk menemukan kata-kata yang bisa menjawab substansi laporan sehingga penyelidikan jaringan antar arsip dapat dilakukan [5].

Analisis sentimen komentar *Youtube* sendiri juga pernah dilakukan Afdhal dkk. [6] dengan algoritma *Random Forest* menghasilkan akurasi sebesar 79%. Ahmadi dkk. [7] juga pernah melakukan analisis sentimen dengan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 74%. Wijayanto dan Defara [8] juga pernah melakukan analisis sentimen dengan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 85,6%. Wahid dan Saputri [9] juga pernah melakukan analisis sentimen dengan algoritma *Naïve Bayes* dan *Decision Tree* menghasilkan akurasi sebesar 66,67% dengan algoritma *Naïve Bayes* dan akurasi sebesar 64,98% dengan algoritma *Decision Tree*. Larasakti dkk. [10] juga pernah melakukan analisis sentimen dengan algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 97% dengan nilai $K=7$. Permana dan Effendi [11] juga pernah melakukan analisis sentimen dengan algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 80%. Maka berdasarkan hasil akurasi dari penelitian sebelumnya, penulis akan melakukan penelitian untuk melakukan analisis sentimen masyarakat tentang Sebelum Pembatalan Piala Dunia U-20 dan Sesudah Pembatalan Piala Dunia U-20 dengan menggunakan algoritma *K-Nearest Neighbor*.

Informasi yang digunakan hanya dari dua video yaitu dari channel *Youtube* PSSI TV dengan judul “Menuju Pentas Piala Dunia U-20 Indonesia 2023 | *World Cup U-20*” dan dari KompasTV dengan judul “Indonesia Batal jadi Tuan Rumah Piala Dunia U-20”, jumlah data sebanyak 500 komentar dengan rincian 250 komentar dari channel *Youtube* KompasTV dan 250 komentar dari channel *Youtube* PSSI TV. Sistem hanya dapat mengetahui opini tentang unggahan dari *Youtube* bernilai positif, negatif dan netral

2. METODE PENELITIAN

2.1 Pengumpulan Data

Proses pengumpulan data menggunakan *API Library*, yakni *Youtube Data API v3*, serta berbagai layanan lainnya seperti *Google Developers Console* dan *Google Colab*. Semua proses dilakukan menggunakan bahasa pemrograman *Python*.

2.2 Preprocessing Data

Preprocessing data adalah proses yang dilakukan pada data mentah. Tujuan dari *preprocessing* data adalah untuk membersihkan, mengubah format, dan menyesuaikan data agar sesuai dengan kebutuhan analisis atau pemodelan. Pada tahapan *preprocessing* data ini terdapat beberapa tahapan, tahapan dari *preprocessing* dalam penelitian ini dapat dilihat pada Gambar 1 berikut:



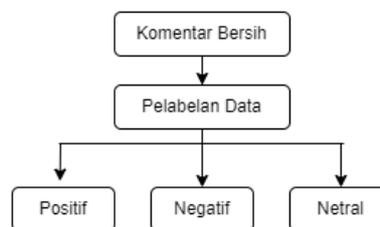
Gambar 1. Tahapan *Preprocessing* Data

- a.) *Cleansing*
Pada tahap *cleansing* ini dilakukan penghapusan atau pembersihan teks adapun pembersihan teks antara lain, yaitu : *URL*, *Hashtag*, angka, tanda baca dan selain huruf a-z dan selain huruf A-Z.
- b.) *Case Folding*
Pada tahap *case folding* ini dilakukan perubahan semua teks dari yang sebelumnya huruf kapital atau huruf besar menjadi huruf kecil agar seragam
- c.) *Mengubah Slangword*
Pada tahap mengubah *slangword* ini dilakukan perubahan kata, yaitu kata tidak baku diubah menjadi kata baku seperti “km” menjadi “kamu”, “udah” menjadi “sudah”, dsb.
- d.) *Stopword Removal*
Pada tahap *stopword removal* ini dilakukan penghapusan kata umum yang biasanya muncul dan tidak memiliki makna seperti kata “dan”, “sudah”, “di”, dsb.
- e.) *Stemming*
Pada tahap *stemming* ini dilakukan mengubah kata menjadi kata dasarnya seperti kata “ditandai” menjadi “tanda”, “dibawa” menjadi “bawa”.

2.3 Pelabelan Data

Pada tahap pelabelan data ini, data komentar yang telah melewati tahap *preprocessing* akan diberi label. Pemberian label dalam penelitian ini menggunakan pakar, yaitu seorang ahli bahasa. Kategori dalam penelitian ini terdiri dari 3 kategori, yaitu Positif, Negatif, dan Netral. Sebelum Pembatalan Piala Dunia U-20, label positif berarti menyatakan mendukung diselenggarakannya Piala Dunia U-20, label negatif berarti menyatakan menolak diselenggarakannya Piala Dunia U-20, dan label netral tidak berpihak.

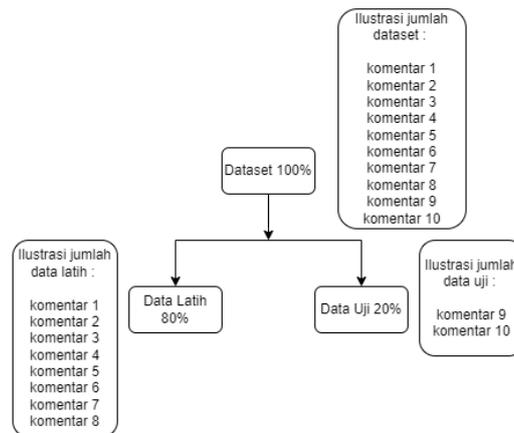
Sedangkan setelah Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20, label positif berarti menyatakan mendukung dibatalkannya Indonesia Menjadi Tuan Rumah Piala Dunia U-20, label negatif berarti menyatakan menolak dibatalkannya Indonesia Jadi Tuan Rumah Piala Dunia U-20, dan label netral tidak berpihak. Berikut ini merupakan gambaran dari tahap pelabelan data pada Gambar 2 berikut:



Gambar 2. Tahapan Pelabelan Data

2.4 Pembagian Data

Pada tahap pembagian data ini, data komentar yang telah diberi label akan dibagi menjadi 2 (dua) bagian, yaitu data uji dan data latih. Pembagiannya akan menjadi 20% data uji dan 80% data latih. Seperti dibawah ini merupakan gambaran dari pembagian data yang dapat dilihat pada Gambar 3 berikut:



Gambar 3. Tahapan Pembagian Data

2.5 Modeling

Pada tahap *modeling* ini, dilakukan ekstraksi komentar dari data latih yang kemudian diubah menjadi vektor menggunakan metode *CountVectorizer*. Tahap ini melibatkan lima proses utama untuk membuat model latih, rincian prosesnya dapat dilihat pada Gambar 4 berikut ini:



Gambar 4. Tahap Modeling

2.6 Klasifikasi K-Nearest Neighbor

Pada tahapan klasifikasi *K-Nearest Neighbor* ini, dilakukan pengklasifikasian dengan menggunakan algoritme *K-Nearest Neighbor*, langkah ini dilakukan setelah ada data latih. Data latih tersebut telah melewati proses pemodelan data. Untuk menerapkan model klasifikasi K-Nearest Neighbor, ada empat proses utama yang dilakukan, Rincian tahapan klasifikasi ini dapat dilihat pada Gambar 5 berikut:



Gambar 5. Klasifikasi K-Nearest Neighbor

Adapun rumus yang digunakan untuk menghitung jarak antara data digunakan perhitungan *Euclidean distance* dengan rumus 1 berikut:

$$d_{(x,y)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2} \quad (1)$$

Keterangan :

- $d_{(x,y)}$: jarak antara data latih dan data uji
- x_i : nilai atribut ke- i dari data uji
- y_i : nilai atribut ke- i dari data latih
- n : jumlah atribut

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Tahap pengumpulan data merupakan tahapan pertama kali yang penulis lakukan. Data berupa komentar “Sebelum Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20” dari *channel Youtube PSSI TV* dan “Setelah Pembatalan Indonesia Jadi Tuan Rumah Piala Dunia U-20” dari *channel Youtube KompasTV*. Pada pengambilan kedua data tersebut penulis menggunakan Google Collab dengan memasukan sebuah Api Key milik penulis dari Google Developer. Data yang digunakan dalam penelitian ini masing-masing berjumlah 250 data komentar.

3.2 Preprocessing Data

Tahap preprocessing data ini dilakukan setelah dataset telah tersedia pada basis data. Preprocessing Data melakukan preprocessing sebanyak 250 data “Sebelum Pembatalan” dan 250 data “Sesudah Pembatalan”. Menurut sub bab (2.2) tahapan preprocessing data meliputi 5 tahap, berikut rincian dari 5 tahapan dalam penelitian ini :

a.) *Cleansing*

Pada proses *cleansing* ini dilakukan penghapusan atau pembersihan teks. Adapun pembersihan teks meliputi: URL, Hastag, angka, tanda baca dan selain huruf a-z dan selain huruf A-Z . Contoh penerapannya dapat dilihat pada Gambar 6 dibawah ini :



Gambar 6. Tahapan *Cleansing*

b.) *Case Folding*

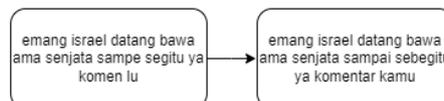
Pada proses *case folding* ini dilakukan penyeragaman kata dengan cara membuat semua teks yang berhuruf kapital atau huruf besar menjadi huruf kecil. Contoh penerapannya dapat dilihat pada Gambar 7 dibawah ini:



Gambar 7. Tahapan *Case Folding*

c.) Mengubah *Slangword*

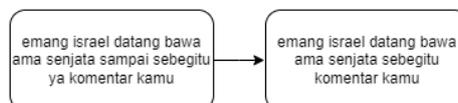
Pada tahap mengubah slangword ini dilakukan pengubahan kata, yaitu kata yang tidak baku menjadi kata baku, Adapun kata yang diubah seperti kata “lu” menjadi kata “kamu”. Contoh penerapannya dapat dilihat pada Gambar 8 dibawah ini :



Gambar 8. Tahapan Mengubah *Slangword*

d.) *Stopword Removal*

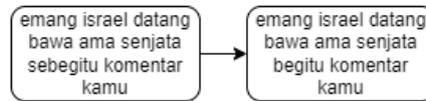
Pada tahap mengubah slangword ini dilakukan pengubahan kata, yaitu kata yang tidak baku menjadi kata baku, Adapun kata yang diubah seperti kata “lu” menjadi kata “kamu”. Contoh penerapannya dapat dilihat pada gambar 9 dibawah ini :



Gambar 9. Tahapan *Stopword Removal*

e.) *Stemming*

Pada tahap stemming ini dilakukan mengubah kata menjadi kata dasarnya seperti kata “sebegitu” menjadi “begitu”. Contoh penerapannya dapat dilihat pada Gambar 10 dibawah ini:



Gambar 10. Tahapan *Stemming*

3.3 Pelabelan Data

Tahapan melabelkan data ini dilakukan setelah tahapan preprocessing data selesai. Pelabelan data dilakukan secara manual oleh seorang pakar adapun pakar tersebut adalah ahli bahasa dengan mempunyai kriteria yaitu sebagai lulusan S2 Linguistik. Pakar tersebut melabelkan data komentar *Youtube* tersebut bernilai positif, negatif atau netral. Data yang dilabelkan sebanyak 250 data “Sebelum Pembatalan” dan 250 data “Sesudah Pembatalan”.

3.4 Pembagian Data

Tahapan pembagian data ini dilakukan setelah tahap pelabelan data selesai. Pembagian data pada dataset ini dibagi menjadi 20:80 yang berarti 20% data uji dan 80% data latih. Pada pembagian data ini akan disaring menjadi label bernilai positif dan negatif, sedangkan label yang bernilai netral tidak diikuti sertakan karena dalam pengujian dan evaluasi data nanti yang melibatkan perhitungan *confusion matrix*, hanya menghitung label positif dan negatif. Lalu setelah disaring data “Sebelum Pembatalan” sebanyak 171 data dengan rincian 34 data uji dan 137 data latih. Sedangkan data “Sesudah Pembatalan” sebanyak 201 data dengan rincian 40 data uji dan 161 data latih yang kemudian dimasukkan ke dalam basis data untuk ke tahap selanjutnya yaitu pemodelan data.

3.5 Modeling Dengan Ekstraksi Fitur *CountVectorizer*

Tahap modeling ini dilakukan setelah tahap pembagian data selesai. Seperti pada sub bab (2.5). Berikut dibawah ini merupakan rincian tahapannya :

3.5.1 Pemilihan Data Latih

Setelah melewati pelabelan data, akan dilakukan pembagian data menjadi data uji dan data latih. Data latih akan dipilih untuk dilakukan pemodelan, data latih yang akan dilakukan pemodelan data dibagi secara seimbang berdasarkan label minoritas, seperti tabel dibawah ini merupakan sampel data latih “Sebelum Pembatalan”, terdapat pada tabel 1 berikut :

Tabel 1. Sampel Data Latih (Sebelum Pembatalan)

latih ke-i (komentar)	Komentar Bersih	Label
latih ke-1	menang please	Positif
latih ke-2	mashaallah indonesia jadi tuan rumah	Positif
latih ke-3	siapa isi theme songnya	Positif
.....
latih ke-8	bilang indonesia negara miskin mana nih indonesia nih bos shenggol dong shenggol dong upin ipin	Negatif

3.5.1 Memecah daftar teks (dari data latih) menjadi daftar kata

Setelah pemilihan data latih, data komentar yang terdapat pada data latih akan dipecah menjadi daftar kata serta menghilangkan kata duplikat. Contoh penerapan pada data “Sebelum Pembatalan” terdapat pada Tabel 2 dibawah ini:

Tabel 2. Sampel Daftar Kata (Sebelum Pembatalan)

Daftar Kata
"menang", "please", "mashaallah", "indonesia", "jadi", "tuan", "rumah", "siapa", "isi", "theme", "songnya", "lutfi", "iya", "informasi", "cameranya", "bukan", "pak", "mas", "cameramannya", "bisa", "pakai", "pildun", "ngeshoot", "orang", "firasat", "tidak", "enak", "kenal", "agama", "dong", "kaya", "qatar", "aji", "dulu", "studion", "bilang", "negara", "miskin", "mana", "nih", "bos", "shenggol", "upin", "ipin"

Tabel 9. Menghitung Jarak (Sebelum pembatalan)

Latih ke-i (Komentar)	Uji ke-i (Komentar)	<i>Euclidean Distance</i> (d latih ke-i – uji ke-i)
Latih ke-1	Uji ke-1	$d(1-1) = 2$
Latih ke-2		$d(2-1) = 2.23606797749979$
Latih ke-3		$d(3-1) = 2.449489742783178$
.....	
Latih ke-8		$d(8-1) = 4.358898943540674$
Latih ke-1	Uji ke-2	$d(1-2) = 1.7320508075688772$
Latih ke-2		$d(2-2) = 2$
Latih ke-3		$d(3-2) = 2.23606797749979$
.....	
Latih ke-8		$d(8-2) = 4.47213595499958$

3.6.5 Mencari K Tetangga Terdekat

Dalam proses ini, penulis akan mencari tetangga terdekat untuk klasifikasi menggunakan *K-Nearest Neighbor*, untuk melakukannya memerlukan nilai K. Dalam penelitian ini menggunakan nilai K yang berbeda, penetapan nilai K yang berbeda lantaran untuk mendapatkan nilai akurasi yang tinggi. Nilai K yang digunakan untuk sebelum pembatalan yaitu K=3 dan sesudah pembatalan yaitu K=5. Kemudian nilai jarak yang sudah didapat sebelumnya diurutkan berdasarkan jarak terkecil. Setelah itu urutan 4-8 pada data sebelum pembatalan akan dihapus karena menggunakan K=3. Rincian hasil pengurutan nilai jarak pada data sebelum pembatalan dapat dilihat pada Tabel 10 dibawah ini :

Tabel 10. Mencari K Tetangga Terdekat (Sebelum Pembatalan)

Urutan	Latih ke-i, Uji ke-i (Komentar)	<i>Euclidean Distance</i> (d latih ke-i – uji ke-i)
1	Latih ke-1, Uji ke-1	$d(1-1) = 2$
2	Latih ke-2, Uji ke-1	$d(2-1) = 2.23606797749979$
3	Latih ke-4, Uji ke-1	$d(4-1) = 2.23606797749979$
1	Latih ke-1, Uji ke-2	$d(1-2) = 1.7320508075688772$
2	Latih ke-2, Uji ke 2	$d(2-2) = 2$
3	Latih ke-4, Uji ke-2	$d(4-2) = 2$

3.6.6 Menghitung Nilai Probabilitas

Setelah selesai mengambil K tetangga terdekat, kemudian menghitung nilai probabilitas label, label positif akan menjadi probabilitas positif dan label negatif akan menjadi probabilitas negatif, pada Tabel 11 merepresentasikan nilai probabilitas sebelum pembatalan. Berikut dibawah ini merupakan rinciannya :

Tabel 11. Menghitung Nilai Probabilitas (Sebelum Pembatalan)

Latih ke-i	Uji ke-i	Label	Probabilitas Positif	Probabilitas Negatif
Latih ke-1	Uji ke-1	Positif	1	0
Latih ke-2		Positif	1	0
Latih ke-4		Positif	1	0
Jumlah			3	0
Latih ke-1	Uji ke-2	Positif	1	0
Latih ke-2		Positif	1	0
Latih ke-4		Positif	1	0
Jumlah			3	0

Berdasarkan Tabel 11 diatas. Nilai probabilitas dengan K=3 pada data sebelum pembatalan dapat diketahui nilai probabilitas pada uji ke 1 dan uji ke 2 mendapat nilai probabilitas yang sama yaitu bernilai positif 3 dan negatif 0. Maka bisa disimpulkan hasil pengujian pada data sebelum pembatalan dengan K=3 memprediksi label positif.

4. KESIMPULAN

Kesimpulan yang didapat dari hasil pengujian dan evaluasi pada penelitian ini didapat bahwa penelitian ini berhasil memprediksi Sebelum Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 didapat 105 komentar positif (42.0%), 66 komentar negatif (26.4%) dan 79 komentar netral (31.6%) Sedangkan Sesudah Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 didapat 27 komentar positif (10.8%), 174 komentar negatif (69.6%) dan 49 komentar netral (19.6%) dengan masing-masing total 250 komentar. Hasil akurasi Sebelum Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 menunjukkan akurasi sebesar 91%, presisi 91% dan *recall* 100% dengan nilai $K=3$, sedangkan hasil akurasi Sesudah Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 menunjukkan akurasi sebesar 80%, presisi 0% dan *recall* 0% dengan nilai $K=5$. Berdasarkan hasil prediksi disimpulkan bahwa sentimen Sebelum Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 bernilai positif sedangkan sentimen Sesudah Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 bernilai negatif

DAFTAR PUSTAKA

- [1] D. K. Aryani, E. W. Putri, M. Rafly, N. Aulia, and P. Maulidina, "Dampak Dibatalkannya Indonesia Sebagai Tuan Rumah Piala Dunia U-20 Tahun 2023 Terhadap Perekonomian Indonesia the Impact of the Cancellation of Indonesia As Host of the 2023 U-20 World Cup on the Indonesian Economy," vol. 1, no. 2, pp. 122–128, 2023.
- [2] (2022) PSSI TV. [Online]. Available: https://www.youtube.com/watch?v=vqKeII0Yhl8&ab_channel=PSSITV. [Accessed 5 April 2023, 11:00 AM]
- [3] (2023) KompasTV. [Online]. Available: https://www.youtube.com/watch?v=ncZ3gYOvMls&ab_channel=KOMPASTV. [Accessed 5 April 2023, 08:13 AM]
- [4] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," *JNKTI (Jurnal Nas. Komputasi dan Teknol. Informasi)*, vol. 5, no. 1, pp. 131–140, 2022, [Online]. Available: <http://repository.uin-suska.ac.id/59746/>
- [5] M. A. Ramadhan and M. I. Wahyudin, "Analisis Sentimen Mengenai Keberhasilan Indonesia di Ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode Naïve Bayes dan Decision Tree," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 505–511, 2022, doi: 10.35870/jtik.v6i4.560.
- [6] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekka.ac.id/jnkkti/article/view/4004/pdf>
- [7] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naïve bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 807–814, 2021.
- [8] A. Wijayanto, "Analisis Sentimen Komentar Youtube Mengenai Vaksin Covid-19 Menggunakan Support Vector Machine," *J. PILAR Teknol. J. Ilm. Ilmu Ilmu Tek.*, vol. 7, no. 1, pp. 24–31, 2022, doi: 10.33319/piltek.v7i1.118.
- [9] A. Wahid and G. Saputri, "Analisis Sentimen Komentar Youtube Tentang Relawan Patwal Ambulance Menggunakan Algoritma Naïve Bayes dan Decision Tree," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 319, 2022, doi: 10.30865/json.v4i2.4941.
- [10] D. N. Larasakti, A. Aziz, and D. Aditya, "Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor," *J. Ilm. Wahana Pendidik.*, vol. 2023, no. 5, pp. 132–142, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.7728573>
- [11] A. Y. Permana and M. Makmun, "Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN," *J. Ilm. Komputasi*, vol. 19, no. 1, pp. 39–50, 2020, doi: 10.32409/jikstik.19.1.154.