

ANALISIS SENTIMEN MASYARAKAT TWITTER TERHADAP PIALA DUNIA U-20 2023 DI INDONESIA MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Saddam Afghan Ramdhani^{1*}, Sejati Waluyo²

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

Email: ^{1*}1911500799@student.budiluhur.ac.id, ²sejati.waluyo@budiluhur.ac.id

(* : *corresponding author*)

Abstrak- Piala Dunia U-20 FIFA 2023 akan menjadi iterasi ke-23 turnamen. Dari 20 Mei hingga 11 Juni 2023, turnamen akan diperebutkan di Argentina. Juara incumbent Ukraina gagal lolos ke Kejuaraan Eropa U-18, membuat mereka tidak dapat mempertahankan gelar mereka. Kemenangan Serbia 2015 mengikuti kemenangan Kejuaraan Dunia Pemuda FIFA 1987 Yugoslavia. Kemudian, Uruguay menjadi juara baru. Pada kenyataannya, Indonesia akan menjadi tuan rumah Piala Dunia U-20 pada tahun 2021. Namun, pada 24 Desember 2020, FIFA mengumumkan pembatalan turnamen 2021 karena pandemi COVID-19. Namun, Indonesia kembali terpilih sebagai negara tuan rumah untuk edisi 2023. Pada 29 Maret 2023, hak tuan rumah Indonesia dicabut karena protes terhadap partisipasi keluarga Israel. Argentina terpilih menjadi tuan rumah Piala Dunia U-20 FIFA 2023 sebagai pengganti Indonesia pada 17 April 2023. Ini adalah kedua kalinya turnamen diadakan di Argentina. Federasi Sepak Bola Dunia (FIFA) menarik Piala Dunia U-20 FIFA 2023 dari Indonesia, meski faktanya Presiden Federasi Sepak Bola Indonesia (PSSI) Erick Thohir bertemu langsung dengan Presiden FIFA di Doha, Qatar untuk berdebat dengan juri. Teknik yang digunakan adalah ekstraksi fitur. CountVectorizer dan algoritma K-Nearest Neighbor untuk mengklasifikasikan data. Dataset yang digunakan terdiri dari query (tweet) dengan istilah 'U20HARUSJADI' dan 'KitaMauMainBola' di jejaring sosial Twitter. Analisis terhadap 232 tweet mengungkapkan sentimen positif 72,84% dan sentimen negatif 27,16% sejak 1 Maret hingga 30 April 2023. Hasil pengujian terbaik diperoleh dengan menggunakan nilai K=9, dengan akurasi 43%, presisi 35%, dan recall 100%.

Kata Kunci: twitter, U20HARUSJADI, KitaMauMainBola, *counvectorizer*, *k-nearest neighbor*

ANALYSIS OF PUBLIC SENTIMENT ON TWITTER SOCIAL MEDIA FOR THE 2023 U20 WORLD CUP IN INDONESIA USING THE *K-NEAREST NEIGHBOR (KNN) ALGORITHM*

Abstract- The 2023 FIFA U-20 World Cup will be the 23rd iteration of the tournament. From May 20 to June 11, 2023, the tournament will be contested in Argentina. The incumbent champion Ukraine failed to qualify for the European Under-18 Championship, rendering them unable to defend their title. Serbia's 2015 victory followed Yugoslavia's 1987 FIFA Youth World Championship triumph. Then, Uruguay became the new champion. In actuality, Indonesia will host the U-20 World Cup in 2021. However, on December 24, 2020, FIFA announced the cancellation of the 2021 tournament due to the COVID-19 pandemic. However, Indonesia was again selected as the host nation for the 2023 edition. On March 29, 2023, Indonesia's hosting rights were revoked due to protests against the participation of Israeli families. Argentina was selected to host the 2023 FIFA U-20 World Cup in lieu of Indonesia on 17 April 2023. This was the second time the tournament had been held in Argentina. The World Football Federation (FIFA) withdrew the 2023 FIFA U-20 World Cup from Indonesia, despite the fact that Indonesian Football Federation (PSSI) President Erick Thohir met directly with the FIFA President in Doha, Qatar to debate the jury. The employed technique is feature extraction. CountVectorizer and the K-Nearest Neighbor algorithm for classifying data. The dataset used consisted of (tweet) queries with the terms 'U20HARUSJADI' and 'KitaMauMainBola' on the social network Twitter. The analysis of 232 tweets revealed a positive sentiment of 72.84% and a negative sentiment of 27.16% from 1 March to 30 April 2023. The best test results were obtained using the value of K=9, with an accuracy of 43%, precision of 35%, and recall of 100%.

Keywords: twitter, U20HARUSJADI, KitaMauMainBola *counvectorizer*, *k-nearest neighbor*

1. PENDAHULUAN

Piala Dunia FIFA secara luas dianggap sebagai kompetisi sepakbola global yang unggul. Kompetisi ini diawasi oleh Federasi Sepak Bola Internasional (FIFA) dan melibatkan partisipasi tim sepak bola nasional dari berbagai negara di seluruh dunia. Piala Dunia FIFA telah diselenggarakan secara empat tahunan sejak didirikan pada tahun 1930, kecuali tahun 1942 hingga 1946, di mana Piala Dunia ditangguhkan karena konflik global Perang Dunia II. Konsep penyelenggaraan acara olahraga global yang dikenal dengan Piala Dunia ini awalnya diusulkan oleh Jules Rimet, yang menjabat sebagai presiden FIFA pada periode tersebut. Edisi perdana Piala Dunia FIFA berlangsung di Uruguay pada tahun 1930, yang berpuncak pada kemenangan bagi negara asalnya. Selanjutnya, turnamen tersebut telah mengumpulkan antusiasme yang signifikan di antara para penggemar sepak bola global. Awalnya, Piala Dunia dipertandingkan secara eksklusif oleh tim nasional dalam jumlah terbatas. Namun demikian, seiring berjalannya waktu, terjadi peningkatan jumlah tim yang ambil bagian dalam acara tersebut, disertai dengan beberapa perubahan format kompetisi. Saat ini, Piala Dunia dipertandingkan oleh total 32 tim nasional yang semuanya berhasil lolos ke turnamen melalui kualifikasi regional yang diadakan di konfederasi masing-masing.

Manfaat dari penelitian ini adalah Penelitian ini merupakan sebuah eksplorasi dari teori yang selama ini dipelajari, serta menambah wawasan, ilmu pengetahuan, dan pengalaman terhadap analisis sentimen *Teks mining*. Bagi universitas, sebagai tolak ukur mahasiswa dalam menguasai ilmu yang sudah dipelajari dan sebagai referensi untuk penelitian selanjutnya. Bagi pembaca, memberikan informasi mengenai sentimen terhadap penyelenggaraan piala dunia U20 2023 di Indonesia dan sebagai referensi penelitian analisis sentimen di bidang sistem informasi.

Berdasarkan analisis riset, telah ditetapkan bahwa Indonesia tidak lagi mampu menjadi tuan rumah putaran final Piala Dunia U20 2023. Keputusan ini diambil menyusul pencabutan hak tuan rumah oleh Federasi Sepak Bola Dunia (FIFA). Khususnya, Presiden Federasi Sepak Bola Seluruh Indonesia (PSSI), Erick Thohir, secara pribadi bertemu dengan Presiden FIFA di Doha, Qatar. Negosiasi.

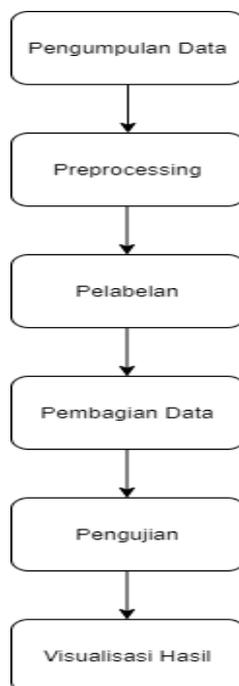
Keputusan untuk menangguhkan Piala Dunia U20 2023 di Indonesia telah menimbulkan kekecewaan di kalangan individu tertentu, dengan wacana publik di media sosial mengaitkan kegagalan sebagian besar dengan munculnya narasi politik.

Penelitian sebelumnya tentang analisis sentimen mencakup penyelidikan perintis yang menerapkan teknik penambangan teks pada aplikasi Tokopedia, menggunakan algoritma K-Nearest Neighbor [1]. Studi ini melaporkan penggunaan 5.254 contoh data, yang dipartisi menggunakan validasi silang K-fold, dan kemudian dievaluasi keakuratannya. Parameter K=200 sesuai dengan nilai 71,12%. Dalam kajian terpisah, dilakukan analisis terhadap sentimen seputar fenomena “new normal” di Indonesia pada masa Covid-19 [2]. Dataset yang digunakan dalam analisis ini terdiri dari 1000 tweet, dengan 811 tweet yang mengekspresikan sentimen positif dan 189 tweet yang mengekspresikan sentimen negatif. Algoritma klasifikasi K-NN, ketika diterapkan dengan nilai K=1, menunjukkan akurasi sempurna 100% pada set pelatihan. Selain itu, akurasi mencapai 92,60% saat dievaluasi menggunakan validasi silang 10 kali lipat dan 94,50% saat menggunakan pembagian persentase 80%. Algoritma K-Nearest Neighbor (KNN) digunakan dalam penelitian bernama “Analisis Sentimen Layanan PT. PLN di Jakarta di Twitter” [3]. Penelitian ini melibatkan klasifikasi data menggunakan KNN dengan metrik jarak Euclidean. Studi ini menggunakan nilai K = 3 untuk menentukan jarak antara titik data yang berdekatan. Mempertimbangkan biaya Angka akurat yang tercatat untuk wilayah DKI Jakarta pada Februari 2020 adalah 89,4 sesuai dengan peringkatnya dari total 500 titik data.

Oleh karena itu, penelitian ini bertujuan untuk mengetahui mood publik terhadap penyelenggaraan Piala Dunia U20 2023 di Indonesia. Kajian ini berpotensi untuk meningkatkan pemahaman terkait penyelenggaraan Piala Dunia U20 2023 di Indonesia.

2. METODE PENELITIAN

Perencanaan Diagram Analisis Sentimen adalah gambaran yang menunjukkan alur penelitian yang dilakukan. Alur pencarian bisa dilihat pada Gambar 1 di bawah ini.



Gambar 1. Alur Penelitian

Proses pengumpulan data melibatkan pemanfaatan Google Collapse untuk merayapi dan mendapatkan kumpulan data tweet. Selanjutnya tweet yang dihasilkan disimpan dalam format Excel dan kemudian dimuat ke dalam database untuk keperluan pre-processing. Hasil dari tahap preprocessing menghasilkan kalimat yang menunjukkan peningkatan struktur dan kebersihan, disebut sebagai teks bersih, yang selanjutnya digunakan pada langkah selanjutnya. Teks yang diproses sebelumnya akan menjalani pemrosesan lebih lanjut selama fase pelabelan untuk memastikan jenis sentimen (label) sebagai positif atau negatif. Selanjutnya, tweet berlabel tersebut akan dipartisi menjadi dua set yang berbeda, yaitu data uji dan data latih. Data pelatihan mengacu pada informasi yang digunakan untuk memfasilitasi perolehan pengetahuan selama proses kategorisasi. Proses ini melibatkan konstruksi dan pemanfaatan model pelatihan, serta pemanfaatan data pelatihan yang ada. Data uji mengacu pada data yang secara khusus dibuat untuk menilai akurasi model pelatihan yang dikembangkan melalui proses pemodelan. Verifikasi kebenaran dilakukan melalui serangkaian tahapan, antara lain pelatihan, pengujian, dan evaluasi. Setelah proses pengujian dan penilaian selesai, hasil implementasi metodologi akan disajikan secara visual, menggunakan persentase dan grafik.

2.1 Pengumpulan data

Proses pengumpulan data dilakukan melalui eksplorasi. Pemanfaatan teknik crawling sangat penting dalam konteks pengumpulan data dan pemantauan evolusi cepat Internet [4]. Prosesnya melibatkan penggalian data dari platform media sosial Twitter menggunakan Google Colabs, yang secara khusus menargetkan tweet. Tweet yang terkumpul kemudian direkam dalam format file Excel (.xlsx), yang kemudian diimpor ke database MySQL. Dataset yang digunakan dalam penelitian ini bersumber dari Twitter, yang terdiri dari data teks tweet yang dikumpulkan antara tanggal 20 Maret 2023 dan 29 Mei 2023, berjumlah total 509 titik data. Adapun atribut data yang dikumpulkan melalui proses penambangan yang meliputi: id tweet, teks_lengkap(tweet), dibuat_at, nama pengguna, akan disimpan dalam file excel (.xlsx)

2.2 Preprocessing

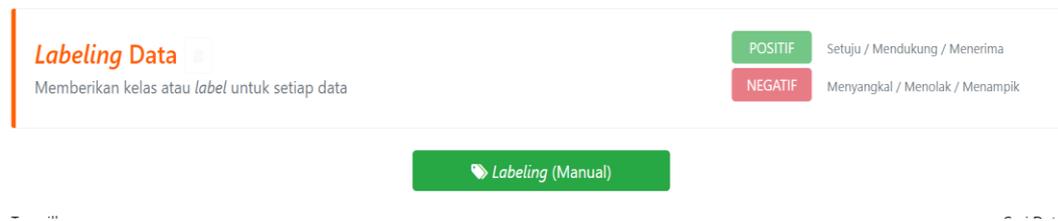
Tahap awal analisis sentimen saat menggunakan data Twitter adalah preprocessing, yang memiliki kepentingan signifikan karena berdampak langsung pada hasil kinerja klasifikasi [5]. Konversi data teks tidak terstruktur

menjadi data teks terstruktur memerlukan prapemrosesan data. Oleh karena itu, penelitian ini akan mencakup beberapa fase preprocessing teks, meliputi:

- 2.2.1. “*Preprocessing* merupakan tahapan awal yang paling penting dalam analisis sentimen saat menggunakan data *twitter*, karena berpengaruh terhadap hasil performansi klasifikasi [5].
- 2.2.2. *Cleaning* merupakan proses untuk menghilangkan tanda baca, angka, simbol, atau link URL, dan username dalam teks[5].
- 2.2.3. Mengubah *slang word* merupakan proses pengubahan kata tidak baku menjadi kata baku, yang dilakukan dengan menggunakan kamus kata slang. Kata-kata dalam kamus slang kemudian diubah menjadi kata baku. Contoh kata slang adalah baper, sotoy, messel mager dan lain-lain [6].
- 2.2.4. Menghapus *stop word* akan menghilangkan kata-kata yang dianggap tidak penting didalam teks. Seperti kata ‘bisa’, ‘tegas’, ‘mampu’, dan ‘tidak’[5].
- 2.2.5. Stemming *Stemming* akan mengubah kata yang ada pada tweet menjadi bentuk dasarnya sehingga dapat mengurangi variasi fitur-fitur yang memiliki makna yang sama namun karena terdapat imbuhan pada kata tersebut menyebabkan adanya perbedaan makna kata [5]”

2.3 Pelabelan

Pelabelan merupakan adalah tahap pemberian label pada dokumen (*tweets*) yang diperoleh dari hasil *twitter* crawling dan telah melewati proses data selection [7]. Pada penelitian ini, proses pelabelan mengklasifikasikan setiap tweet menjadi positif atau negatif (2) kelas yang dilakukan dengan melakukan pelabelan manual dengan (3) orang yang melakukan pelabelan dan (1) orang yang melakukan validasi data tersebut.



Gambar 2. Labeling Manual

2.4 Pembagian data

Selama fase berbagi data, tweet berlabel dipartisi menjadi dua subset yang berbeda: data pengujian dan data pelatihan. Proses berbagi data melibatkan partisi kumpulan data menjadi dua subset, yaitu data pelatihan, yang merupakan 90% dari data asli, dan data uji, yang merupakan 10% sisanya.

2.5 Pemodelan

Pemodelan mengacu pada prosedur sistematis untuk memperoleh informasi atau wawasan dari data pelatihan yang ada. Data latih model dipilih melalui pemanfaatan teknik kuota sampling. Sampling kuota adalah strategi pengambilan sampel yang digunakan untuk memastikan jumlah sampel dari suatu populasi yang memiliki sifat atau kriteria tertentu, sampai kuota yang diinginkan tercapai. [8]

2.6 Pengujian dan evaluasi

Pengujian model menggunakan data uji [9]. Pekerjaan ini melibatkan melakukan tes pada model yang dikembangkan menggunakan data eksperimen yang ada. Evaluasi temuan pengujian akan dilakukan dengan menggunakan matriks perancu untuk menilai tingkat akurasi, presisi, dan daya ingat.

Teknik K-Nearest Neighbor (KNN) umumnya digunakan dalam tugas klasifikasi. Algoritma K-Nearest Neighbor (KNN) adalah pendekatan langsung yang digunakan untuk mengatasi masalah klasifikasi. Premis operasinya melibatkan penentuan jarak terkecil antara data yang akan dievaluasi dan tetangga terdekat dalam data penelitian. Meskipun sederhana, teknik KNN sering menghasilkan hasil yang kompetitif dan signifikan secara statistik. Untuk menghitung jarak melalui metode Euclidean [10]. Rumus jarak Euclidean:

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Penelitian ini melakukan pengujian untuk mengevaluasi akurasi, presisi, dan recall implementasi algoritma K-Nearest Neighbor (KNN) dalam memprediksi label untuk data uji. Selain mengevaluasi akurasi, presisi, dan recall, penelitian ini juga menggabungkan penilaian nilai K dengan menggunakan variasi yang telah ditentukan, yaitu K=3, K=5, K=7, K=9, dan K=11. Proses selanjutnya setelah menghitung jarak untuk setiap data latih adalah mencari data latih dengan nilai jarak terkecil (ketetanggaan terdekat) sebanyak nilai K yang telah ditentukan. Proses akhir setelah ditemukannya data tetangga terdekat adalah pemungutan suara (*voting*). Voting bertujuan untuk menentukan kelas atau label dari suatu data uji.

3. Hasil dan pembahasan

3.1 Dataset

Pengumpulan dataset dilakukan pada tanggal 1 maret 2023 hingga 30 April 2023. Tercatat sejumlah 232 tweet. Properti kepemilikan meliputi tweet_id, teks lengkap (tweet), dibuat di, nama pengguna .

3.2 Preprocessing Data

Seperti yang dijelaskan pada bagian sebelumnya, kumpulan data awal yang bersumber dari platform media sosial Twitter telah mengalami pra-pemrosesan. Prosedur preprocessing yang dilakukan dalam penelitian ini meliputi case folding, cleaning, slang modification, stop word eliminasi, dan stemming. Tabel 1 menyajikan strategi untuk preprocessing data.

Tabel 1. Proses *preprocessing* data

Tahap	Sebelum	Sesudah
Case Folding	AKU TITIPKAN NEGARA INI PADAMU..	Aku titipkan negeri ini padamu
Cleansing	Ketua DPP PDIP ke Presiden Jokowi: Piala Dunia U-20 Tak Mungkin Dipisahkan dari Politik https://t.co/pc3VoLkKKU	Ketua DPP PDIP ke Presiden Jokowi: Piala Dunia U-20 Tak Mungkin Dipisahkan dari Politik
Merubah slang word	Gw setuju piala dunia di indonesia	Saya setuju piala dunia di indonesia
Menghapus stop word	Sejak awal melihat ada ketulusan dan cita-cita dalam dirinya untuk memajukan sepakbola indonesia	Sejak awal melihat ketulusan cita-cita dalam dirinya untuk memajukan sepakbola indonesia
Stemming	Kalo liat dari gaya jalannya sih mendebarakan	Kalo liat dari gaya jalannya sih debar

3.3 Pengujian dan Evaluasi

Selama fase pengujian dan evaluasi ini, model pelatihan dan data uji digunakan. Berdasarkan model pelatihan yang terdiri dari 209 data pelatihan dengan karakteristik sebagai berikut 47 tweet positif dan 47 tweet pelatihan negatif dan 23 data uji, berikut pada Tabel 2 adalah sampel hasil prediksi algoritma KNN dengan nilai K=9.

Tabel 2. Sampel data hasil prediksi

No	Tweet	Label aktual	Label prediksi
1	Waspada! gaslighting ala pendukung U20 harus di Indonesia	Positif	Positif

2	Gue termasuk yang mendukung Timnas Israel main di Indonesia Keputusan FIFA sudah keluar ... kecewa dan sedih Tidak baik juga berlarut terus dlm kekecewaan akhirnya sesama anak bangsa saling perang opini, saling menyalahkan Gue ambil hikmahnya aja	Positif	Positif
3	Yang kemarin ngasih pernyataan bikin repot panitia, Apakah suruh buat pernyataan lagi setelah terjadi begini	Positif	Negatif
4	kurg ajar masak kerjaan anies di analogikan	Positif	Negatif
5	Tetap Berjuang Papuans	Postif	Positif
.....
50	nih ga usah ribut sendiri, permasalahnya ada sama dia si paling kotor	Negatif	Negatif

Pada Tabel 2 sampel data prediksi dan aktual sesuai dengan data yang diperoleh melalui proses pelabelan, sedangkan label prediksi sesuai dengan data label yang dihasilkan oleh prosedur klasifikasi dengan bantuan KNN. Hasil prediksi keseluruhan dari data uji (50 data tweet) kemudian diplot dalam matriks konfusi. Representasi dari matriks konfusi untuk K=9 yang terbentuk pada Tabel 3 dapat dilihat di bawah ini.

Tabel 3. Nilai confusion matrix pengujian K=9

		Nilai aktual	
		Positif	Negatif
Nilai prediksi	Positif	7	13
	Negatif	0	3

Berdasarkan Tabel 3 Uji Confusion Matriks K=9, nilai akurasi, presisi dan recall dapat diperoleh pada Tabel 4 di bawah ini.

Tabel 4. Nilai pengujian K=9

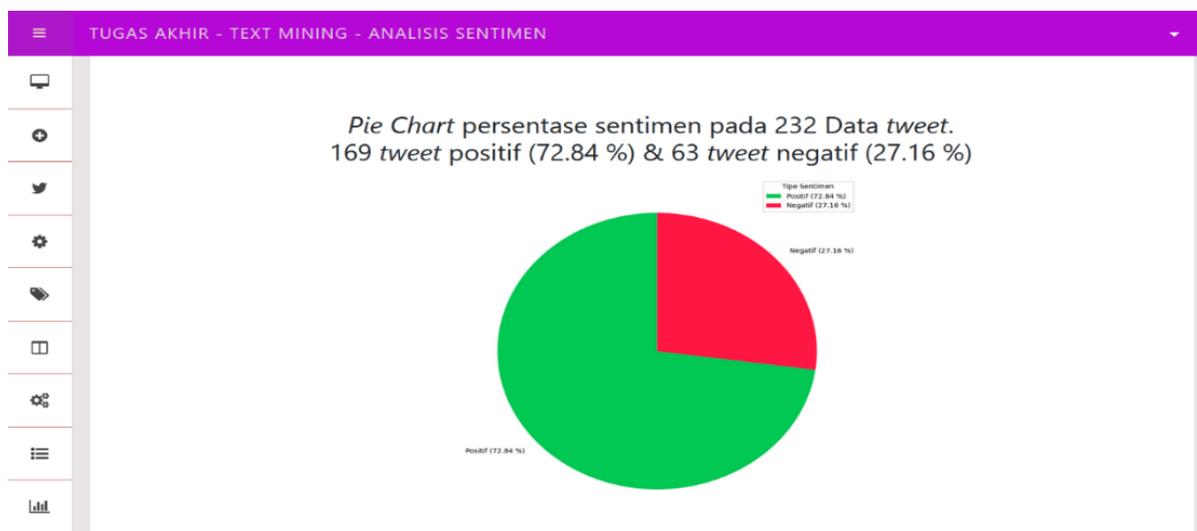
Pengujian		
Akurasi	=	$\frac{10}{23}$ 0.43 (43%)
Presisi	=	$\frac{7}{20}$ 0.35 (35%)
Recall	=	$\frac{7}{7}$ 1 (100%)

Pengujian di atas dilakukan dengan variasi nilai K yang berbeda, sehingga hasil pengujian secara keseluruhan dapat dilihat seperti pada Tabel 5 berikut ini.

Tabel 5. Hasil pengujian dan evaluasi

	K=3	K=5	K=7	K=9	K=11
Akurasi	0.30	0.30	0.30	0.43	0.35
Presisi	0.30	0.30	0.30	0.35	0.32
Recall	1	1	1	1	1

Berdasarkan tabel 5 di atas diketahui hasil pengujian, dapat diketahui bahwa hasil pengujian menunjukkan bahwa algoritma KNN mampu memperoleh nilai pengujian tertinggi menggunakan nilai K=9, dengan akurasi 43%, presisi 35%, dan *recall* 100%. Sementara itu hasil analisis sentiment terhadap 232 tweet menunjukkan arah pandangan (sentiment) masyarakat Indonesia cenderung positif 72.84%, sementara sentimen negatif sebesar 27.16 % pada periode Maret hingga April 2023.



Gambar 3. Pie Chart

4. KESIMPULAN

Berdasarkan 232 tweet, arah pandangan (sentimen) masyarakat Indonesia terhadap penyelenggaraan Piala Dunia U-20 2023 di Indonesia cenderung ke arah sentimen positif sebesar 72.84%, sementara sentimen negatif sebesar 27.16% pada periode Maret hingga April 2023. Langkah-langkah utama dalam penelitian ini meliputi: scraping, preprocessing, pelabelan, pemodelan, pembagian data, dan klasifikasi K-Nearest Neighbor (KNN). Langkah preprocessing yang baik akan menjadi faktor penentu dalam membentuk hasil yang optimal untuk langkah selanjutnya. Menggunakan ekstraksi CountVectorizer dan algoritma K-Nearest Neighbor (KNN) dalam analisis sentiment dapat bekerja dengan baik, dengan skor tes dan peringkat tertinggi dicapai oleh: Akurasi 43%, akurasi 35% dan pemulihan 100% dengan K=9.

DAFTAR PUSTAKA

- [1] M. Afdal and L. Rahma Elita, “Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, pp. 78–87, 2022.
- [2] M. Furqan, S. Sriani, and S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [3] M. S. Alrajak, I. Ernawati, and I. Nurlaili, “Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN),” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 1, no. 2, pp. 110–122, 2020.
- [4] N. A. Rakhmawati, R. B. Waskitho, D. A. Rahman, and M. F. A. U. Nuha, “Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 78–83, 2021, doi: 10.26740/jieet.v5n2.p78-83.
- [5] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [6] N. Nofiyani and W. Wulandari, “Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1621, 2022, doi: 10.30865/mib.v6i3.4332.
- [7] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, “Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi,” *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 239, 2020, doi: 10.26418/justin.v8i3.39118.
- [8] M. Priandi and Painem, “Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 311–319, 2021.
- [9] 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.
- [10] I. A. Nikmatun and I. Waspada, “Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.