

ANALISIS SENTIMEN MASYARAKAT TENTANG PENGARUH POLITIK IDENTITAS PADA PEMILU 2024 TERHADAP TOLERANSI BERAGAMA MENGGUNAKAN METODE K – NEAREST NEIGHBOR

Muhammad Jeral Palepa¹⁾, Nunik Pratiwi^{*2)}, Reva Qintara Rohmansa³⁾

1. Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
2. Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
3. Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; K – Nearest Neighbor; Politik Identitas; Youtube

Keywords: Identity Politics; K-Nearest Neighbor; Sentiment Analysis; Youtube

Article history:

Received 6 December 2023

Revised 3 December 2023

Accepted 2 January 2024

Available online 1 March 2024

DOI :

<https://doi.org/10.29100/jipi.v9i1.4957>

* Corresponding author.

Nunik Pratiwi

E-mail address:

npratiwi@uhamka.ac.id

ABSTRAK

Menjelang pemilu 2024 mendatang politik identitas kembali muncul sebagai polemik dan senjata utama dalam berpolitik demi mendapatkan suara. 2023 sebagai tahun politik menggunakan agama sebagai alat bagi setiap calon dalam berkampanye. Hal tersebut tentu berpengaruh terhadap toleransi beragama dalam bermasyarakat, baik sekarang maupun kedepannya. Tanggapan – tanggapan masyarakat mengenai politik identitas tersebut dapat ditemui di banyak media sosial terutama pada akun – akun youtube yang mengangkat tema politik identitas. Penelitian ini bertujuan untuk mengetahui bagaimana politik identitas mempengaruhi toleransi beragama di masyarakat melalui hasil nilai akurasi dan evaluasi dari penerapan metode KNN dalam pengujian analisis sentimen pada komentar video youtube Indonesia lawyers club yang bertemakan politik identitas. Pada penelitian ini terdapat 2000 data pada *dataset* yang akan diproses dalam tahapan *preprocessing* dan *labeling* dengan dilakukan implementasi algoritma dengan metode K – Nearest Neighbor. Pengujian menggunakan teknik *splitting* data 80 : 20 pada *range* K-1 sampai K-10 dengan *accuracy* sebesar 65% serta melakukan pengujian evaluasi hasil penelitian dengan menggunakan *confusion matrix*.

ABSTRACT

In the run-up to the 2024 elections, identity politics is once again emerging as a polemic and the main weapon in politics in order to gain votes. 2023, as a political year, uses religion as a tool for every candidate in their campaigns. This certainly has an impact on religious tolerance in society, both now and in the future. Public responses to this identity politics can be found on many social media platforms, especially on YouTube channels that raise the theme of identity politics. This study aims to determine how identity politics affects religious tolerance in society through the results of accuracy and evaluation values from the application of the KNN method in the sentiment analysis test of the comments on the YouTube video of the Indonesia Lawyers Club on the topic of identity politics. In this study, there are 2,000 data in the dataset that will be processed in the preprocessing and labeling stages by implementing the algorithm with the K-Nearest Neighbor method. The test uses the k-fold cross validation data splitting technique to obtain the best model from limited data, resulting in the best K value in the range K-1 to K-10 with an accuracy of 65%, and conducting an evaluation test of the research results using a confusion matrix.

I. PENDAHULUAN

PADA tahun 2016, terdapat sebuah peristiwa yang mengangkat isu politik identitas di Indonesia, Basuki Tjahaja Purnama (Ahok) diduga melakukan penistaan agama pada saat melakukan kunjungan kerja ke Kepulauan Seribu, Jakarta pada tanggal 27 September 2016. Atas dugaan tersebut, Basuki Tjahaja Purnama (Ahok) ditetapkan sebagai tersangka pada 16 November 2016. Hingga Basuki Tjahaja Purnama atau Ahok

mendapat putusan hakim dengan hukuman dua tahun penjara dan divonis bersalah melanggar Pasal 156 KUHP tentang tindak pidana penodaan agama terhadap sekelompok orang, khususnya agama Islam. Pasca kejadian tersebut, di mata politik, agama menjadi salah satu sarana utama untuk meraih suara atau dukungan dalam politik. Dalam bukunya Cherian George menunjukkan bagaimana isu penistaan agama digunakan oleh para *entrepreneur* politik untuk mencapai tujuan mereka, yang dimana hal tersebut terjadi dan digunakan oleh para pejabat partai politik di Indonesia dalam menjatuhkan suatu kekuasaan. Ahmad Syafi'i Ma'arif berpendapat bahwa bentuk politik identitas Indonesia didominasi oleh politik identitas berdasarkan agama, etnis, dan ideologi politik[1]. Perpecahan mendasar, kelompok suku dan nasional, memunculkan gerakan politik dan sosial yang luas. Pada kondisi ini mobilisasi ideologi diprakarsai oleh para pemimpin. Tujuannya adalah untuk merebut kekuasaan dari satu pemimpin ke pemimpin baru.

Seperti penelitian – penelitian serupa yang telah dilakukan sebelumnya, pada penelitian ini melakukan analisis sentimen terhadap opini atau pendapat masyarakat dari komentar youtube yang bertemakan politik identitas terhadap pengaruhnya ke toleransi beragama di lingkungan masyarakat. Analisis tersebut digunakan untuk mengetahui keberpihakan dari opini seseorang perihal apakah teridentifikasi golongan setuju atau tidak setuju akan persoalan yang terjadi. [2]. Analisis sentimen adalah teknik dalam mengidentifikasi sentimen yang terkandung pada teks, baik positif maupun negatif. Hasil analisis sentimen dari sistem prototipe mencapai akurasi tinggi, yaitu 75% hingga 95%, dalam mencari sentimen pada artikel berita dan halaman[3]. Analisis sentimen juga merupakan kegiatan pengelompokan polaritas pada sebuah kalimat serta menentukan apakah kalimat yang dinyatakan tersebut bersifat positif, negatif atau netral[4].

Youtube merupakan salah satu database video terbesar di dunia dengan hanya memerlukan koneksi internet dalam penggunaannya, di mana seorang penggunanya dapat menonton, komentar serta mengunggah video dengan *account* youtube mereka[5]. Pada dasarnya youtube merupakan platform praktis, gratis dan mudah dalam penggunaannya sehingga membuat youtube dapat digunakan dari anak – anak, dewasa hingga orang tua[6]. Pemilihan youtube sebagai media dalam pengambilan data pada penelitian ini adalah youtube menjadi media sosial yang paling banyak digunakan di Indonesia, mulai dari kalangan bawah, menengah hingga atas menggunakan media sosial youtube untuk kepentingan informasi maupun hiburan mereka[7]. Dengan adanya kolom komentar pada setiap unggahan video youtube tersebut menjadi wadah diskusi antara sesama pengguna atau akun yang saling memberikan tanggapan dan opini yang diharapkan sesuai dengan tema video tersebut. Pada tahun 2021 terdapat penelitian dengan metode analisis sentimen yang menggunakan youtube sebagai media sumber pengambilan data berdasarkan hasil riset dari platform manajemen media sosial HootSuit bahwa youtube menjadi media sosial dengan pengguna terbanyak, hal tersebut yang menjadi salah satu bahan dasar pertimbangan youtube menjadi pilihan sebagai sumber pengambilan data[8].

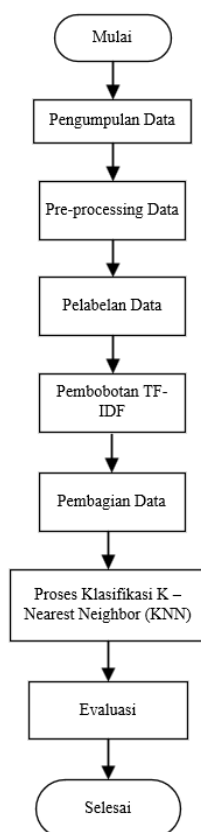
Dalam penelitian ini pengambilan data pada media youtube dengan teknik pengumpulan data youtube *crawler* pada Google Colaboratory dengan bahasa pemrograman python menggunakan API key serta id video yang akan dilakukan penarikan data komentar. Validasi data yang dihasilkan dari penarikan data menggunakan youtube *crawler* tersebut sesuai dengan komentar pada video tersebut dengan pertimbangan etika penarikan data tersebut tetap menampilkan nama akun berdasarkan komentarnya pada hasil penarikan data awal dan tidak menampilkan nama akun tersebut pada hasil akhir pemrosesan data komentar. Penelitian ini menggunakan variabel *Dependent Variable* atau variabel terikat yang akan diuji nilainya berdasarkan positif dan negatifnya kata di suatu kalimat tanpa menghitung atau mempertimbangkan jumlah *like* pada komentar[9].

Penelitian yang telah dilakukan mengenai analisis sentimen dengan metode *K – Nearest Neighbor* untuk menganalisis komentar atau *tweet* dari akun twitter terkait isu khilafah serta radikalisme di Indonesia pada tahun 2022 dengan hasil akurasi sebesar 88,2 %[2]. Algoritma KNN adalah algoritma klasifikasi dan regresi yang tidak menggunakan angka dalam data mining. Untuk menggunakan algoritma ini, diperlukan data latih dan uji pada dataset[10]. Algoritma KNN adalah proses klasifikasi yang menggunakan data pelatihan untuk menentukan kelas objek berdasarkan jarak terdekatnya[11]. *Cross validation* adalah teknik untuk menilai seberapa baik model dapat digunakan untuk memprediksi data baru, dengan cara membagi data menjadi beberapa bagian dan menggunakan sebagian data untuk melatih model, serta sebagian lagi untuk menguji model. Teknik dari *cross validation* adalah *k-fold cross validation*, dimana memecah data menjadi 'k' ke beberapa bagian set data dengan jumlah data yang sama. Penggunaan *k-fold cross validation* dapat membantu mengurangi bias dalam data, hal ini dilakukan dengan membagi data menjadi k bagian yang sama ukurannya, lalu menggunakan k-1 bagian data untuk melatih model dan 1 bagian data untuk menguji model. Proses ini dilakukan sebanyak k yang ditentukan sehingga model akan diuji menggunakan semua bagian data[12].

Tujuan penelitian ini untuk mengetahui pengaruh politik identitas terhadap toleransi beragama di masyarakat melalui hasil nilai akurasi dan evaluasi dari penerapan metode KNN dalam pengujian analisis pada sentimen positif dan negatif berdasarkan komentar masyarakat pada salah satu unggahan video youtube Indonesia lawyers club yang bertemakan politik identitas, yang dimana jumlah dari sentimen positif dan negatif tersebut menjadi salah satu hasil penelitian yang akan memberikan informasi terkait respon masyarakat terhadap pengaruh tidaknya politik identitas terhadap toleransi mereka dalam beragama dilingkungan masyarakat.

II. METODE PENELITIAN

Pada metode penelitian terdapat tahapan – tahapan yang dilakukan selama proses penelitian berlangsung dari proses pengambilan data hingga pada tahap akhir yaitu melakukan evaluasi dari hasil penelitian yang telah dilakukan. Gambar 1 menunjukkan tahapan - tahapan yang dilakukan pada penelitian ini.



Gambar. 1. Alur Penelitian *Pengumpulan Data*

Tahap penarikan data opini masyarakat berupa komentar pada video youtube channel Indonesia lawyers club yang bertemakan politik identitas dengan judul “DOSAKAH POLITIK IDENTITAS?? // GANJALAN UNTUK ANIES?!” sesuai dengan apa yang menjadi topik penelitian ini. Proses tahapan *crawling* dilakukan sebagai berikut:

1. Data komentar yang diambil menggunakan bahasa Indonesia
2. Komentar diambil pada salah satu unggahan video pada channel youtube Indonesia Lawyers Club yang bertemakan politik identitas dengan link dibawah.
<https://www.youtube.com/watch?v=DTS0JNRcOn4&t=4348s>
3. Data diambil secara keseluruhan berdasarkan setiap akun yang berkomentar
4. kumpulan data komentar tersebut akan menjadi dataset.

Gambar 2 menunjukkan proses pengambilan data komentar.



Gambar. 2. Ilustrasi pengambilan data

A. Pre-processing

Preprocessing adalah proses untuk membersihkan data dari hal-hal yang tidak relevan seperti kata yang tidak memiliki makna, simbol, dan angka. Selain itu, preprocessing juga dapat digunakan untuk mengubah semua huruf kapital menjadi huruf kecil, memisahkan data teks menjadi per kata, dan mengubah kata imbuhan menjadi kata dasar. Dilakukannya proses tersebut agar komputer dapat lebih mudah mengenali data sehingga hasil pelatihan menjadi lebih baik[13]. Berikut ini merupakan bentuk preprocessing yang dilakukan yaitu:

1. *Cleaning* adalah proses untuk menghilangkan kata yang tidak relevan dari dokumen. Kata yang dihilangkan adalah kata yang berfungsi sebagai tanda atau simbol, seperti mention, emoticon, hashtag, dan link URL., dan simbol atau kode karakter numeric dengan memberikan perintah def cleaningulasan berisikan karakter – karater yang akan dihilangkan.
2. *Case folding* adalah proses untuk menstandarkan huruf dalam dokumen, yaitu mengubah semua huruf kapital menjadi huruf kecil. Proses ini dapat dilakukan dengan perintah def lower text.
3. *Tokenizing* adalah proses pememisahan data teks yang berupa bab, paragraf dan kalimat menjadi kata-kata dengan perintah memasukkan package tokenizingtext.
4. Formalisasi adalah proses untuk menghilangkan karakter khusus, angka numerik, dan simbol lainnya dari data teks [12]
5. *Stopword* adalah tahap untuk memilih kata-kata penting dari dokumen yang sudah ditokenisasi, dengan menghapus kata-kata yang tidak memiliki makna dengan perintah memasukkan daftar stopwords yang berisikan kumpulan kata yang tidak memiliki makna
6. *Stemming* adalah proses menghilangkan imbuhan pada kata menjadi kata dasar.

B. Pelabelan Data

Setelah data pre-processing selesai dilakukan, data tersebut akan diklasifikasikan berdasarkan sentimennya dengan menggunakan lexicon sebagai metode pelabelan data sehingga menghasilkan nilai sentimen sebagai dasar pembagian sentimen positif dan negatif pada komentar tersebut. Klasifikasi ini dilakukan untuk menentukan nilai sentimen dari hasil data komentar. Penilaian sentimen pada komentar tersebut bersifat umum tanpa adanya kamus khusus tentang politik identitas pada lexicon.

C. Pembobotan TF – IDF

Proses pembobotan setiap kata (Term weighting) untuk mengoptimalkan kinerja analisis sentimen. TF (Term Frequency) memiliki kemampuan *balancing* kepentingan sesuai total kemunculan pada teks sedangkan IDF (Inverse Document Frequency) yaitu metode pembobotan token yang berfungsi untuk memonitor munculnya token pada himpunan teks[14].

D. Pembagian Data

Data dalam dataset dibagi menjadi dua kelompok, yaitu data latih dan data uji. Data latih digunakan untuk melatih mesin agar dapat mengenali pola, sedangkan data uji digunakan untuk menguji kinerja mesin tersebut[15]. Dalam penelitian ini, pembagian data didasarkan pada Prinsip Pareto atau aturan 80/20. Prinsip ini menyatakan bahwa 80% dari hasil berasal dari 20% penyebab[16]. Berdasarkan hasil pengujian dengan pembagian data 60:40, 70:30, 80:20 serta 90:10, tingkat akurasi tertinggi didapatkan dalam pembagian data 80:20.

E. Implementasi Klasifikasi K – Nearest Neighbor (KNN)

Selanjutnya yaitu mengklasifikasikan dataset dengan metode KNN. Tujuan utama dari algoritma ini yaitu membagi objek berdasarkan atribut dan sampel latihan. Algoritma KNN menggunakan klasifikasi kedekatan pada titik sebagai nilai perkiraan dari query instance yang baru. Pada penelitian ini menggunakan validasi model *cross validation* untuk penilaian pada kinerja KNN dengan membagi nilai K pada seluruh bagian dataset. Bagian K tersebut digunakan sebagai data uji yang digunakan dalam penilaian kinerja KNN dengan range K yang telah ditentukan.

Proses pada K-NN untuk mendapatkan *class category*:

1. Menghitung kemiripan dari data baru dengan setiap data yang telah dikategorikan (algoritma tersebut bekerja berdasarkan data baru terhadap data latih).
2. Berikutnya, data latih tersebut disusun dari data yang memiliki kemiripan paling besar dengan data yang akan dikategorikan.
3. Data dipilih sebanyak k data dengan kemiripan nilai terbesar
4. Memprediksi data baru ke dalam kategori data tersebut.

Rumus untuk perhitungan kedekatan antara dua kasus dapat dilihat pada persamaan 3 sebagai berikut:

$$\text{Similarity}(T, S) = x \frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i} \quad (1)$$

Keterangan:

- T : Kasus baru
- S : Kasus yang ada dalam penyimpangan
- n : Jumlah atribut pada setiap kasus
- i : Atribut individu antara 1 sampai n
- f : Fungsi *similarity* atribut i antara kasus T dan kasus S
- w : Nilai yang diberikan pada atribut ke i

Kedekatan nilai berada pada 0 sampai 1. Apabila kedua kasus tidak memiliki similarity maka akan bernilai 0, sebaliknya maka apabila kedua kasus memiliki similarity mutlak maka akan bernilai 1 [17].

F. Evaluasi

Hasil dari implementasi dengan klasifikasi algoritma K – Nearest Neighbor (KNN) tersebut kemudian dievaluasi pada tahap ini. Evaluasi menggunakan *confusion matrix* untuk mengetahui kerja dari model klasifikasi yang digunakan yaitu K – Nearest Neighbor dengan dasar pada tingkat akurasi yang dihasilkan. Pada proses evaluasi ini akan menunjukkan jumlah baris pada data komentar dengan perbandingan label antara data aktual dengan hasil prediksi. Selanjutnya terlebih dahulu menentukan dan menghitung jumlah kondisi true positive, false positive, false negative, dan true negative untuk setiap baris pada data komentar yang digunakan dalam menghitung untuk mendapatkan evaluasi model dengan ukuran yang digunakan pada penelitian ini yaitu *accuracy*, *precision*, *recall* dan *f1-score*. Penggunaan *confusion matrix* dalam evaluasi kinerja KNN pada penelitian ini dengan memberikan informasi tentang kesalahan pada model serta kesalahan – kesalahan yang timbul pada pengujian.

Rumus perhitungan *accuracy* pada evaluasi model.

$$\text{Accuracy} = \frac{TP+TN}{TP + TN+ FP + FN} \quad (2)$$

Rumus perhitungan *precision* pada evaluasi model.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

Rumus perhitungan *recall* pada evaluasi model.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

Rumus perhitungan *f1-score* pada evaluasi model.

$$f1 - score = 2x \frac{\text{Recall} x \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

Keterangan:

TP (True Positive) adalah nilai dari sentimen positif yang diprediksi dengan benar. TN (True Negative) adalah nilai dari sentimen negatif yang diprediksi dengan benar. FP (False Positive) adalah nilai sentimen negatif yang diprediksi sebagai sentimen positif. Dan FN (False Negative) adalah nilai sentimen positif yang diprediksi sebagai sentimen negatif [18].

III. HASIL DAN PEMBAHASAN

A. Pengambilan Data

Proses pengambilan data menggunakan *youtube crawler* dengan proses crawling menggunakan media Google Colaboratory dan python sebagai bahasa pemrograman. Penelitian ini menggunakan data berupa komentar youtube dari beberapa akun pada unggahan video *channel* youtube Indonesia lawyers club pada desember 2022 yang bertemakan politik identitas dengan judul “DOSAKAH POLITIK IDENTITAS?? // GANJALAN UNTUK ANIES?!” yang penarikan data komentar dimulai dari bulan desember 2022 sampai november 2023 dengan total penarikan data sebanyak 9697 data dengan pemilihan data kembali secara manual berdasarkan isi komentar yang memiliki kata kunci identitas sehingga didapat data sejumlah 2142 data. Gambar 3 menunjukkan proses dari pengambilan data dengan *youtube crawler*.

```
# isikan dengan api key Anda
api_key = 'AIzaSyA3hhNZcIZetBCPn5n8usGZHb1cg3tyXag'

#]Enter video id
# contoh url video = https://www.youtube.com/watch?v=5tucmKjOGi8 https://youtu.be/LHo1czHzTB4?si=orMOD30WaJHZR6ov
video_id = "DTS0JNRcOn4" #isikan dengan kode / ID video

# Call function
comments = video_comments(video_id)

comments
```

Gambar. 3. Pengambilan Data Dengan Youtube Crawler

Hasil dari pengambilan data awal komentar dengan menampilkan nomor komentar, tanggal komentar tersebut saat di kirim, nama akun serta komentar yang di buat. Tabel 1 menunjukkan hasil awal dari pengambilan data.

TABEL I
 PENGAMBILAN DATA

No	Date	Username	comment
1.	2023-10	*****	Politik tanpa identitas itu palsu yg rugi nasib rakyat.
2.	2023-09	*****	Macam - Macam Identitas: 1. Identitas pekerjaan 2. Identitas kesukuan 3. Identitas bangsa 4. Identitas organisasi 5. Identitas alumni 6. Identitas paguyuban 7. Identitas agama 8. Identitas jenis kelamin 9. dan lain lain
3.	2023-08	*****	Teori identitas dalam berpolitik adalah bagaikan air dan minyak.

B. Pre-processing

Preprocessing adalah proses pembersihan data yang telah diperoleh. Proses ini untuk menghilangkan karakter yang tidak relevan, seperti tanda baca dan simbol, serta mengganti huruf kapital menjadi huruf kecil [16]. Proses preprocessing pada penelitian ini menggunakan Google Colaboratory dengan Bahasa pemrograman Python. Berikut tahapan – tahapan dalam proses preprocessing data pada penelitian ini:

1) Cleaning

Tahap pertama dalam preprocessing yaitu Cleaning pada data. Pembersihan data dilakukan untuk menghilangkan karakter non-teks, seperti angka, tanda baca, link, hashtag, mention, dan emoticon. Proses cleaning menghilangkan karakter yang tidak relevan dari kata dan menggantinya dengan karakter spasi [18]. Kata-kata yang telah dibersihkan siap untuk diolah pada proses selanjutnya. Tabel 2 menunjukkan hasil dari data setelah dan sesudah proses *cleaning*.

TABEL II
 HASIL PROSES *CLEANING*

Full_text	Clean_text
ha,,jokowi omdo dia sendiri lakuin politik identitas. mana bukti nya...? bukti nya ketika dia jadi imam sholat terus di sorot kamera itu	ha jokowi omdo dia sendiri lakuin politik identitas mana bukti nya bukti nya ketika dia jadi imam sholat terus di sorot kamera itu
Terus argumentasi pelarangan penggunaan identitas seperti apa yang harus ditabukan dalam perpolitikan dan proses pencarian kepemimpinan di Indonesia? Maksudnya kaum muslim tidak boleh memilih pemimpin muslim, sesuai ajaran agamanya?	Terus argumentasi pelarangan penggunaan identitas seperti apa yang harus ditabukan dalam perpolitikan dan proses pencarian kepemimpinan di Indonesia br Maksudnya kaum muslim tidak boleh memilih pemimpin muslim sesuai ajaran agamanya

2) Case folding

Case folding untuk mengubah karakter pada tanggapan menjadi karakter huruf kecil seluruhnya (lower-case).

Tabel 3 menunjukkan hasil dari data setelah dan sesudah proses *case folding*.

TABEL III
 HASIL PROSES CASE FOLDING

Clean_text	Text_lower
Politik identitas kok takut malah jd kaya identitas phobia Misal negara A mayoritas Islam ya wajar cari pemimpin yg agama islamnya kental kalau mayoritas nasrani wajar pula cari yg nasraninya kental yg ga wajar itu mengatakan anti identitas tp gaetnya suhunya identitas Misal slogan anti politik nasrani gaetnya malah pendeta nah itu yg munafik	politik identitas kok takut malah jd kaya identitas phobia misal negara a mayoritas islam ya wajar cari pemimpin yg agama islamnya kental kalau mayoritas nasrani wajar pula cari yg nasraninya kental yg ga wajar itu mengatakan anti identitas tp gaetnya suhunya identitas misal slogan anti politik nasrani gaetnya malah pendeta nah itu yg munafik
Yang pilitik Identitas itu Ahok Jauh sebelum Anis masuk DKI Jakarta Lihat rekam hejak ahok pada tahun jauh kasus Ahok yg kedua ttg Q S Al Maidah	yang pilitik identitas itu ahok jauh sebelum anis masuk dki jakarta lihat rekam hejak ahok pada tahun jauh kasus ahok yg kedua ttg qs al maidah

3) Tokenizing

Proses *tokenizing* untuk memisahkan teks menjadi bagian-bagian yang lebih kecil atau per kata. Proses ini juga menghilangkan karakter selain huruf. Tabel 4 menunjukkan hasil dari data setelah dan sesudah proses *tokenizing*.

TABEL IV
 HASIL PROSES TOKENIZING

Text_lower	Comment_tokenized
terus argumentasi pelarangan penggunaan identitas seperti apa yang harus ditabukan dalam perpolitikan dan proses pencarian kepemimpinan di indonesia br maksudnya kaum muslim tidak boleh memilih pemimpin muslim sesuai ajaran agamanya	['terus', 'argumentasi', 'pelarangan', 'penggunaan', 'identitas', 'seperti', 'apa', 'yang', 'harus', 'ditabukan', 'dalam', 'perpolitikan', 'dan', 'proses', 'pencarian', 'kepemimpinan', 'di', 'indonesia', 'br', 'maksudnya', 'kaum', 'muslim', 'tidak', 'boleh', 'memilih', 'pemimpin', 'muslim', 'sesuai', 'ajaran', 'agamanya']
gk salah ngomong tuh presiden smua politik pasti ada identitas pinter dikit napa rakyat sudah bisa menilai kenapa pk presiden pidato sprti itu karena mau menjegal anies spy a anies tidk lolos ikut capres klo gk ada identitas berti tidak ada golongan partai dong tidak ada warna merah kuning hijau agama suku nama orang itu namanya identitas pak pres	['gk', 'salah', 'ngomong', 'tuh', 'presiden', 'smua', 'politik', 'pasti', 'ada', 'identitas', 'pinter', 'dikit', 'napa', 'rakyat', 'sudah', 'bisa', 'menilai', 'kenapa', 'pk', 'presiden', 'pidato', 'sprti', 'itu', 'karena', 'mau', 'menjegal', 'anies', 'spya', 'anies', 'tidk', 'lolos', 'ikut', 'capres', 'klo', 'gk', 'ada', 'identitas', 'berti', 'tidak', 'ada', 'golongan', 'partai', 'dong', 'tidak', 'ada', 'warna', 'merah', 'kuning', 'hijau', 'agama', 'suku', 'nama', 'orang', 'itu', 'namanya', 'identitas', 'pak', 'pres']

4) Formalisasi

Proses ini mengubah kata yang tidak baku menjadi kata baku berdasarkan kamus baku untuk menghindari duplikasi kata dengan arti yang sama. Tabel 5 menunjukkan hasil dari data setelah dan sesudah proses formalisasi.

TABEL V
 HASIL PROSES FORMALISASI

Comment_tokenized	Comment_formalisasi
['itu', 'argumen', 'arya', 'bima', 'justru', 'identitas', 'dia', 'yg', 'jauh', 'lebih', 'benderang', 'dgn', 'dalih', 'nya']	['itu', 'argumen', 'arya', 'bima', 'justru', 'identitas', 'dia', 'yang', 'jauh', 'lebih', 'benderang', 'dengan', 'dalih', 'nya']
['menuduh', 'politik', 'identitas', 'tapi', 'nyatanya', 'mereka', 'lebih', 'licik', 'dengan', 'melakukan', 'identitas', 'politik', 'br', 'mereka', 'menuduh', 'kubu', 'sebelah', 'melakukan', 'politik', 'identitas', 'untuk', 'menyembunyikan', 'identitas', 'politik', 'mereka', 'itulah', 'kelicikan', 'kelompok', 'merah']	['menuduh', 'politik', 'identitas', 'tapi', 'nyatanya', 'mereka', 'lebih', 'licik', 'dengan', 'melakukan', 'identitas', 'politik', 'baru', 'mereka', 'menuduh', 'kubu', 'sebelah', 'melakukan', 'politik', 'identitas', 'untuk', 'menyembunyikan', 'identitas', 'politik', 'mereka', 'itulah', 'kelicikan', 'kelompok', 'merah']

5) Stopword

Kata-kata dari hasil tokenisasi dibandingkan pada kata yang terdapat dalam database *stopword* kemudian kata-kata yang sama dengan kata *stopword* akan dihilangkan dari hasil tokenisasi. Proses tersebut bertujuan untuk menghilangkan kata - kata yang tidak relevan dalam analisis sentimen. Menggunakan fungsi `stopwords.words('Indonesia')` pada library NLTK sebagai dasar dalam menghapus kata yang tidak berpengaruh atau tidak dibutuhkan. Tabel 6 menunjukkan hasil dari data setelah dan sesudah proses *stopword*.

TABEL VI
 HASIL PROSES STOPWORD

Comment_formalisasi	Comment_stopword
['justru', 'yang', 'kita', 'pilih', 'mereka', 'yang', 'punya', 'identitas', 'yaitu', 'nasionalis', 'agamis', 'bukan', 'komunis', 'jangan', 'pilih', 'mereka', 'yang', 'bodong', 'alias', 'gak', 'punya', 'identitas']	['pilih', 'identitas', 'nasionalis', 'agamis', 'komunis', 'pilih', 'bodong', 'alias', 'gak', 'identitas']
['politik', 'identitas', 'akal', 'akalan', 'memojokkan', 'islam', 'mempersempit', 'ruang', 'gerak', 'isiam', 'lebih', 'berlaku', 'hanya', 'untuk', 'islam', 'sementara', 'non', 'islam', 'bebas', 'melakukan', 'politik', 'identitas']	['politik', 'identitas', 'akal', 'akalan', 'memojokkan', 'islam', 'mempersempit', 'ruang', 'gerak', 'isiam', 'berlaku', 'islam', 'non', 'islam', 'bebas', 'politik', 'identitas']

6) Stemming

Pada proses *stemming* yaitu mengubah kata imbuhan menjadi kata dasar. Proses ini bertujuan untuk memudahkan pada proses pembobotan. Proses ini menggunakan algoritma Nazief dan Adriani yang terdapat dalam library Sastrawi.

Tabel 7 menunjukkan hasil dari data setelah dan sesudah proses *stemming*.

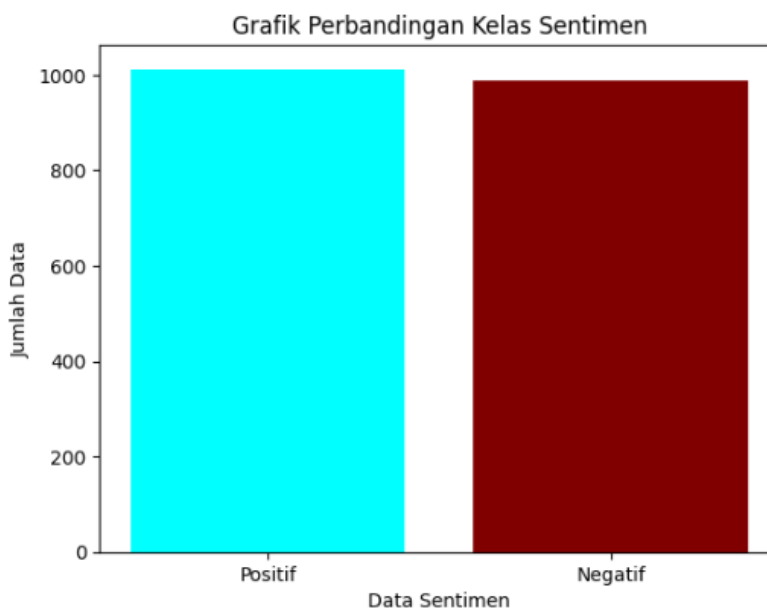
TABEL VII
 HASIL PROSES STEMMING

Comment_stopword	Stem_text
['narasi', 'politik', 'udentitas', 'judul', 'hembuskan', 'oligarki', 'menghabisi', 'orang', 'menguasai', 'si', 'penguasa', 'kelak', 'logikanya', 'kitaau', 'memilih', 'pemimpin', 'memiliki', 'identitas', 'orang', 'bodoh', 'menelan', 'bulat', 'hembusan', 'oligarki']	narasi politik udentitas judul hembus oligarki habis orang kuasa si kuasa kelak logika kitaau pilih pemimpin milik identitas orang bodoh tel bulat hembus oligarki
['menjelekan', 'jelek', 'politik', 'identitas', 'mreka', 'mengutamakan', 'kepentingan', 'kelompok']	jelek jelek politik identitas mreka utama penting kelompok

C. Pelabelan Data

Penggunaan lexicon dalam pelabelan data pada penelitian ini dengan memberikan nilai sentimen pada setiap komentar yang nantinya digunakan sebagai dasar pembagian sentimen positif dan negatif pada komentar tersebut. Pemberian nilai pada komentar tersebut bersifat umum dengan menggunakan kamus lexicon dalam menemukan kata baku dan tidak adanya konteks khusus seperti penggunaan kamus khusus politik identitas. Pelabelan data dengan membagi data menjadi kelas sentimen positif dan negatif dengan hasil data positif sejumlah 1012 dan negatif sejumlah 988 data.

Gambar 4 menunjukkan grafik perbandingan sentimen positif dan negatif.



Gambar. 4. Perbandingan Sentimen Positif dan Negatif

D. Pembobotan TF-IDF

Proses selanjutnya yaitu pemberian nilai terhadap setiap term pada komentar yang telah melewati proses sebelumnya yaitu Preprocessing. Tujuan dari pembobotan ini yaitu pemberian nilai pada term yang dimana nilai tersebut akan digunakan nanti sebagai input pada proses klasifikasi. Gambar 5 menunjukkan hasil pembobotan dengan menggunakan TF – IDF.

term	TF	TF-IDF
0 0	{"wajar": 0.0625, "jokowi": 0.0625, "beru...	{"wajar": 0.28782313662425574, "jokowi": 0...
1 1	{"komentar": 0.025, "hukum": 0.075, "keli...	{"komentar": 0.14978661367769955, "hukum": "...
2 2	{"identitas": 0.030303030303030304, "mainan...	{"identitas": 0.0007516702702718543, "maina...
3 3	{"pdip": 0.2, "tidak": 0.2, "menggunakan"...	{"pdip": 0.767940468769704, "tidak": 0.182...
4 4	{"politik": 0.16666666666666666, "identitas"...	{"politik": 0.06329956022659777, "identitas"...
...
1995 1995	{"pakai": 0.16666666666666666, "identitas"...	{"pakai": 0.47745066852459783, "identitas"...
1996 1996	{"aria": 0.11111111111111111, "bima": 0.111...	{"aria": 0.4666338975422141, "bima": 0.399...
1997 1997	{"hampir": 0.08333333333333333, "rasakan"...	{"hampir": 0.46012174315518717, "rasakan"...
1998 1998	{"memang": 0.125, "anies": 0.0625, "ident...	{"memang": 0.44473139896344155, "anies": 0...
1999 1999	{"pdahal": 0.045454545454545456, "diwilayah"...	{"pdahal": 0.3139888763173699, "diwilayah"...

2000 rows x 3 columns

Gambar. 5. Hasil Pembobotan TF-IDF

E. Klasifikasi dan Pengujian

Proses klasifikasi K – Nearest Neighbor ini membutuhkan nilai pada setiap term yang diperoleh dari proses pembobotan sebelumnya. Proses klasifikasi KNN menggunakan validasi model *k-fold cross validation*. Gambar 6 menunjukkan proses klasifikasi dengan KNN.

```
#splitting data untuk data train dan data test
from sklearn.model_selection import train_test_split
#pengujian dengan perbandingan 20:80
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=80)

from sklearn.neighbors import KNeighborsClassifier
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import cross_val_score
from sklearn.metrics import accuracy_score, classification_report

# Inisialisasi vektorisasi TF-IDF dan mengubah data pelatihan
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)

# Tentukan nilai k yang akan dievaluasi
k_values = range(1, 10)
best_accuracy = 0
best_k = 0

for k in k_values:
    knn = KNeighborsClassifier(n_neighbors=k)

    # Lakukan validasi silang k-fold
    scores = cross_val_score(knn, X_train_tfidf, y_train, cv=5, scoring='accuracy')

    # Hitung akurasi rata-rata dari validasi silang
    avg_score = scores.mean()

    # Pilih nilai k terbaik berdasarkan hasil evaluasi
    if avg_score > best_accuracy:
        best_accuracy = avg_score
        best_k = k

# Latih model KNN dengan nilai k terbaik pada subset pelatihan
knn = KNeighborsClassifier(n_neighbors=best_k)
knn.fit(X_train_tfidf, y_train)

# Ubah data pengujian menggunakan vektorisasi TF-IDF yang telah dilatih
X_test_tfidf = vectorizer.transform(X_test)

# Uji model pada subset pengujian dan evaluasi
y_pred = knn.predict(X_test_tfidf)
accuracy = accuracy_score(y_test, y_pred)

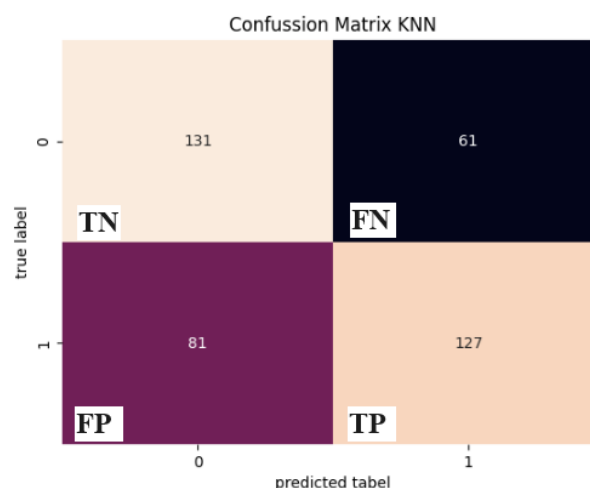
# Cetak nilai k terbaik dan akurasi pada subset pengujian
print(f"Nilai k terbaik: {best_k}")
print(f"Akurasi pada subset pengujian: {accuracy}")
```

Gambar. 6. Proses Klasifikasi KNN

Pengujian ini dilakukan dengan memasukkan data latih dan uji yang ditentukan yaitu 80:20 sebagai *splitting* data. Nilai K yang akan di evaluasi yaitu dengan *range* k-1 hingga k-10 dengan hasil yang didapat nantinya berupa nilai K terbaik yaitu k-9.

F. Evaluasi

Evaluasi kinerja dari metode K – Nearest Neighbor dengan k-9 menggunakan *confusion matrix* berdasarkan data uji. Pada evaluasi menggunakan *confusion matrix* ini didapatkan empat nilai yang terbagi menjadi True Positif, True Negative, False Positif dan False Negative. Hasil evaluasi dengan *confusion matrix* terhadap algoritma KNN dari pengujian dengan data latih 80% dan data uji 20%. Gambar 7 menunjukkan hasil dari evaluasi dengan metode *confusion matrix*.



Gambar. 7. Hasil Evaluasi Confusion Matrix

Dari nilai *confusion matrix* yang diperoleh diatas maka menghasilkan laporan klasifikasi hasil evaluasi dari kinerja KNN yang terbagi dari nilai *precision*, *recall* dan *f1-score* tersebut. Gambar 8 menunjukkan hasil evaluasi dari kinerja pada klasifikasi KNN.

Classification Report:				
	precision	recall	f1-score	support
-1	0.68	0.62	0.65	212
1	0.61	0.68	0.64	188
accuracy			0.65	400
macro avg	0.65	0.65	0.64	400
weighted avg	0.65	0.65	0.65	400

Gambar. 8. Hasil Evaluasi Kinerja KNN

Perhitungan *accuracy* dari hasil pengujian klasifikasi.

$$Accuracy = \frac{TP+TN}{TP + TN+ FP + FN} \quad (6)$$

$$Accuracy = \frac{127+131}{127 + 131+ 81 + 61} = 0,645 \approx 0,65 \quad (7)$$

Perhitungan *precision* dari hasil pengujian klasifikasi.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Precision = \frac{127}{127 + 81} = 0,61 \quad (9)$$

Perhitungan *recall* dari hasil pengujian klasifikasi.

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$Recall = \frac{127}{127 + 61} = 0,68 \quad (11)$$

Perhitungan *f1-score* dari hasil pengujian klasifikasi.

$$f1 - score = 2x \frac{Recall \times Precision}{Recall + Precision} \quad (12)$$

$$f1 - score = 2x \frac{0,61 \times 0,68}{0,61 + 0,68} = 0,64 \quad (13)$$

Keterangan:

TN : Hasil dari 400 data yang diuji, terdapat 131 data yang menjadi True Negative (TN).

TP : Hasil dari 400 data yang diuji, terdapat 127 data yang menjadi True Positive (TP).

FN : Hasil dari 400 data yang diuji, terdapat 61 data yang menjadi False Negative (FN).

FP : Hasil dari 400 data yang diuji, terdapat 81 data yang menjadi False Positive (FP).

Analisis sentimen yang digunakan sebagai metode pada penelitian kali ini yaitu sebagai cara untuk memahami apakah data dari opini atau komentar tersebut bersifat positif atau negatif. Hasil dari analisis sentimen dengan menggunakan algoritma KNN pada sentimen negatif tersebut mendapatkan hasil *precision* yaitu 68%, *recall* 62% dan *f1-score* 65% dan pada sentimen positif dengan hasil *precision* 61, *recall* 68% dan *f1-score* 64%. Keterbatasan kamus kata menjadi penyebab kurangnya kinerja pada pra-pemrosesan data dikarenakan pada isi komentar – komentar dari akun youtube tersebut banyak menggunakan kata yang tidak baku, kata gaul serta kata yang telah disingkat secara tidak umum. Perbandingan hasil dari penelitian serupa yang dilakukan sebelumnya yaitu analisis sentimen menggunakan metode KNN dengan 2000 data *tweet* diperoleh hasil dengan range nilai K=1 hingga k=30 dengan hanya bilangan ganjil pada setiap percobaan k tersebut maka didapatkan akurasi 80% pada k=23[19]. Alasan accuracy dibawah penelitian perbandingan.

IV. KESIMPULAN

Berdasarkan hasil kesimpulan dari analisis sentimen menggunakan metode KNN dengan dataset berupa komentar dari akun – akun pada sebuah video youtube yang bertemakan politik identitas memiliki hasil nilai akurasi yang relatif cukup baik. Hasil sentimen dari dataset yang cukup *balance* pada sentimen positif sejumlah 1012 dan sentimen negatif sejumlah 988. Hasil accuracy pada subset pengujian dengan algoritma KNN ini mendapatkan nilai accuracy sebesar 65% dengan nilai K terbaik dari *range* k-1 hingga k-10 yaitu k-9. pada *splitting* data perbandingan latih dan uji sebesar 80:20 dengan *random state* 80. Hasil dari analisis sentimen ini diperoleh bahwa politik identitas berpengaruh terhadap toleransi beragama di masyarakat berdasarkan dari komentar – komentar yang didapatkan dalam pengumpulan data serta terdapat *balancing* pada hasil perbandingan sentimen positif dan negatif. Dengan banyaknya sentimen negatif dari penelitian ini dapat disimpulkan apabila terdapat banyak masyarakat yang setuju akan dampak negatif terhadap toleransi beragama dari politik identitas menjelang pemilu 2024 ini serta terdapat juga komentar bersentimen positif yang beranggapan bahwa politik identitas merupakan bagian dari politik itu sendiri dan tidak berpengaruh terhadap toleransi beragama.

DAFTAR PUSTAKA

- [1] [1] D. Ardiyanti, A. M. G. P. Laia, and M. H. Nabiyin, "Demokrasi, Penegakan Hukum dan Politik Identitas di Indonesia," *Resolusi J. Sos. Polit.*, vol. 2, no. 2, pp. 94–106, 2019, doi: 10.32699/resolusi.v2i2.974.
- [2] [2] L. Legito et al., "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 324–330, Nov. 2023, doi: 10.57152/malcom.v3i2.893.
- [3] [3] P. Pasek, O. Mahawardana, G. M. A. Sasmita, P. A. H. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *JITTER J. Ilm. Teknol. dan Komput.*, 2022, [Online]. Available: <https://api.semanticscholar.org/CorpusID:264186311>
- [4] [4] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [5] [5] I. P. Hendika Permana, "Analisis Rasio Pada Akun Youtube Untuk Penelitian Kualitatif Menggunakan Metode Eksploratif," *J. Ilm. Media Sisfo*, vol. 15, no. 1, pp. 40–48, Apr. 2021, doi: 10.33998/mediasisfo.2021.15.1.970.
- [6] [6] D. A. Lestari and T. W. Apoko, "Efektivitas Video Animasi melalui YouTube terhadap Minat Belajar Bahasa Indonesia pada Siswa Sekolah Dasar," *J. Basicedu*, vol. 6, no. 4, pp. 5953–5960, May 2022, doi: 10.31004/basicedu.v6i4.3180.
- [7] [7] L. Indrianiingsih and B. Budiarsih, "ANALISIS HUKUM KONTEN NEGATIF DI PLATFORM YOUTUBE DI INDONESIA," *Bur. J. Indones. J. Law Soc. Gov.*, vol. 2, no. 3, pp. 892–916, Dec. 2022, doi: 10.53363/bureau.v2i3.71.
- [8] [8] F. Fauzi Abdulloh and I. R. Pambudi, "Analisis sentimen pengguna youtube terhadap program vaksin covid-19," *Csrid (computer Sci. Res. its Dev. journal)*, vol. 13, no. 3, pp. 141–148, 2021.
- [9] [9] A. R. Widangsa and A. R. Pratama, "Analisis Sentimen Kebijakan Pendidikan Di Masa Pandemi COVID-19 dengan CrowdTangle di Facebook," *AUTOMATA*, vol. 2, no. 2, 2021.
- [10] [10] N. Nuraeni, "Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 01, pp. 65–75, 2021.
- [11] [11] Q. A. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, Mar. 2023, doi: 10.57152/ijirse.v3i1.484.
- [12] [12] R. Tuntun, K. Kusriani, and K. Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 4, p. 2111, Oct. 2022, doi: 10.30865/mib.v6i4.4681.
- [13] [13] R. Naquitasia, D. H. Fudholi, and L. Iswari, "Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning," *J. Teknoinfo*, 2022, [Online]. Available: <https://api.semanticscholar.org/CorpusID:250402144>
- [14] [14] O. I. Gifari, M. Adha, I. R. Hendrawan, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022.
- [15] [15] W. P. P. Soenjoto, "Eksplorasi Isu Politik Identitas terhadap Identitas Politik pada Generasi Milenial Indonesia di Era 4.0," *J. Islam. Stud. Humanit.*, vol. 4, no. 2, pp. 187–217, 2019, doi: 10.21580/jish.42.5223.
- [16] [16] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [17] [17] Y. D. Atma and A. Setyanto, "Perbandingan Algoritma C4. 5 Dan K-Nn Berbasis Fitur Seleksi Forward Selection Dalam Identifikasi Mahasiswa Berpotensi Drop Out," *Metik J.*, vol. 2, no. 2, pp. 31–37, 2018.
- [18] [18] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, May 2021, doi: 10.30812/matrik.v20i2.1130.

- [19] [19] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” J. Intell. Syst. Comput., vol. 1, no. 1, pp. 43–49, Aug. 2019, doi: 10.52985/insyst.v1i1.36.