JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 11 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 *DOI: https://doi.org/10.23960/ jitet.v11i3.3150*

KLASIFIKASI PENGGUNA HASHTAG PADA APLIKASI TIKTOK MENGGUNAKAN PERBANDINGAN METODE *K-NEAREST NEIGHBORS* DAN *NAÏVE BAYES CLASSIFIER*

Moh. Aulia Miftakhurahmat[1*], Nur Safitri[2], Putri Aulia Kusnadi[3], Chaerur Rozikin[4]

[1,2,3,4] Informatika; Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Karawang

Riwayat artikel:
Received: 3 Juni 2023
Accepted: 10 Juli 2023
Published: 1 Agustus 2023

Keywords:
K-Nearest Neighbors; Naive Bayes classifier; Hashtag; Tiktok.

Corespondent Email:
2010631170087@student.unsika.ac.id

Abstrak. Penelitian ini bertujuan untuk membandingkan penggunaan algoritma machine learning yaitu, *K-Nearest Neighbors* (KNN) dengan algoritma *Naive Bayes classifier* dalam memberikan rekomendasi hashtag untuk pengguna aplikasi TikTok. Dataset yang digunakan dalam penelitian ini adalah dataset hashtag TikTok yang diambil dari sebuah website yang berdasarkan dari setiap kategori. Pada penelitian ini dilakukan pemodelan dengan menggunakan algoritma *K-Nearest Neighbors* (KNN) dan Naive Bayes classifier untuk memprediksi hashtag yang sesuai untuk pengguna Tiktok berdasarkan pada hashtag yang sedang populer digunakan. Kemudian dilakukan evaluasi kinerja kedua metode dengan menggunakan *precision, recall, f1 score dan accuracy.* Pada penelitian ini penulis akan membandingkan performa klasifikasi model yang telah dibuat menggunakan metode *K-Nearest Neighbors* dan *Naive Bayes Classifier*, tujuan perbandingan kinerja ini adalah untuk mempelajari metode mana yang memiliki kinerja terbaik dalam hal merekomendasikan penggunaan hashtag. Hasil dari penelitian ini menunjukan bahwa perbandingan dari kedua metode dapat memberikan klasifikasi rekomendasi hastag yang baik dengan nilai *f1 score* dan *accuracy* yang cukup tinggi.

Abstract. *This study aims to compare the use of machine learning algorithms, namely, K-Nearest Neighbors (KNN) with the Naive Bayes classifier algorithm in providing hashtag recommendations for TikTok application users. The dataset used in this study is the TikTok hashtag dataset taken from a website based on each category. In this study, modeling was carried out using the K-Nearest Neighbors (KNN) algorithm and the Naive Bayes classifier to predict the appropriate hashtags for Tiktok users based on the currently popular hashtags used. Then evaluate the performance of both methods using precision, recall, f1 score and accuracy. In this study the authors will compare the performance of the classification models that have been created using the K-Nearest Neighbors and Naive Bayes Classifier methods, the purpose of this performance comparison is to learn which method has the best performance in terms of recommending the use of hashtags. The results of this study indicate that the comparison of the two methods can provide a good classification of hashtag recommendations with a fairly high f1 score and accuracy.*

Vol. 11 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 *DOI: https://doi.org/10.23960/ jitet.v11i3.3150*

1. PENDAHULUAN

Tiktok adalah salah satu media sosial yang paling populer dan disukai. Tiktok menawarkan kemampuan untuk membuat video dengan durasi 15 detik hingga 1 menit dengan filter, musik, dan fitur kreatif lainnya, menjadikannya lebih baik daripada media sosial lainnya. Tiktok juga memungkinkan orang dari berbagai kalangan untuk membuat konten karena kesederhanaannya dan kemudahan penggunaan [1].

Tiktok semakin populer, dengan 1,5 miliar unduhan sejak peluncuran awal tahun 2017 dan berbagai prestasi. Generasi Z lebih suka hal-hal praktis dan bergantung pada teknologi [2]. Di sisi lain, mereka lebih sering melakukan kegiatan sosial melalui dunia maya dan lebih cepat dalam mencari dan menemukan informasi [3]. Tiktok berhasil menyalip unduhan Instragram dan Facebook pada kuartil ketiga dari januari hingga September 2019 [4].

Dalam hal ini, gaya hidup mamnusia berkembang seiring perkembangan zaman. Sebagian besar masyarakat Indonesia, terutama generasi milenial, memiliki sejumlah sikap Hedon dalam kehidupan mereka. Media sosial seperti Tiktok sekarang lebih dari sekedar menampilkan video hiburan. Mereka juga menambahkan fitur baru, seperti penggunaan hashtag untuk membantu pengguna menemukan konten yang relevan. Terkadang, pengguna tidak tahu hashtag apa yang harus digunakan, jadi sistem rekomendasi hashtag sangat penting untuk membantu mereka menemukan konten yang relevan.

Studi ini akan membandingkan dua metode klasifikasi, *K-Nearest Neighbours* dan *Naïve Bayes Classifier*, untuk tugas klasifikasi penggunaan hashtag di Tiktok. Tujuan dari penelitian ini adalah untuk mengetahui keunggulan dan kelemahan masing-masing metode klasifikasi untuk tugas klasifikasi penggunaan hashtag di Tiktok.

Dalam penelitian ini, akan digunakan dataset hashtag yang dikumpulkan dari salah satu website di internet yaitu website hastag fyp tiktok 2023, dan dilakukan evaluasi terhadap kinerja dari kedua metode klasifikasi dengan menggunakan evaluasi metrik-metrik yang umum digunakan dalam tugas klasifikasi, seperti *precision, recall, f1 score dan accuracy*. Selain itu, akan dilakukan analisis terhadap faktor-faktor yang mempengaruhi kinerja dari kedua metode klasifikasi.

Hasil yang diharapkan dari penelitian ini dapat memberikan kontribusi bagi pengembangan sistem rekomendasi hashtag di aplikasi Tiktok, serta memberikan pandangan yang lebih jelas mengenai keunggulan dan kelemahan dari masing-masing metode klasifikasi dalam tugas klasifikasi rekomendasi penggunaan hashtag.

2. TINJAUAN PUSTAKA

2.1 Klasifiasi

Klasifikasi adalah metode data mining yang membantu dalam melakukan prediksi berdasarkan label kelas sampel yang diklasifikasi. Klasifikasi adalah proses analisis data yang menghasilkan model-model yang nantinya digambarkan melalui kelas-kelas yang ada dalam data. Metode ini memiliki persyaratan bahwa attribut data harus numerik atau nominal dan label data harus nominal [5].

2.2 K-Nearest Neighbors

K-Nearest Neighbor adalah metode untuk mengklasifikasikan objek berdasarkan data pelatihan dengan menggunakan jarak terdekat atau kemiripan. Algoritma ini hanya mengimpan vektor-vektor fitur dan mengklasifikasikan data pembelajaran selama fase pembelajaran. Selama fase klasifikasi, fitur-fitur yang sama dihitung untuk data test, yang klasifikasinya tidak diketahui. Setelah menghitung jarak antara vektor data pembelajaran dan vektor baru, sejumlah K yang paling dekat ditemukan. Snagat mungkin bahwa titik yang paling banyak dikasifikasikan akan termasuk dalam kumpulan titik tersebut [6]. Akurasi waktu terbaik untuk klasifikasi adalah K-NN jika dibandingkan dengan metode lain dengan hasil pemeriksaan dan aktifitas lain yang memengaruhi penetapan aturan tertentu [7].

2.3 Naïve Bayes Classifier

Metode klasifikasi *Naïve Bayes* digunakan dalam *text mining*. Ini berpotensi baik untuk klasifikasi dalam hal presisis dan komputasi data [8]. Metode ini banyak digunakan dalam teknik klasifikasi

dengan menggunakan metode seperti *Maximum Entropy Classification*, *Unigram Naïve Bayes*, dan *Multinomial Naïve Bayes* [9]. Karakteristik utama klasifikasi *Naive Bayes* adalah untuk membuat hipotesis yang kuat dari setiap situasi atau kejadian [10]. Salah satu keuntungan dari *Naive Bayes Classifier* adalah bahwa ia akan hanya membutuhkan jumlah data pelatihan yang relatif kecil untuk memperkirakan parameter yang diperlukan untuk klasifikasi, yaitu sarana dan varians dari variable-variable [11].

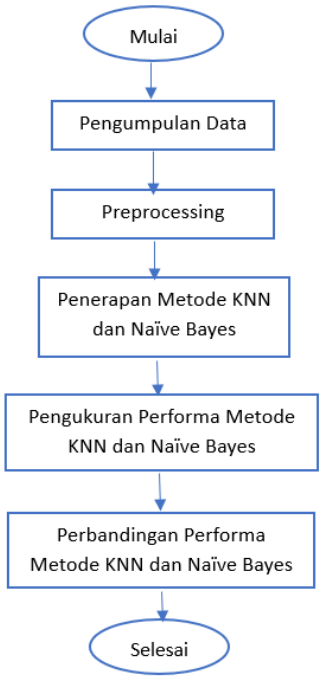
2.4 Text Mining

Text mining adalah fase menemukan informasi yang memungkinkan pengguna berinteraksi dengan berbagai dokumen melalui berbagai alat analisis [12]. Proses utama *text mining* adalah menemukan teks yang dapat menggantikan isi dokumen, dan setelah itu, proses review dilakukan untuk mengidentifikasi hubungan antara setiap dokumen**.**

3. METODE PENELITIAN

3.1 Prosedur Penelitian

Perhatikan Prosedur penelitian pada gambar 1.



Gambar 1. *Flowchart* Prosedur Penelitian

3.2 Pengumpulan data

Proses pengumpulan data memiliki peran krusial dalam penelitian ini guna menghimpun data pendukung yang sangat dibutuhkan. Dataset yang dipergunakan diambil dari beragam platform website melalui proses crawling data. Dataset tersebut diarsipkan dalam format file csv untuk kemudahan pengolahan. Dalam penelitian ini, total terdapat 359 hashtag yang menjadi bagian dari dataset yang digunakan. Hal ini menjadi esensial dalam melakukan klasifikasi untuk merekomendasikan penggunaan hashtag pada aplikasi Tiktok.

3.3 Prepocessing

Prepocessing adalah tahap pertama dalam pengolahan data input sebelum memulai proses. Ini adalah prosedur dan prosedur yang digunakan untuk mengubah data mentah menjadi data yang berkualitas tinggi dan menerima input yang baik [12]. Beberapa metode *preprocessing* termasuk menghilangkan nilai kosong atau mengisi celah dengan rata-rata.

3.4 Penerapan Metode K-Nearest Neighbor dan Naïve Bayes Classifier

Metode *K-Nearest Neighbor* menggunakan langkah-langkah penerapan berikut [13][14]: pertama, menentukan jumlah tetangga terdekat K; kedua, menghitung jarak antara data uji (test data) dan data pelatihan (training data); ketiga, mengurutkan data berdasarkan jarak geometri terkecil; dan keempat, menentukan kelompok data uji berdasarkan mayoritas label pada tetangga terdekat K.

Sedangkan untuk metode *Naïve Bayes* menggunakan langka-langkah berikut: untuk data kategorikal, hitung jumlah dan kemungkinan setiap peristiwa; kemudian buat tabel kemungkinan. Untuk data numerik, tentukan mean (rata-rata) dari setiap parameter; kemudian buat tabel mean dan standar deviasi.

3.5 Pengukuran Performa Metode K-Nearest Neighbor dan Naïve Bayes Classifier

Untuk melakukan evaluasi performa metode, kita dapat menggunakan metrik-metrik seperti *Accuracy, Recall, dan F-Score*.

1. *Accuracy*: Digunakan untuk mengukur sejauh mana nilai prediksi cocok dengan nilai sebenarnya [15].

2. *Precision* Menunjukkan seberapa akurat atau tepat model dalam memprediksi hasil positif. Presisi juga merupakan metrik yang baik untuk mengidentifikasi tingkat keberadaan nilai positif palsu pada suatu model [16].
3. *Recall*: Menghitung berapa banyak nilai positif sebenarnya yang berhasil diidentifikasi oleh model sebagai True Positive. Recall juga digunakan sebagai metrik untuk memilih model terbaik ketika tingkat nilai false negatif tinggi [17].
4. *F1-Score*: Merupakan rata-rata tertimbang antara presisi dan *recall* [18].

Pengukuran performa metode digunakan untuk mengukur seberapa baik metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier* dalam menghasilkan prediksi yang sesuai dengan nilai atau data yang sebenarnya. Tujuannya adalah untuk mengevaluasi tingkat kesamaan atau kedekatan antara hasil perhitungan metode dengan nilai sebenarnya.

3.6 Perbandingan Performa Metode K-Nearest Neighbor dan Naïve Bayes Classifier

Setelah tahap pengukuran performa selesai, metode dievaluasi untuk kinerjanya. Ini dilakukan untuk mengetahui hasil dari dua pendekatan: *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Ini dilakukan untuk menentukan kategori yang lebih baik untuk penggunaan hashtag di aplikasi Tiktok.

4 HASIL DAN PEMBAHASAN

4.1 Penerapan Metode K-Nearest Neighbors dan Naïve Bayes Classifier

Dalam penelitian ini, kami menerapkan metode *K-Nearest Neighbors* (KNN) dan *Naive Bayes Classifier* untuk melakukan pengujian prediksi penggunaan hashtag di aplikasi Tiktok. Dataset yang digunakan dalam penelitian ini diambil dari berbagai platform website melalui proses crawling data. Dataset ini terdiri dari 359 hashtag yang telah diberi label berdasarkan kategorinya. *Review* dataset sebagai berikut ditampilkan secara singkat:

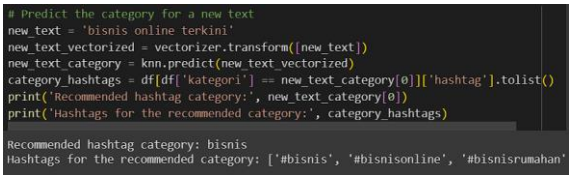
Tabel 1. Dataset Hashtag dan Kategori

no	hashtag	kategori
1	#FYP	rekomendasi
2	#trending	rekomendasi
.	.	.
.	.	.
359	#dautych beautes	kecantikan

Dataset berhasil dikumpulkan dan langkah selanjutnya adalah melakukan prepocessing data yaitu dengan fungsi *CountVectorizer(),* yaitu Mengubah teks menjadi vektor dengan menghitung frekuensi kemunculan setiap kata dalam dataset pada kolom hashtag.

4.2 Penerapan Metode K-Nearest Neighbors

Pada metode *K-Nearest Neighbors* Untuk melakukan evaluasi terhadap model yang kami bangun, kami membagi dataset menjadi dua bagian, yaitu data latih (80%) dan data uji (20%). Bagian data latih digunakan untuk melatih model, sedangkan bagian data uji digunakan untuk menguji kinerja model yang telah dilatih. Kami melakukan klasifikasi dengan menggunakan metode *K-Nearest Neighbors* dengan memasukan parameter n-neighbors = 5. Untuk penerapannya, kami memasukan sebuah kalimat dan selanjutnya akan diprediksi kalimat tersebut masuk kedalam kategori apa dan merekomendasikan hashtag yang relevan.

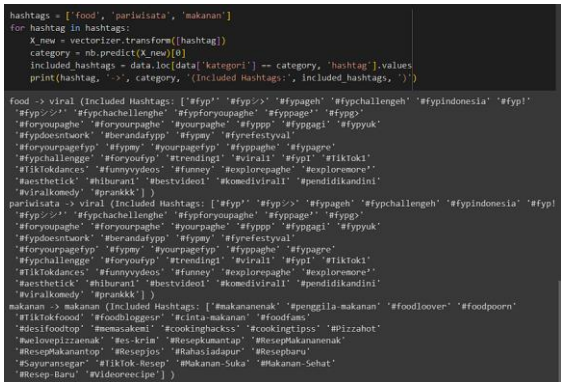


Gambar 2 .Prediksi hashtag menggunakan metode *K-Nearest Neighbors*

Dari gambar diatas dapat menggambarkan bahwa prediksi hashtag menggunakan metode *K-Nearest Neighbors* cukup berhasil.

4.3 Penerapan Metode Naïve Bayes Classifier

Pada metode *Naïve Bayes Classifier* Untuk melakukan evaluasi terhadap model yang kami bangun, kami membagi dataset menjadi dua bagian, yaitu data latih (60%) dan data uji (40%). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji kinerja model yang telah dilatih. Setelah melakukan pembagian dataset, kami melanjutkan dengan proses training pada model menggunakan metode *Naïve Bayes Classifier*. Pada metode ini, kami tidak melakukan pengaturan parameter khusus atau pencarian parameter untuk mencapai performa yang paling optimal. Hal ini disebabkan oleh karakteristik metode *Naïve Bayes Classifier* yang tidak membutuhkan pencarian parameter yang kompleks seperti pada metode *K-Nearest Neighbors.* Untuk penerapannya, kami memasukan beberapa kata dan selanjutnya akan diprediksi kata tersebut masuk kedalam kategori apa dan merekomendasikan hashtag yang relevan.



Gambar 2. Prediksi hashtag menggunakan metode *Naïve Bayes Classifier*

4.4 Pengukuran Performa Metode K-Nearest Neighbors dan Naïve Bayes Classifier

Pada titik ini, kinerja klasifikasi dinilai menggunakan metode *K-Nearest Neighbors* dan *Naïve Bayes Classifier*. Pengujian ini mencakup pengukuran performa berdasarkan *precision, recall, f1 score, dan accuracy.*

1. Performa Metode *K-Nearest Neighbors*: Tabel di bawah menunjukkan hasil model klasifikasi metode *K-Nearest Neighbors*.

Tabel 2. Performa *K-Nearest Neighbors*

Metode	Accuracy	Precision	Recall	F1 Score
K-Nearest Neighbors	0,16	0,01	0,07	0,017

2. Performa Metode *Naïve Bayes Classifier:* Tabel berikut menunjukkan hasil model klasifikasi metode *Naïve Bayes Classifier*.

Tabel 3. Performa *Naïve Bayes Classifier*

Metode	Accuracy	Precision	Recall	F1 Score
Naïve Bayes Classifier	0,98	0,98	0,98	0,98

4.5 Perbandingan Performa Metode K-Nearest Neighbors dan Naïve Bayes Classifier

Tujuan selanjutnya adalah untuk membandingkan kinerja model klasifikasi yang telah dibangun menggunakan metode *K-Nearest Neighbors* dan *Naïve Bayes* Classifier. Tujuan dari perbandingan ini adalah untuk menentukan metode mana yang menangani masalah klasifikasi rekomendasi penggunaan hashtag dengan lebih baik. Hasil perbandingan ditunjukkan dalam tabel berikut.

Tabel 4. Perbandingan Performa *K-Nearest Neighbors* dan *Naïve Bayes Classifier*

Metode	Accuracy	Precision	Recall	F1 Score
K-Nearest Neighbors	0,16	0,01	0,07	0,017
Naïve Bayes Classifier	0,98	0,98	0,98	0,98

Dari tabel diatas, menunjukkan bahwa metode *Naïve Bayes* lebih tinggi dengan tingkat *accuracy* sebesar 98% dibandingkan dengan *K-Nearest Neighbors* yang hanya mendapatkan tingkat *accuracy* 16%. Namun, tidak hanya dilihat dari tingkat *accuracy* saja tetapi dapat dilihat dari *precision*, *recall* dan terutama *f1 score* dari *naive bayes* yang lebih unggul dari *K-Nearest Neighbors* dengan selisih 96%. Jadi, dapat dikatakan bahwa untuk mengklasifikasikan rekomendasi penggunaan hashtag di aplikasi tiktok menggunakan metode *Naïve Bayes* mendapatkan hasil performa yang lebih baik.

5 KESIMPULAN

a. Berdasarkan hasil analisis keseluruhan, dapat disimpulkan bahwa perbandingan kedua metode *K-Nearest Neighbors* (KNN) dan *Naïve Bayes Classifier* memiliki kemampuan untuk melakukan klasifikasi pada data dan dapat digunakan untuk rekomendasi penggunaan hashtag pada aplikasi Tiktok. Hasil uji coba menunjukkan bahwa Naïve Bayes Classifier memiliki performa yang relatif tinggi dalam melakukan klasifikasi dan rekomendasi penggunaan hashtag pada aplikasi Tiktok.

b. Kedua metode memiliki kelebihan dan kelemahan masing-masing, KNN memerlukan waktu yang lebih lama untuk memproses data dan melakukan klasifikasi, sementara Naïve Bayes lebih cepat dalam melakukan klasifikasi pada data. Oleh karena itu, pemilihan metode yang tepat dalam melakukan rekomendasi penggunaan hashtag pada aplikasi Tiktok tergantung pada karakteristik data dan tujuan dari rekomendasi yang akan dilakukan. Penelitian lebih lanjut dapat dilakukan untuk mengoptimalkan kinerja kedua metode dan menambah wawasan perbandingan dengan metode lainnya dalam klasifikasi rekomendasi hashtag pengguna di aplikasi Tiktok.

DAFTAR PUSTAKA

[1] Putri, N. W. (2022). Persepsi mahasiswa Universitas Islam Negeri Sunan Ampel Surabaya mengenai konten lgbt di aplikasi tik tok. Novita Wardaini Putri_I73218045.pdf

[2] Turner, A. (2015). Generation Z: Technology and social interest. The Journal of Individual Psychology, 71(2), 103–113.

[3] Noordiono, A. (2016). Karakter Generasi Z Dan Proses Pembelajaran Pada Program Studi Akuntansi [Doctoral Dissertation]. Universitas Airlangga

[4] Bulele, Y. N., & Wibowo, T. (2020). Analisis Fenomena Sosial Media Dan Kaum Milenial : Studi Kasus Tiktok, 1, 565–572.

[5] F. Ariani, Amir, N. Alam, and K. Rizal, "Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naïve Bayes," Paradig. - J. Komput. dan Inform., vol. 20, no. 2, pp. 33–38, 2018, doi: 10.31294/p.v.

[6] Feldman, R. d. (2007). The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. New York.: Cambridge University Press.

[7] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," Syntax J. Inform., vol. 10, no. 01, pp. 46–55, 2021, [Online]. Available: https://journal.unsika.ac.id/index.php/syntax/article/download/5173/2749.

[8] J. M and V. H, "Opinion Mining For Sentiment Data Classification," Int. J. Res. Inf. Technol., vol. 3, no. 1, pp. 1–13,2014.

[9] R. P and M. M, "Sentiment Analysis of User Generated Twitter Updates using Various Classification," 2009.

[10] N. Rochmawati and S. C. Wibawa, "Opinion Analysis on Rohing

[11] V. Narayanan, I. Arora, and A. Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model", IDEAL.Varanasi: India, 2013, pp.194–201.

[12] J. Ling, I. P. E. N. Kencana, And T. B. Oka, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square," E-Jurnal Mat., vol. 3, no. 3, p. 92, 2014, doi: 10.24843/mtk.2014.v03.i03.p070.

[13] Razzaghi, T., Roderick, O., Safro, I., Marko, N, Multilevel weighted support vector machine for classification on healthcare data with missing values. PloS ONE, 11(5), e0155,119.P. K. L. Utama. 2018. Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning", Widya Duta Vol. 13, No. 1. hal 69-76.

[14] Ramadhan, R. S. Junta, Z., Ardytha, L. 2016. Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir. Journal of Applied Intelligent System, Vol. 1, No. 2. 123-133

[15] V. A. Permadi, "Analisis sentimen menggunakan algoritma Naïve Bayes terhadap review restoran di Singapura," Jurnal Buana Informatika, vol. 11, no. 2, pp. 141– 151, 2020.

[16] F. Septianingrum, J. H. Jaman, and U. Enri, "Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier," Jurnal Media Informatika Budidarma, vol. 5, no. 4, pp. 1431–1437, 2021.
[17] T. Nugraha, P. Purwantoro, and Y. Umaidah, "Analisis Sentimen terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan Algoritma Naïve Bayes," Jurnal Pendidikan dan Konseling (JPDK), vol. 4, no. 4, pp. 4625–4635, 2022.
[18] A. Aisyah and S. Anraeni, "Analisis penerapan metode K-Nearest Neighbor (K-NN) pada dataset citra penyakit malaria," Indonesian Journal of Data and Science, vol. 3, no. 1, pp. 17–29, 2022.