



## **Implementasi Algoritma K-Nearest Neighbors untuk Menganalisis Pendapat Pakar AI tentang Kemajuan Kecerdasan Buatan**

**Muhammad Akmal Hakim<sup>1</sup>, Hermawan<sup>2</sup>, Peri Gunawan<sup>3</sup>, Rendi Pratama<sup>4</sup>, Fandi Kurniawam<sup>5</sup>**

Teknik dan Ilmu Komputer, Universitas Muhammadiyah Kotabumi, Lampung, Indonesia.  
Sistem dan Teknologi Informasi Ilmu Komputer, Universitas Muhammadiyah Kotabumi,  
Lampung, Indonesia

E-mail: muham.2059201103@umko.ac.id<sup>1</sup>, herma.2059201025@umko.ac.id<sup>2</sup>,  
perig.2059201071@umko.ac.id<sup>3</sup>, rendipratama95061@gmail.com<sup>4</sup>,  
fandi.kurniawan@umko.ac.id<sup>5</sup>

### **Abstrak**

Perkembangan teknologi berkembang dengan sangat pesat, terutama dibidang informasi seperti media sosial. Saat ini sudah banyak platform media sosial seperti YouTube yang bisa digunakan untuk mengunggah dan menonton video serta streaming video di Internet. Video yang disalurkan di YouTube biasanya mendapat ratusan ribu, bahkan jutaan penonton dan puluhan ribu komentar dari para netizen. Ada banyak video di YouTube yang membahas tentang AI (Artificial Intelligence/Kecerdasan Buatan, dan beberapa pakar serta netizen di seluruh dunia mendukung AI, sementara yang lain menentangnya. Oleh karena itu pada penelitian ini kami menggunakan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan sentimen komentar menjadi positif dan negatif. Hasil akurasi dari menggunakan klasifikasi K-NN yaitu, akurasi yang didapat yaitu sebesar 70.82%, class precision untuk pred.positif dengan skor 57.80%, dan class.precision pred.negatif dengan skor 82.26%. dan untuk class.recall true positif dengan skor 74.12%, dan class.recall true negatif dengan skor 68.92%.

**Kata Kunci :** Analisis Sentimen, Klasifikasi, K-Nearest Neighbor

### **1. PENDAHULUAN**

Teknologi informasi berkembang sangatlah cepat, sehingga hampir dari semua kegiatan atau bahkan pekerjaan berhubungan dengan sebuah teknologi informasi. Dan dengan adanya teknologi informasi kita dengan sangat mudah mendapatkan informasi, juga dapat meringankan dan mempercepat pekerjaan seseorang.



Media sosial adalah komunitas online yang memungkinkan penggunanya berkomunikasi, berpartisipasi, dan bertukar informasi tanpa dibatasi oleh ruang dan waktu [1]. Media sosial yang umum digunakan adalah Instagram, Twitter, Facebook, YouTube dan lain-lain. Dalam media sosial banyak yang membahas tentang Artificial Intelligence (AI), Kecerdasan buatan (AI) adalah program komputer dengan algoritma yang bertugas memeriksa data dan menggunakannya untuk melakukan proses berpikir dan berperilaku seperti manusia [2]. YouTube adalah media sosial streaming video online atau situs web untuk mengunduh dan menonton video. Dalam penelitian ini dilakukan analisis opini yang mendukung dan menentang kecerdasan buatan [3].

Data mining merupakan suatu teknik menggali data secara mendalam untuk memperoleh pengetahuan atau wawasan. Data mining menggunakan pengalaman atau bahkan kesalahan sebelumnya untuk meningkatkan kualitas model dan hasil analisis, salah satunya adalah kemampuan teknik data mining atau klasifikasi [4]. Data mining merupakan suatu proses yang dapat digunakan untuk mengolah data komentar emosional dari pengguna di Twitter, Youtube, dan Instagram.

Penelitian kali ini menggunakan algoritma K-Nearest Neighbor adalah salah satu metode klasifikasi data mining. KNN adalah pengklasifikasi data mining yang mengklasifikasikan kumpulan data yang belajar dari informasi berlabel [5]. Beberapa pembahasan singkat beberapa penelitian terdahulu yang relevan untuk mendukung alasan dan motivasi perlunya dilakukan penelitian seperti yang diusulkan.

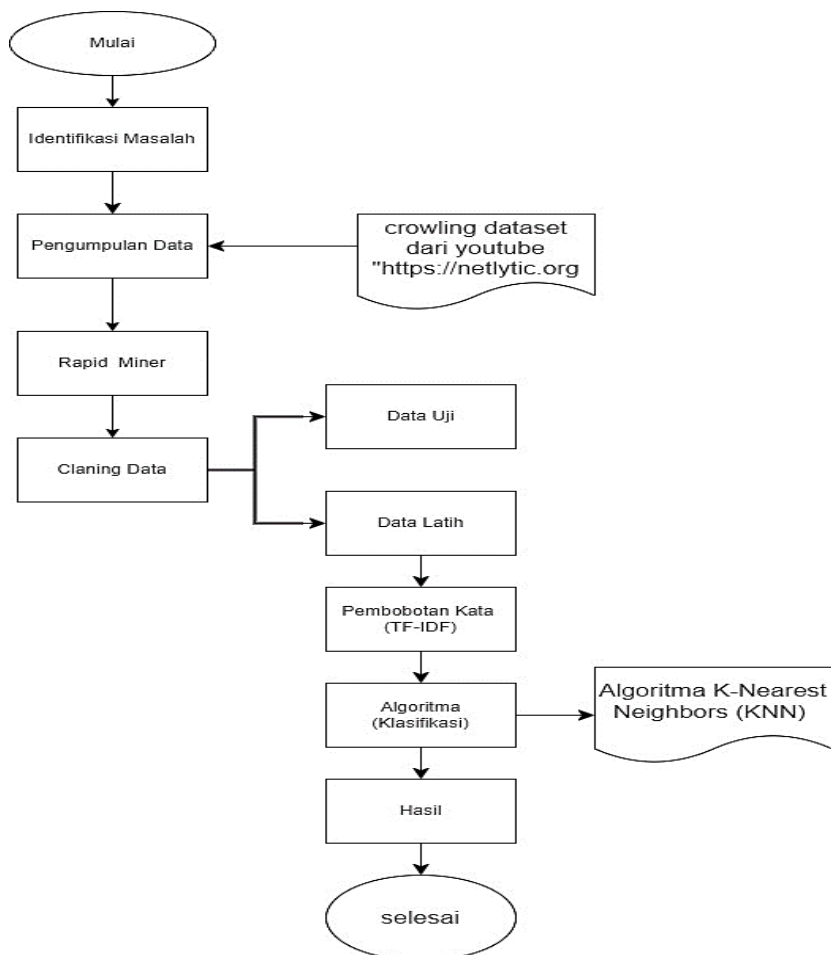
Penelitian dengan judul Metode K-Nearest Neighbor Untuk Analisis Sentimen Tentang Penghapusan Ujian Nasional [6]. Pada penelitian ini dijelaskan penelitian tersebut di ajukan untuk menganalisis opini masyarakat, ditentukan dari menganalisis sentimen pada twitter, youtube dan instagram dalam komentar masyarakat menggunakan Algoritma K-Nearest Neighbor. Hasil dari pembahasannya, di dapatkan akurasi sebesar 75,45% dan AUC sebesar 0,733%.

Berdasarkan dari penelitian sebelumnya penulis bermaksud melakukan analisis sentimen pada komentar youtube, yaitu bertujuan untuk melihat opini dari masyarakat atau warganet tentang kemajuan dari kecerdasan buatan (AI). Yang menghasilkan opini dengan kategori opini positif atau opini negatif. Peneliti berharap dalam penelitian ini dapat bermanfaat

untuk menyelidiki opini-opini masyarakat yang termasuk sentimen positif, negatif, netral, dan lainnya.

## 2. METODE

Berikut merupakan langkah-langkah yang dilakukan dalam metode penelitian, digambarkan menjadi alur penelitian bersumber dari [7] yang dijelaskan pada gambar di bawah ini:



**Gambar 1.** Alur penelitian

## **2.1 Identifikasi masalah**

Pada tahap ini peneliti melakukan mengidentifikasi masalah yang sering terjadi pada media sosial, seperti twitter, instagram, youtube dan lainnya. Data yang digunakan berasal dari opini masyarakat YouTube dan konten unduhan. Konten seputar perkembangan teknologi pun banyak disukai warganet. Oleh karena itu, perlu adanya proses untuk menentukan komentar mana yang merupakan sentimen positif dan mana yang merupakan sentimen negatif.

## **2.2 Metode Pengumpulan data**

Dalam tahap pengumpulan data, penelitian ini menggunakan data yang didapat dari dari komentar video yaoutube <https://youtu.be/YaM-Y8RT4us?si=JsVHJgIyJ0QJhqIu> data tersebut diambil menggunakan website netlytic.org.

## **2.3 Analisis sentimen**

Sentiment Analysis merupakan perpaduan dari data mining dan teks mining, dimana Sentiment Analysis sendiri digunakan untuk mengolah berbagai macam opini yang telah diberikan oleh masyarakat atau para pakar melalui berbagai media yang ada, opini tersebut diberikan untuk sebuah produk, jasa maupun sebuah instansi [8].

## **2.4 Texts mining**

Texts mining adalah sebuah analisis terhadap teks, yaitu dengan mengestrak sebuah teks informasi penting dan bermutu atau teks yang tidak terstruktur [7]. Ini bertujuan untuk mengembangkan sebuah teks yaitu mengidentifikasi peola, tren, kandungan, serta wawasan dalam teks.

## **2.5 Kasifikasi K-Nearest Neighbors**

K-Nearest Neighbors merupakan algoritma atau metode yang digunakan untuk pembelajaran mesin, yang fokus utamanya untuk pengkasifikasian atau data berdasarkan kesamaan dari tetangga terdekatnya. KNN ini berkerja denagn cara mengidentifikasi kategori dari data yang baru dengan berdasarkan mayoritas label terdekatnya.

Berikut adalah Langkah untuk menghitung metode K-Nearest Neighbor [9] sebagai berikut:

- a) Menentukan parameter K
- b) Menghitung jarak antara data training dan data testing Perhitungan jarak yang paling umum dipakai pada perhitungan pada algoritma KNN adalah menggunakan perhitungan jarak Euclidean. Rumusannya adalah sebagai berikut:

$$euc = \sqrt{(\sum_{i=0}^n (p_i - q_i)^2)}$$

dimana:

$p_i$  = sample data / data training

$q_i$  = data uji / data testing

$i$  = variabel data

$n$  = dimensi data

- c) Mengurutkan jarak yang terbentuk
- d) Menentukan jarak terdekat sampai urutan K
- e) Memasangkan kelas yang bersesuaian
- f) Mencari jumlah kelas dari tetangga terdekat Anda dan pilih satu sebagai kelas data untuk dievaluasi.

## 2.6 Evaluasi klasifikasi

Proses evaluasi klasifikasi didasarkan pada pengujian pada item yang benar dan item yang salah. Jenis model yang paling cocok untuk hasil klasifikasi ditemukan dengan menggunakan validasi ini ialah confusion matrix. Confusion matrix adalah informasi mengenai hasil klasifikasi aktual yang dapat diprediksi oleh sebuah sistem klasifikasi [10].

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<b>TP</b> (True Positive)	<b>FP</b> (False Positive) <i>Type I Error</i>
	0 (Negative)	<b>FN</b> (False Negative) <i>Type II Error</i>	<b>TN</b> (True Negative)

tabel 1. Confusion Matrix

- a) True positif (TP) merupakan jumlah record positif dalam dataset yang diklasifikasikan positif.
- b) True negatif (TN) merupakan jumlah record negative dalam dataset yang diklasifikasikan negatif.
- c) False positif (FP). merupakan jumlah record negatif dalam dataset yang diklasifikasikan positif.
- d) False negatif (FN) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif.

### 3. HASIL DAN PEMBAHASAN

Dari hasil pengumpulan data dari komentar video youtube data tersebut dikumpulkan dalam file csv.

#### 3.1. Crowling dataset

Yaitu sebuah proses yang menggali jauh ke dalam website atau target tertentu untuk mengambil data. Disini peneliti mengambil dataset dari salah satu konten pada youtube, dengan menggunakan website netlytic.org. Netlytic adalah salah satu perangkat untuk menganalisis jaringan pada media sosial, yaitu dalam percakapan di media sosial seperti (twitter, FB, youtube, dan lain-lain). Penelitian ini menggunakan 957 opini yang bersifat positif, dan negatif, dengan melabeli secara manual yaitu sekitar 30% data latih dan 70% data uji. Sebagai contoh komentar youtube yang telah dilabeli, Seperti pada Tabel 1

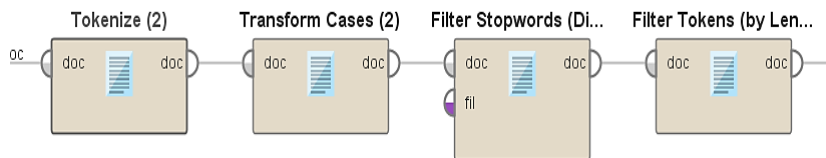
Tabel 1. Contoh Sentimen pada dataset

deskripsi	sentimen
Gpp deh semua pekerjaan manusia di kerjakan oleh AI jdi manusia fokus sm ibadah sm meperbaiki Ahlaknya?	Positif
Ntar ai buat dirinya sendiri dan menjadi robot dan dunia ini di penuh robot era cyberpunk pun di mulai waktunya survival melawan robot	Negatif

Label dengan sentimen positif berarti yang setuju atau tidak mempermasalahkan adanya AI, sedangkan sentimen negatif berarti yang tidak menyetujui adanya AI.

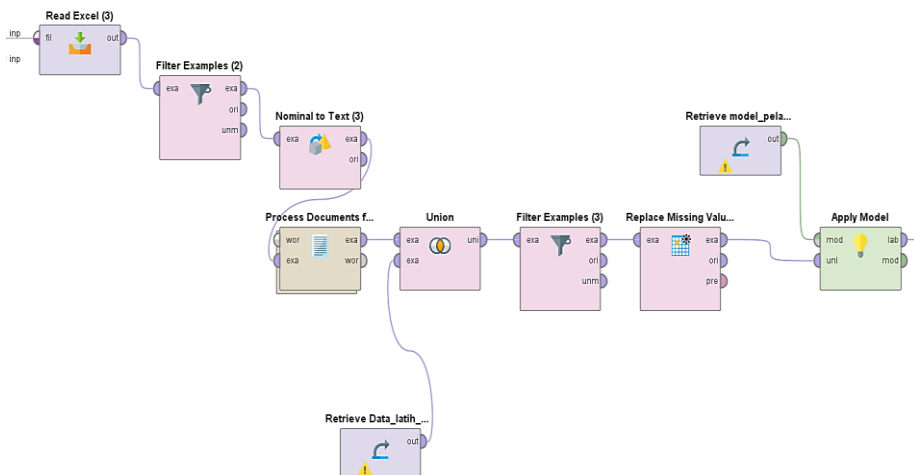
### 3.2. Pre-processing

Pada tahap ini peneliti harus lebih dulu melakukan subproses agar teks dalam keadaan clean atau bersih melakukannya dengan aplikasi Rapidminer. Subproses adalah langkah untuk menghilangkan URL, simbol-simbol dalam description, emoticon, dan lainnya. Setelah subproses selesai selanjutnya pre-pemrosesan data yang telah di clean kemudian tokenizing, stopwords, dan stemming. Pada aplikasi Rapidminer. Gambar 2 adalah pre-pemrosesan pada penelitian ini.



**Gambar 2.** Premrosesan data

Tranform casses berfungsi untuk mengubah seluruh teks yang ada dari huruf kapital menjadi huruf kecil. Lalu tokenize bertujuan untuk memisahkan kalimat menjadi kata untuk diproses lebih lanjut. filter stopwords berfungsi sebagai menghilangkan kata-kata yang tidak diperlukan. Selanjutnya adalah stemming adalah menghilangkan kata imbuhan pada setiap kata yang diproses pada tahap sebelumnya.



**Gambar 3.** Rangkaian Proses Rapidminer

Gambar 3 merupakan tahapan-tahapan untuk menghilangkan kata imbuhan menjadi kata dasarnya yaitu seperti 'pekerjaan' menjadi kerja. Dan pada tahap ini untuk pengujian dari data latih dan model yang telah dilakukan pada tahap sebelumnya.

### 3.3. TF-IDF

Tahapan ini merupakan pembobotan kata yang menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), yaitu bertujuan untuk mengubah data teks menjadi numerik. Bidang pembelajaran mesin atau Machine learning dan deep learning akan bekerja dengan baik dan maksimal jika data yang diolah berupa data numerik [11]. Yaitu sebuah fitur yang berfungsi sebagai pengukuran kata dalam dokumen dengan konteks tertentu dari dokumen yang lebih besar. Biasanya metode ini digunakan untuk memproses bahasa natural, untuk pengembangan teks dan mengambil informasi. Dengan urutan rumus sebagai berikut:

#### a) TF (Term Frequency)

TF adalah frekuensi nilai untuk kemunculan kata dalam dokumen. Dengan rumus perhitungan sebagai berikut:

$$TF_{Kata} = \left( \frac{\text{total jumlah dokumen}}{\text{Total jumlah kata dalam dokumen}} \right)$$

Nilai dari TF tidak dapat memberikan informasi penting dari kata tersebut dalam dokumen. Yaitu kata-kata konjungsi yang sering muncul pada dokumen atau kata umum yang bisa jadi memiliki nilai TF tinggi, sebaliknya kata tersebut tidak mempunyai makna yang penting dalam sebuah dokumen. Karena hal ini dibutuhkan teknik lain, yaitu IDF (Inverse Document Frequency) yaitu yang akan memberikan pembobotan pada kata yang sering muncul dalam sebuah dokumen. Dengan rumus perhitungan sebagai berikut:

#### b) IDF (Inverse Document Frequency)

$$IDF_{Kata} = \log \left( \frac{\text{total jumlah dokumen (N)}}{\text{jumlah dokumen yang mengandung kata (n)}} \right)$$

Yaitu total jumlah dokumen didalam kumpulan dari dokumen dan jumlahdokumen yang mengandung kata-kata tersebut. dari hasil nilai TF



dan IDF yang didapat selanjutnya adalah TF-IDF dapat dihitung dengan rumus dibawah ini:

c) TF-IDF (Term Frequency-Inverse Document Frequency)

$$TF - IDF_{Kata} = TF_{Kata} \times IDF_{Kata}$$

Yaitu kata yang memiliki nilai tinggi terhadap TF-IDF dianggap penting yang akan memberikan sebuah kontribusi yang besar untuk menentukan suatu topik dalam kumpulan dokumen tersebut.

Row No.	word	in documents	total
1	manusia	280	550
2	robot	91	137
3	kesadaran	66	100
4	mesin	57	78
5	dunia	61	71
6	data	35	68
7	tuhan	51	65
8	film	55	64
9	otak	40	62
10	teknologi	44	62

**Gambar 4.** Result All TF-IDF

Pada gambar 4 dapat dijelaskan yaitu gabungan dari kata dalam dokumen yang mana kata-kata tersebut sering muncul atau sebagai top word yang dijabarkan dalam bentuk tabel.

### 3.4. Word Cloud

Tahap selanjutnya yaitu word cloud, Word Cloud adalah visualisasi dari kumpulan beberapa kata dengan ukuran teks yang lebih besar, dan berupa awan kata, yang mana kata yang sering digunakan dalam teks.

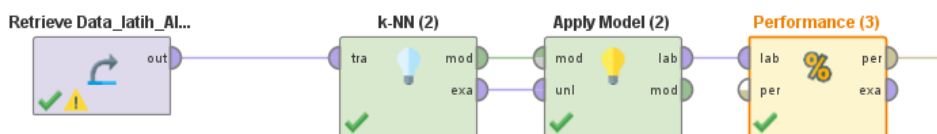


Gambar 5. Word Cloud

Dari gambar diatas dapat disimpulkan word cloud merupakan sebuah visualisasi yang kuat dengan menggambarkan sebuah informasi dalam teks secara ringkas dan juga mudah dipahami.

### 3.5. Kasifikasi K-Nearest Neighbors

Untuk penelitian ini menggunakan algoritma KNN (K-Nearest Neighbors) yang merupakan salah satu algoritma untuk Machine Learning. Guna melihat tingkatan akurasi dari algoritma KNN untuk proses analisis sentimen dengan pengaplikasian Rapidminer. Dibawa adalah proses untuk melihat akurasi menggunakan algoritma KNN.



Gambar 6. Proses melihat akurasi algoritma KNN.

Ini merupakan step yang diperlukan untuk pengujian tingkat akurasi atau performa dari model algoritma KNN. Data untuk pengujian tingkat akurasi ini didapat dari data latih.

### 3.6. Evaluasi klasifikasi

accuracy: 70.82%

	true positif	true negatif	class precision
pred. positif	63	46	57.80%
pred. negatif	22	102	82.26%
class recall	74.12%	68.92%	

**Gambar 7.** Hasil sentimen menggunakan algoritma KNN.

Dari hasil training pada gambar 7 dengan menggunakan algoritma K-Nearest Neighbors (KNN) mendapatkan hasil antara lain: akurasi yang didapat yaitu sebesar 70.82%, class precision untuk pred.positif dengan skor 57.80%, dan class.precision pred.negatif dengan skor 82.26%. dan untuk class.recall true positif dengan skor 74.12%, dan clasa.recall true negatif dengan skor 68.92%.

## 4. KESIMPULAN

Berdasarkan hasil penelitian dengan menggunakan metode K-Nearest Neighbor mendapatkan akurasi sebesar 70,82%, dengan sentimen positif sebesar 57.80% dan negatif 68.92%. Klasifikasi cocok untuk data dengan distribusi yang jelas dan dapat memprediksikan dengan cepat. Namun harus menyesuaikan kebutuhan yang baik. penelitian diawali dengan mengcrawling data, dan kemudian data diolah dalam proses pre-processing. Lalu, data bersih diproses menggunakan pemrosesan cepat. Selanjutnya, algoritma K-Nearest Neighbor, yang memiliki tingkat akurasi yang tinggi, digunakan untuk klasifikasi. Tahap terakhir, evaluasi metode yang digunakan dengan pemrosesan cepat, didasarkan pada hasil perhitungan kata yang diprediksi benar positif dan benar negatif.

## REFERENSI

- [1] A. Suryaningsih, "Dampak Media Sosial Terhadap Prestasi Belajar Peserta Didik," *Edusaintek J. Pendidikan, Sains Dan Teknol.*, vol. 7,

- no. 1, pp. 1–10, 2020, doi: 10.47668/edusaintek.v7i1.45.
- [2] E. I. Supriyadi and D. B. Asih, “Key to artificial intelligence (AI),” *Adv. Intell. Syst. Comput.*, vol. 1252 AISC, no. 2, pp. 647–656, 2021, doi: 10.1007/978-3-030-55190-2\_49.
- [3] R. Asrianto and M. Herwinanda, “Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 431–440, 2022, doi: 10.37859/coscitech.v3i3.4368.
- [4] E. S. Susanto, K. Kusriani, and H. Al Fatta, “Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor,” *Respati*, vol. 13, no. 2, pp. 67–72, 2018, doi: 10.35842/jtir.v13i2.260.
- [5] F. Rozi, M. Bagoes, and S. Junianto, “Penerapan Machine Learning Untuk Prediksi Harga Saham PT.Telekomunikasi Indonesia Tbk Menggunakan Algoritma K-Nearest Neighbors,” *J. Inform. MULTI*, vol. 1, no. 1, pp. 18–24, 2023.
- [6] F. Ahluna *et al.*, “Metode K-Nearest Neighbor Untuk Analisis Sentimen Tentang Penghapusan Ujian Nasional,” *J. Ikraith-Informatika*, vol. 7, no. 2, pp. 1–6, 2023.
- [7] F. Kurniawan and Q. Al Qorni, “Exploring Sentimen Analysis Using Machine Learning: A Case Study on Partai Demokrasi Indonesia Perjuangan (PDIP) in the 2024 General Election,” *Sinomics J. / Vol.*, vol. 2, no. 4, pp. 911–920, 2023.
- [8] A. Salam, J. Zeniarja, and R. S. U. Khasanah, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.
- [9] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [10] A. N. Kasanah, M. Muladi, and U. Pujiyanto, “Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 196–201, 2019, doi: 10.29207/resti.v3i2.945.
- [11] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes Pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH,” vol. 13, no. April, pp. 84–93, 2023.