

# ANALISIS PERBANDINGAN ANGKA PUTUS SEKOLAH DASAR NEGERI DAN SWASTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (K-NN)

Yayi Naulia Silalahi<sup>1\*</sup>, Rosita Panjaitan<sup>2</sup>, Lailan Sofinah Harahap<sup>3</sup>

<sup>1,2</sup>Sistem Informasi, Universitas Islam Negeri Sumatera Utara Medan; Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353; Telp & Fax: (+6261) 6615683

<sup>3</sup>Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara; Jl. Kapten Muchtar Basri No.3, Glugur Darat II, Kec. Medan Tim., Kota Medan, Sumatera Utara 20238; Telp & Fax: 061- 6625474

Received: 2 Januari 2025  
Accepted: 14 Januari 2025  
Published: 20 Januari 2025

## Keywords:

K-Nearest Neighbors; School Dropout Rate; Data Analysis.

## Correspondent Email:

yayinauliasilalahi@gmail.com

**Abstrak.** Artikel ini mengulas perbandingan tingkat putus sekolah antara sekolah dasar negeri dan swasta di Indonesia, dengan memanfaatkan algoritma K-Nearest Neighbors (KNN) untuk menganalisis data. Studi ini bertujuan untuk menemukan provinsi yang memiliki tingkat putus sekolah tertinggi serta menganalisis pola-pola yang membedakan status pendidikan berdasarkan informasi yang ada. Data menunjukkan bahwa tingkat putus sekolah lebih tinggi di sekolah negeri daripada di sekolah swasta. Algoritma KNN telah berhasil digunakan dalam klasifikasi dan menghasilkan evaluasi yang memuaskan, yang menunjukkan kemampuannya sebagai alat bantu dalam pengambilan keputusan di bidang pendidikan. Artikel ini ditutup dengan saran kebijakan untuk mengurangi tingkat putus sekolah, khususnya pada sekolah negeri.

**Abstract. Abstract.** This article reviews the comparison of dropout rates between public and private elementary schools in Indonesia, using the K-Nearest Neighbors (KNN) algorithm to analyze the data. This study aims to find provinces that have the highest school dropout rates and analyze patterns that differentiate educational status based on existing information. Data shows that the dropout rate is higher in public schools than in private schools. The KNN algorithm has been successfully used in classification and produced satisfactory evaluations, which shows its ability as an aid in decision making in the field of education. This article concludes with policy suggestions to reduce school dropout rates, especially in public schools.

## 1. PENDAHULUAN

Angka putus sekolah merupakan salah satu tantangan besar yang dihadapi sistem pendidikan di Indonesia, khususnya di tingkat dasar. Fenomena ini tidak hanya berdampak pada individu yang terpaksa menghentikan studinya, namun juga mempunyai konsekuensi jangka panjang bagi masyarakat dan negara. Ada banyak faktor berbeda yang berkontribusi terhadap angka putus sekolah tinggi, termasuk

kondisi ekonomi keluarga yang buruk, rendahnya tingkat pendidikan orang tua, dan kondisi geografis yang sulit [1]. Penelitian menunjukkan bahwa terdapat perbedaan karakteristik yang signifikan antara sekolah negeri dan swasta di Indonesia. Sekolah swasta cenderung memiliki fasilitas dan sumber daya yang lebih baik, sedangkan sekolah negeri menghadapi tantangan seperti rasio guru-siswa yang tinggi, yang dapat mempengaruhi

kualitas pendidikan yang diterima dan diterima siswa [2].

Faktor pembeda kedua jenis sekolah ini antara lain afiliasi, bahasa, kurikulum, fasilitas, dan biaya pendidikan [3]. Selain faktor-faktor di atas, pernikahan dini, kurangnya motivasi belajar dan perundungan di sekolah juga menjadi penyebab penting yang meningkatkan risiko putus sekolah [4]. Angka putus sekolah dipengaruhi oleh banyak faktor yang berbeda, antara lain jenis kelamin, jumlah anggota keluarga, umur, jenis sekolah, rasio guru-murid, pendapatan keluarga, dan tingkat pendidikan [5]. Dampak dari tingginya angka putus sekolah ini sangat luas, antara lain meningkatnya kemiskinan, kesenjangan sosial, dan rendahnya kualitas sumber daya manusia, yang pada akhirnya menghambat pembangunan nasional [4].

Di Indonesia, perbedaan antara sekolah negeri dan swasta memberikan dinamika unik. Sekolah swasta biasanya lebih eksklusif dengan sumber daya yang lebih baik, tetapi sering menghadapi tantangan seperti rasio guru-siswa yang tinggi dan keterbatasan fasilitas [6]. Oleh karena itu, untuk membuat kebijakan pendidikan yang lebih efisien, penting untuk memahami pola putus sekolah di kedua jenis sekolah ini. Dalam konteks ini, penting untuk melakukan analisis lebih lanjut mengenai perbandingan angka putus sekolah antara sekolah dasar negeri dan swasta. Studi ini bertujuan untuk mengidentifikasi provinsi dengan angka putus sekolah tertinggi dan menganalisis tren yang membedakan status pendidikan berdasarkan informasi yang tersedia.

Data mining adalah proses mengubah data implisit dan tidak berguna menjadi informasi dari kumpulan data besar. Penambangan data melibatkan penggunaan teknologi pengenalan pola serta teknik statistik dan matematika untuk menemukan korelasi, pola, dan tren baru yang bermakna dengan menambang sejumlah besar data yang disimpan dalam repository [7]. Algoritma KNN terbukti efektif dalam mengklasifikasikan data dan memberikan evaluasi yang memuaskan, sehingga dapat digunakan sebagai alat pendukung keputusan di bidang pendidikan. Perbedaan kualitas antara sekolah negeri dan swasta mencerminkan penerapan kebijakan

pendidikan yang tidak merata oleh pemerintah [8].

Algoritma K-Nearest Neighbor (K-NN) adalah pengklasifikasi berbasis centroid terdekat. Menghitung jarak atau tingkat kesamaan antara data uji dan seluruh data pelatihan adalah prinsip dasar dari algoritma ini. Selanjutnya, algoritma akan mengidentifikasi K tetangga terdekat dari data uji dalam kumpulan data pelatihan. Kemudian, algoritma akan menentukan kategori data uji berdasarkan kategori tetangga-tetangga tersebut. Karena proses klasifikasinya didasarkan pada mayoritas tetangga terdekat, K-NN memiliki keunggulan dalam mengatasi ketidakseimbangan data [9].

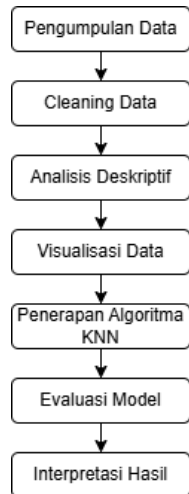
Algoritma K-Nearest Neighbors (KNN) merupakan metode analisis berbasis data yang dapat digunakan untuk menemukan pola-pola tersebut. Dengan menggunakan pendekatan berbasis jarak, KNN dapat mengklasifikasikan status pendidikan berdasarkan angka putus sekolah di berbagai provinsi.

Dalam penelitian ini, algoritma KNN diterapkan untuk menganalisis perbedaan angka putus sekolah antara sekolah negeri dan swasta di Indonesia. Selain itu, penelitian ini bertujuan untuk mengidentifikasi provinsi dengan angka putus sekolah tertinggi dan mengevaluasi faktor-faktor yang mungkin mempengaruhi hasil tersebut. Hasil analisis menunjukkan bahwa angka putus sekolah di sekolah negeri lebih tinggi dibandingkan di sekolah swasta. Dalam konteks ini, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan terhadap upaya peningkatan mutu pendidikan di Indonesia dan menurunkan angka putus sekolah yang masih menjadi permasalahan serius.

## 2. METODE PENELITIAN

Metode penelitian kuantitatif adalah pendekatan sistematis dan terstruktur yang menggunakan data numerik untuk menguji hipotesis. Langkah-langkah dalam penelitian kuantitatif antara lain memilih masalah, mengkaji literatur, menetapkan hipotesis, mengidentifikasi variabel, mengumpulkan dan menganalisis data, dan menarik kesimpulan [10]. Strategi penting dalam penelitian kuantitatif antara lain pemilihan instrumen pengukuran yang tepat dan teknik pengumpulan data yang efektif. Tantangan

paling umum dalam penelitian kuantitatif mencakup analisis data dan memastikan keandalan dan validitas hasil [11].



**Gambar 1.** Diagram Penelitian

### 2.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari portal data resmi Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek) Republik Indonesia. Data ini penting karena mencakup informasi komprehensif mengenai angka putus sekolah di tingkat sekolah dasar, yang dikelompokkan berdasarkan provinsi dan status sekolah, baik negeri maupun swasta. Oleh karena itu, data ini memberikan gambaran yang jelas mengenai kondisi pendidikan di berbagai wilayah di Indonesia.

Kumpulan data diunduh dalam format CSV, memfasilitasi analisis data menggunakan berbagai pemrograman dan perangkat lunak statistik. Kumpulan data ini berisi beberapa atribut penting yang relevan dengan penelitian, termasuk tingkat kelas, putus sekolah, dan status akademik. Atribut-atribut ini memungkinkan peneliti untuk melakukan analisis mendalam terhadap faktor-faktor yang mempengaruhi angka putus sekolah, serta membandingkan perbedaan antara sekolah negeri dan swasta [12].

### 2.2. Cleaning Data

Pemrosesan awal data adalah teknik yang mengekstraksi informasi dari data mentah. Teknik ini digunakan untuk menghilangkan noise yang ada pada data

mentah sehingga dapat diproses lebih lanjut. Data mentah diubah menjadi kumpulan data yang mempertimbangkan struktur, gaya penulisan, dan akurasi penulisan penelitian yang dilakukan oleh [13]. Cleaning data atau pembersihan data merupakan proses penting dalam pengolahan data yang bertujuan untuk meningkatkan kualitas data agar dapat digunakan secara efektif dalam analisis dan pemodelan. Proses ini mencakup identifikasi dan koreksi kesalahan, ketidaksesuaian, atau cacat pada kumpulan data, seperti: duplikat, data hilang, atau nilai tidak relevan.

### 2.3. Analisis Deskriptif

Analisis deskriptif dilakukan untuk memahami pola dan sebaran angka putus sekolah berdasarkan kategori sekolah negeri dan swasta. Langkah pertama adalah mengidentifikasi 10 Provinsi dengan angka putus sekolah tertinggi di setiap kategori, sekolah negeri dan swasta, dan mengidentifikasi wilayah mana yang paling membutuhkan perhatian. Perhitungan statistik deskriptif kemudian dilakukan, termasuk menghitung mean, median, dan standar deviasi angka putus sekolah. Tujuannya adalah untuk memahami distribusi data dan mengukur variasi angka putus sekolah di berbagai negara bagian. Terakhir, membandingkan tren angka putus sekolah di sekolah negeri dan swasta untuk mengidentifikasi kemungkinan perbedaan pola antara kedua kategori tersebut. Analisis ini memberikan informasi lebih rinci tentang status pendidikan masing-masing Provinsi dan perbedaan yang disebabkan oleh jenis sekolah.

### 2.4. Visualisasi Data

Proses visualisasi data merupakan proses gambaran hasil olah data agar lebih mudah dipahami. Ada beberapa diagram yang dapat digunakan, yaitu:

Buat diagram batang untuk menunjukkan distribusi angka putus sekolah berdasarkan Provinsi dan status sekolah.

Boxplot digunakan untuk menunjukkan sebaran data angka putus sekolah berdasarkan kategori. Bagan heatmap dibuat untuk menunjukkan korelasi antar atribut dalam kumpulan data. Pada penelitian ini, penulis menggunakan diagram batang.

### 2.5. Penerapan Algoritma K-NN

Dataset dibagi menjadi data latih (train) dan data uji (test) dengan perbandingan tertentu, misal 80:20.

Algoritma KNN dilatih dengan data latih untuk mengenali pola klasifikasi status.

Pengujian dilakukan dengan data uji untuk evaluasi kinerja model

### 2.6. Evaluasi Model

Pada tahap evaluasi model, yang dilakukan ialah menghitung metric evaluasi seperti:

#### Akurasi

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data}$$

Atau secara detail:

$$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{TP + TN + False\ Positives(FP) + False\ Negatives(FN)}$$

#### Precision

$$Precision = \frac{TP}{TP + FP}$$

#### Recall

$$Recall = \frac{TP}{TP + FN}$$

#### F1-Score

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

#### Support

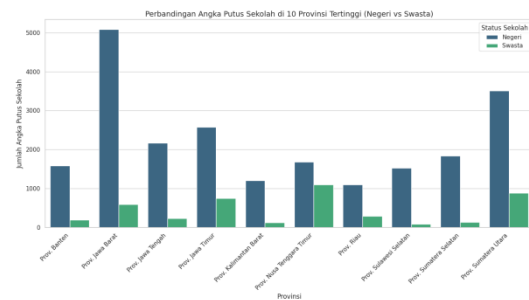
Support = Jumlah sampel aktual untuk kelas tertentu

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil

Proses cleaning data dilakukan untuk memastikan bahwa kumpulan data yang digunakan tidak mengandung nilai kosong atau baris yang tidak relevan. Kumpulan data akhir mencakup informasi tentang Provinsi, angka putus sekolah untuk setiap jenjang pendidikan dasar (Tingkat I hingga Tingkat VI), jumlah putus sekolah, dan status sekolah (negeri atau swasta). Kemudian penulis menganalisis data ini untuk mengidentifikasi 10 Provinsi dengan tingkat putus sekolah tertinggi. Hasil penelitian menunjukkan bahwa angka putus sekolah di sekolah negeri secara umum lebih tinggi dibandingkan di sekolah swasta. Hal ini berlaku di sebagian besar Provinsi, dengan beberapa pengecualian dimana perbandingannya tampak lebih seimbang.

Bila divisualisasikan dalam diagram batang (Biru: Negeri, Hijau: Swasta), perbedaan angka putus sekolah antara sekolah negeri dan swasta di setiap Provinsi langsung terlihat jelas. Data ini menyoroti bahwa sebagian besar Provinsi dengan tingkat putus sekolah tinggi menghadapi tantangan besar dalam pendidikan dasar, terutama di sekolah negeri.



**Gambar 2.** Diagram Batang Angka Putus Sekolah Dasar 10 Provinsi Tertinggi

Berdasarkan diagram batang di atas, Provinsi dengan angka putus sekolah dasar Negeri tertinggi ialah Provinsi Jawa Barat dengan jumlah mencapai 5000 mahasiswa dan angka putus sekolah dasar negeri terendah ialah Provinsi Riau dengan jumlah 1000 mahasiswa.

Sedangkan untuk angka putus sekolah dasar Swasta tertinggi ialah Provinsi Nusa Tenggara Timur dengan jumlah hampir mencapai 2000 mahasiswa dan angka putus sekolah dasar negeri terendah ialah Provinsi Sulawesi Selatan dengan jumlah kurang dari 1000 mahasiswa.

Model klasifikasi K-Nearest Neighbor (KNN) digunakan untuk memprediksi status sekolah berdasarkan tingkat putus sekolah di setiap jenjang pendidikan. Model dilatih pada 80% data pelatihan dan 20% data digunakan untuk pengujian. Hasil evaluasi menunjukkan bahwa akurasi model ANN adalah 87,5%.

Accuracy: 0.875

Classification Report:

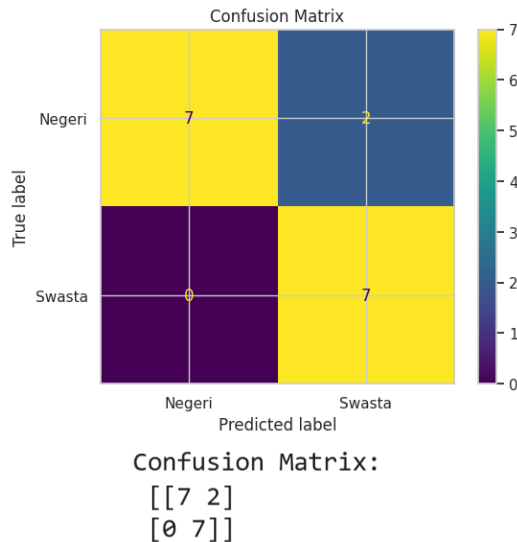
Negeri: {'precision': 1.0, 'recall': 0.7777777777777778, 'f1-score': 0.875, 'support': 9.0}  
Swasta: {'precision': 0.7777777777777778, 'recall': 1.0, 'f1-score': 0.875, 'support': 7.0}

accuracy: 0.875

**Gambar 3.** Hasil K-NN

Matriks konfusi menunjukkan bahwa model tersebut memprediksi sekolah negeri (7 benar, 2 salah) dan sekolah swasta (7 benar, 0

salah) dengan baik. Selanjutnya, skor presisi, recall, dan F1 masing-masing menunjukkan performa tinggi pada kedua kategori tersebut.



**Gambar 4.** Confusion Matrix

### 3.2. Pembahasan

Analisis menemukan bahwa tingginya angka putus sekolah di sekolah negeri kemungkinan dipengaruhi oleh beberapa faktor. Salah satu alasan utamanya adalah perbedaan kualitas fasilitas dan sumber daya antara sekolah negeri dan swasta. Sekolah negeri sering kali memiliki pilihan terbatas untuk dukungan akademis, sementara sekolah swasta cenderung memiliki sumber daya yang lebih baik karena dikelola secara independen. Selain itu, faktor sosial ekonomi juga memainkan peran penting. Siswa sekolah umum umumnya berasal dari latar belakang sosial ekonomi yang beragam, yang dapat memengaruhi kemampuan mereka untuk menyelesaikan pendidikan dasar.

Model JST yang digunakan untuk klasifikasi menunjukkan hasil yang memuaskan dengan akurasi sebesar 87,5%. Karya ini menunjukkan bahwa algoritma ANN dapat mendeteksi pola angka putus sekolah pada setiap jenjang pendidikan untuk menentukan status sekolah. Akurasi yang tinggi (100%) untuk kategori negeri menunjukkan bahwa model tersebut sangat akurat dalam memprediksi sekolah negeri. Namun, rendahnya tingkat ingatan (77,8%) untuk kategori ini menunjukkan bahwa beberapa data sekolah negeri tidak

teridentifikasi dengan benar. Sebaliknya, untuk kategori privat, recall mencapai 100%, namun akurasinya rendah (77,8%), menunjukkan adanya kesalahan klasifikasi.

Temuan-temuan ini memperkuat perlunya memberikan perhatian khusus pada sekolah-sekolah negeri di negara-negara dengan tingkat putus sekolah tertinggi. Peningkatan kualitas fasilitas, pelatihan guru, dan pemberian beasiswa dapat menjadi solusi untuk menurunkan angka putus sekolah. Selain itu, penelitian lebih lanjut juga disarankan, termasuk data tambahan seperti faktor sosial ekonomi, rasio guru-siswa, dan akses terhadap teknologi. Mempertimbangkan algoritme pembelajaran mesin lain seperti pohon keputusan dan hutan acak juga dapat membantu meningkatkan performa model

## 4. KESIMPULAN

Kesimpulan dari artikel ini ialah:

1. Angka putus Sekolah Dasar Negeri cenderung lebih tinggi dibandingkan di sekolah swasta, terutama di provinsi-provinsi dengan angka putus sekolah tertinggi. Faktor seperti keterbatasan fasilitas dan kondisi sosial ekonomi menjadi penyebab utamanya.
2. Model K-Nearest Neighbors (KNN) memiliki akurasi sebesar 87,5% yang menunjukkan kemampuan yang baik dalam membedakan status sekolah berdasarkan model angka putus sekolah.
3. Kajian ini menyoroti pentingnya memberikan perhatian pada sekolah negeri melalui peningkatan fasilitas, pemberian beasiswa dan program intervensi, terutama di daerah dengan angka putus sekolah yang tinggi. Penelitian lebih lanjut dapat memasukkan faktor sosial ekonomi untuk memberikan hasil yang lebih komprehensif.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada segenap tim yang membantu dalam penelitian ini, sehingga dapat terselesaikan. Terima kasih pula kepada Tim dari JITET Universitas

Lampung yang telah meluangkan waktunya untuk melakukan proses paper ini untuk bisa dipublikasikan.

#### DAFTAR PUSTAKA

- [1] Mujiati, Nasir, dan A. Ashari, “FAKTOR-FAKTOR PENYEBAB SISWA PUTUS SEKOLAH,” 2018.
- [2] U. Kulsum, “DISPARITAS KESIAPAN SEKOLAH NEGERI DAN SEKOLAH SWASTADALAM MENINGKATKAN MUTU PENDIDIKAN VOKASI,” vol. 4, 2020.
- [3] P. A. Cakranegara, “Faktor Diferensiasi Sekolah di Indonesia,” *Ideas J. Pendidikan, Sos. dan Budaya*, vol. 7, no. 2, hal. 13, 2021, doi: 10.32884/ideas.v7i2.335.
- [4] I. Widiyanti, A. Virgiani Abdul, A. Nirwana, A. Della Arlita, dan L. Nopita Sari, “Ancaman Melawan Putus Sekolah Dengan Dilema Kualitas Pendidikan Indonesia,” *J. Ilmu Sos. dan Pendidik.*, vol. 7, 2023, doi: 10.58258/jisip.v7i1.5228/http.
- [5] L. Kusbudiyanto dan A. I. Munandar, “Karakteristik Siswa Putus Sekolah Pada Jenjang Sekolah Menengah Kejuruan (Smk) Di Kota Bekasi,” *Sos. Horiz. J. Pendidik. Sos.*, vol. 6, no. 2, hal. 298–318, 2020, doi: 10.31571/sosial.v6i2.1153.
- [6] M. Khoiriyah, M. Humaira Fauziah, S. A. Zahra, dan N. W. Abdul Majid, “Pengaruh faktor ekonomi terhadap angka putus sekolah di kalangan mahasiswa indonesia,” *Integr. (Information Technol. Vocat. Educ.)*, vol. 3, 2021.
- [7] Lisda, N. Isnaeni, dan M. R. Firmansyah, “Analisa Slope Wilayah Kebakaran Hutan menggunakan Metode Naive Bayes,” *J. Sist. Inf. Galuh*, vol. 2, no. 2, 2024, doi: 10.25157/jsig.v2i2.3934.
- [8] Burhanuddin, E. Yusuf, I. Yurni, dan E. Maulani, “Analisis Pengembangan Model Pembelajaran Dalam Kelayakan Bangunan Sekolah Dengan Metode K-Nearest Neighbors,” *Sisfo J. Ilm. Sist. Inf.*, vol. 7, no. 2, hal. 97, 2023, doi: 10.29103/sisfo.v7i2.14788.
- [9] K. D. Attarik, N. Safriadi, dan Yulianti, “ANALISIS SENTIMEN KEBIJAKAN PEMBERLAKUAN TERHADAP PERTUMBUHAN EKONOMI SEKTOR E-COMMERCE DI INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBOR ( KNN ),” vol. 12, no. 3, 2024.
- [10] M. I. Syahroni, “Prosedur Penelitian Kuantitatif,” *J. Al-Musthafa STIT Al-Aziziyah Lomb. Barat*, vol. 2, no. 3, 2022.
- [11] E. Rosyidah dan E. Masykuroh, “Memahami Strategi dan Mengatasi Tantangan dalam Penelitian Metode Kuantitatif,” *Syntax Idea*, vol. 6, no. 6, hal. 2787–2803, 2024, doi: 10.46799/syntax-idea.v6i6.3748.
- [12] “Jumlah Siswa Putus Sekolah Menurut Tingkat Tiap Provinsi • Semua Wilayah • 2023 • SD/MI/Sederajat,” Portal Data Kemendikbudristek. Diakses: 7 Januari 2025. [Daring]. Tersedia pada: <https://data.kemdikbud.go.id/dataset/p/peserta-didik/jumlah-siswa-putus-sekolah-menurut-tingkat-tiap-provinsi-indonesia-sd-2023>
- [13] A. N. Kasanah, M. Muladi, dan 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, hal. 196–201, 2019, doi: 10.29207/resti.v3i2.945.