

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI TINGKAT BULLYING DI YAYASAN PENDIDIKAN DHARMA BAKTI SELESAI

Yeni Wahyuningsih¹, Imran Lubis², Kristina Annatasia Br Sitepu³

Sistem Informasi, STMIK Kaputama, Binjai

E-mail: *yenizakiaskana@gmail.com¹, imran.loebis.medan@gmail.com², kannatasia88@gmail.com³

ABSTRAK

Bullying merupakan salah satu permasalahan serius di lingkungan sekolah yang dapat memengaruhi kondisi psikologis dan prestasi siswa. Penelitian ini bertujuan untuk mengklasifikasikan tingkat bullying berdasarkan data kuesioner menggunakan algoritma K-Nearest Neighbor (KNN). Dataset penelitian terdiri dari 118 responden dengan 16 atribut yang mewakili faktor-faktor terkait bullying serta satu atribut label dengan tiga kategori tingkat bullying, yaitu Rendah, Sedang, dan Tinggi. Tahapan penelitian meliputi *preprocessing* data berupa penghapusan atribut yang tidak relevan, transformasi data kategorikal ke numerik dengan *Label Encoding*, serta normalisasi menggunakan *MinMaxScaler*. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk memastikan pemerataan distribusi kelas. Implementasi algoritma KNN dilakukan dalam dua pendekatan, yaitu KNN manual dengan perhitungan jarak Euclidean dan KNN menggunakan pustaka *scikit-learn*. Hasil pengujian menunjukkan bahwa akurasi yang diperoleh masih rendah, yaitu 37,5%, dengan nilai *precision*, *recall*, dan F1-score yang juga kurang dari 40%. Pengujian *cross-validation 5-fold* memperkuat hasil tersebut dengan rata-rata akurasi sebesar 34,89%, yang menunjukkan performa model belum optimal. Faktor yang memengaruhi hasil ini adalah jumlah data yang terbatas serta distribusi kelas yang tidak seimbang. Dengan demikian, algoritma KNN dapat digunakan untuk klasifikasi tingkat bullying, namun perlu dilakukan pengembangan lebih lanjut dengan penambahan data, penyeimbangan kelas, serta perbandingan dengan algoritma lain agar hasil yang diperoleh lebih akurat dan stabil.

Kata kunci

Bullying, K-Nearest Neighbor, Klasifikasi, Data Mining

ABSTRACT

Bullying is one of the serious problems in the school environment that can affect students' psychological conditions and academic performance. This study aims to classify the level of bullying based on questionnaire data using the K-Nearest Neighbor (KNN) algorithm. The research dataset consists of 118 respondents with 16 attributes representing bullying-related factors and one label attribute with three categories of bullying levels, namely Low, Moderate, and Severe. The research stages include data preprocessing by removing irrelevant attributes, transforming categorical data into numerical form using Label Encoding, and normalization using MinMaxScaler. The data were then divided into training and testing sets with an 80:20 ratio to ensure balanced class distribution. The KNN algorithm was implemented in two approaches, namely manual KNN with Euclidean distance calculation and KNN using the scikit-learn library. The testing results show that the obtained accuracy is still low, namely 37.5%, with precision, recall, and F1-score values also below 40%. The 5-fold cross-validation further confirmed these findings with an average accuracy of 34.89%, indicating that the model performance is not yet optimal. Factors influencing these results include the limited amount of data and the imbalance in class distribution. Therefore, the KNN algorithm can be applied for bullying classification, but further development is needed by increasing the dataset size, balancing the classes, and comparing with other algorithms to achieve more accurate and stable results.

Keywords

Bullying, K-Nearest Neighbor, Classification, Data Mining

1. PENDAHULUAN

Masa remaja merupakan fase transisi dari anak-anak menuju dewasa, di mana kondisi psikis sangat labil. Pada masa ini, peran lingkungan sangat diperlukan untuk membentuk kepribadian seorang remaja. Namun, kekerasan di dunia pendidikan, terutama kasus bullying, masih sering ditemui (Firdaus et al., 2025). Bullying didefinisikan sebagai perilaku agresif yang dilakukan berulang-ulang oleh individu atau kelompok yang memiliki kekuasaan, terhadap orang lain yang lebih lemah, dengan tujuan menyakiti orang tersebut (Olweus, 1993; Ramadhan et al., 2021). Salah satu hak anak yang penting untuk dipenuhi adalah hak anak untuk mendapatkan pendidikan dasar, namun masalah seperti bullying mampu merenggut masa depan anak (Magdalena et al., 2021).

Penerapan *data mining* menjadi kunci dalam mengidentifikasi pola tersembunyi. Dengan fokus pada identifikasi pola kombinasi *datasets*, penelitian ini diharapkan dapat memberikan wawasan baru terhadap faktor-faktor penyebab dan pola terjadinya bullying (Tarigan et al., 2022; Sugianto & Astita, 2017).

Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (KNN) untuk memprediksi tingkat bullying. KNN adalah salah satu algoritma *machine learning* yang paling sederhana dan intuitif, mengklasifikasikan titik data baru berdasarkan mayoritas label dari k tetangga terdekatnya (Sahar et al., 2020; Dwiki et al., 2021). KNN bersifat non-parametrik sehingga fleksibel untuk data yang distribusinya tidak teratur. Pada *dataset* yang ukurannya tidak terlalu besar, KNN dapat bekerja dengan efisien dan memberikan hasil yang akurat (Wahyudi & Kusumawardhana, 2021).

Penerapan algoritma K-NN diharapkan dapat memberikan prediksi yang lebih akurat terkait tingkat bullying di Yayasan Pendidikan Dharma Bhakti Selesai. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat memprediksi tingkat bullying di lingkungan Yayasan Pendidikan Dharma Bhakti Selesai dengan menggunakan algoritma K-NN, sebagai langkah untuk menciptakan lingkungan yang lebih aman dan nyaman bagi seluruh siswa.

2. METODE PENELITIAN

Penelitian ini menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang memiliki enam fase. Namun, penelitian ini berfokus pada fase pemodelan dan evaluasi. Lokasi penelitian adalah Yayasan Pendidikan Dharma Bakti Selesai, dengan waktu penelitian dari Maret hingga Mei 2025 (Alkhussayid & Ferdiansyah, 2022; Hasanah et al., 2021).

2.1. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh siswa SMA dan SMK di Yayasan Pendidikan Dharma Bakti Selesai. Teknik pengambilan sampel yang digunakan adalah *purposive sampling*, dengan sampel sebanyak 118 responden.

2.2. Variabel Penelitian

Variabel yang digunakan terbagi menjadi dua jenis:

- Variabel Independen (X): Terdapat 16 variabel independen (K1-K16) yang mencakup faktor-faktor terkait bullying (jenis kelamin, pengalaman, jenis bullying, frekuensi, dll.).
- Variabel Dependen (Y): Tingkat Bullying, yang memiliki tiga kategori: Rendah, Sedang, dan Tinggi. Kategorisasi dilakukan berdasarkan skor total dari kuesioner,

dengan rentang yang ditetapkan sebagai berikut: Rendah (3–16), Sedang (17–30), dan Tinggi (31–44).

2.3. Penerapan Metode

Tahapan penelitian meliputi:

- Pra-Pemrosesan Data: Data mentah dibersihkan dan ditransformasi. Data kategorikal diubah menjadi numerik menggunakan *Label Encoding* dan dinormalisasi menggunakan *MinMaxScaler* (Alfarizi et al., 2023).
- Membagi Data: Dataset dibagi menjadi data latih (80%) dan data uji (20%) secara *stratified* untuk mempertahankan distribusi kelas (Jtik et al., 2023).
- Penerapan Algoritma KNN: Algoritma KNN diimplementasikan dengan metrik jarak Euclidean. Jarak antara setiap data uji dan data latih dihitung menggunakan rumus:

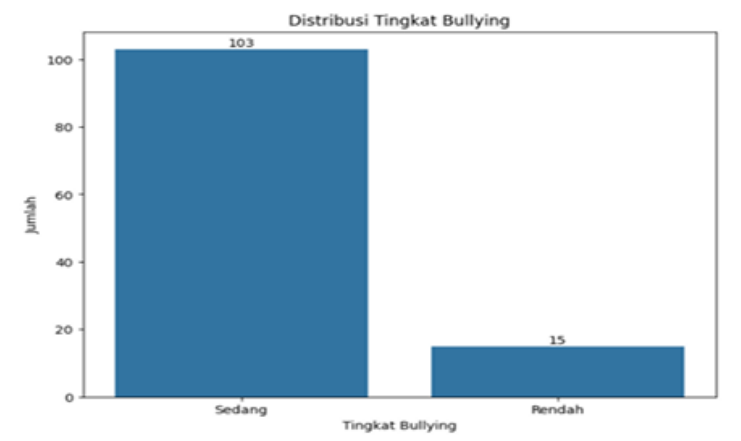
$$d_{(i,j)} = \sqrt{|x_{i1} - x_{j1}|^2 + \dots + |x_{ip} - x_{jp}|^2}$$

- Evaluasi Kinerja Model: Performa model diukur menggunakan *Confusion Matrix*, Akurasi, Presisi, *Recall*, dan F1-Score (Ridwansyah, 2022). Akurasi dihitung dengan rumus:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

3. HASIL DAN PEMBAHASAN

Berdasarkan analisis data kuesioner, mayoritas responden (103 orang) berada pada kategori tingkat bullying Sedang, sedangkan sisanya (15 orang) berada di kategori Rendah (Gambar 1). Ini menunjukkan bahwa kasus bullying dengan intensitas sedang lebih dominan dan memerlukan perhatian khusus. Ketidakseimbangan data ini berpotensi memengaruhi kinerja model klasifikasi.



Gambar 3. Hasil Tingkat Bullying Siswa

Hasil pengujian model KNN menunjukkan akurasi yang rendah, yaitu 37,5% baik pada perhitungan manual maupun menggunakan pustaka *scikit-learn* (Gambar 2). Nilai *precision*, *recall*, dan F1-score juga berada di bawah 40%. Rata-rata akurasi dari *cross-validation 5-fold* hanya 34,89%, yang mengindikasikan model tidak stabil dan sulit melakukan generalisasi pada data yang berbeda (Gambar 3).

```

=== Sklearn KNN ===
Confusion Matrix:
[[8 4]
 [6 6]]
Accuracy : 0.5833
Precision: 0.5857
Recall   : 0.5833
F1-Score : 0.5804

      precision    recall  f1-score   support

     0       0.57       0.67       0.62         12
     1       0.60       0.50       0.55         12

 accuracy         0.58         0.58         0.58         24
 macro avg       0.59       0.58       0.58         24
 weighted avg    0.59       0.58       0.58         24

```

Gambar 2: Hasil KNN *scikit-learn*

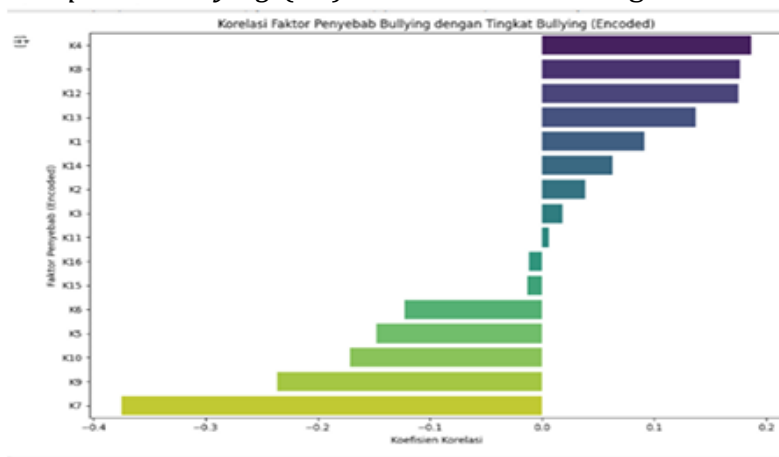
```

=== Cross Validation (5-Fold) ===
Scores: [0.33333333 0.33333333 0.20833333 0.34782609 0.52173913]
Mean Accuracy: 0.34891304347826085

```

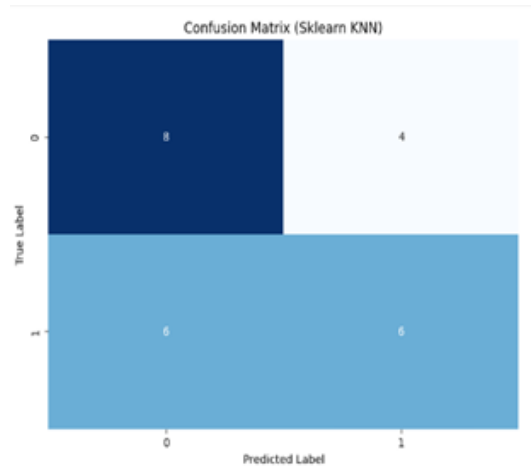
Gambar 3: Hasil *Cross Validation* (5-fold)

Analisis korelasi antara faktor-faktor penyebab *bullying* (K1-K16) dengan tingkat *bullying* menunjukkan hubungan yang bervariasi. Faktor dengan korelasi positif (misalnya K4, K8, K12, K13) cenderung meningkatkan tingkat *bullying*, sedangkan faktor dengan korelasi negatif (misalnya K7, K9, K10, K5, K6) menunjukkan hubungan berlawanan (Gambar 4). Faktor *pengalaman bullying* (K4) memiliki korelasi positif terkuat, sementara *pelaku bullying* (K7) memiliki korelasi negatif terkuat.



Gambar 4: Kekuatan dan arah hubungan antar variable

Kinerja model klasifikasi KNN secara keseluruhan dapat dikatakan kurang optimal. *Confusion Matrix* (Gambar 5) menunjukkan bahwa model lebih baik dalam mengenali kelas Rendah (*True Positives* = 8) daripada kelas Sedang (*True Negatives* = 6), namun terdapat kesalahan prediksi yang cukup tinggi. Hal ini disebabkan oleh keterbatasan jumlah data dan ketidakseimbangan distribusi kelas yang memengaruhi kemampuan model untuk belajar dan memprediksi dengan akurat (Daqiqil, 2021).



Gambar 5: Confusion Matrix

4. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma K-Nearest Neighbor (KNN) baik secara manual maupun menggunakan pustaka *scikit-learn* berhasil diterapkan untuk mengklasifikasikan tingkat *bullying* berdasarkan data kuesioner siswa dengan atribut K1-K16. Meskipun demikian, hasil pengujian menunjukkan nilai akurasi yang rendah, yaitu sekitar 37,5%. Hal ini disebabkan oleh keterbatasan jumlah data, distribusi kelas yang tidak seimbang, serta representasi atribut kuesioner yang masih sederhana. Algoritma KNN dapat digunakan untuk klasifikasi *bullying*, namun performanya pada dataset ini belum optimal. Oleh karena itu, diperlukan perbaikan lebih lanjut untuk meningkatkan akurasi model.

5. DAFTAR PUSTAKA

- Alfarizi, M. Riziq Sirfatullah, Muhamad Zidan Al-farish, Muhamad Taufiqurrahman, Ginan Ardiansah, and Muhamad Elgar. 2023. "Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning." Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID) 2(1): 1–6.
- Arsi, Primandani, and Retno Waluyo. 2021. "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)." Jurnal Teknologi Informasi dan Ilmu Komputer 8(1): 147. doi:10.25126/jtiik.0813944.
- Aulia, Nadira, Sarinah Sarinah, and Juanda Juanda. 2023. "Analisis Kurikulum Merdeka Dan Kurikulum 2013." Jurnal Literasi dan Pembelajaran Indonesia 3(1): 14–20.
- Dwiki, Aluisius, Adhi Putra, Safitri Juanita, Program Studi, Sistem Informasi, Fakultas Teknologi, Informasi Universitas, and Budi Luhur. 2021. "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN." 8(2): 636–46.
- Fauzi, Achmad. 2022. "Implementasi Kurikulum Merdeka Di Sekolah Penggerak." Pahlawan: Jurnal Pendidikan-Sosial-Budaya 18(2): 18–22. doi:10.57216/pah.v18i2.480.
- Gifari, Okta Ihza, Muh Adha, Ivan Rifky Hendrawan, Fernandito Freddy, and Setlight Durrand. 2022. "Analisis Sentimen Review Film Menggunakan TF-IDF Dan Support Vector Machine." 2(1): 36–40.

- Handayani, Arfina, and Ilka Zufria. 2024. "Analisis Sentimen Terhadap Bakal Capres RI 2024 Di Twitter Menggunakan Algoritma SVM." 5(1): 53–63. doi:10.47065/josh.v5i1.4379.
- Hasanah, Msy Aulia, Sopian Soim, and Ade Silvia Handayani. 2021. "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree Dengan Algoritma CART Untuk Prediksi Curah Hujan Berpotensi Banjir." *Journal of Applied Informatics and Computing* 5(2): 103–8. doi:10.30871/jaic.v5i2.3200.
- Hendrastuty, Nirwana, Auliya Rahman Isnain, Ari Yanti Rahmadhani, Program Studi, Sistem Informasi, Universitas Teknokrat Indonesia, Program Studi Informatika, Universitas Teknokrat Indonesia, and Kota Bandar Lampung. 2021. "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine." 6(3): 150–55.
- Husada, Hendry Cipta, and Adi Suryaputra Paramita. 2021. "Analisis Sentimen Pada Maskapai Penerbangan Di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)." *Teknika* 10(1): 18–26. doi:10.34148/teknika.v10i1.311.
- Idris, Irma Surya Kumala, Yasin Aril Mustofa, and Irvan Abraham Salihi. 2023. "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)." *Jambura Journal of Electrical and Electronics Engineering* 5(1): 32–35. doi:10.37905/jjee.v5i1.16830.
- Informatika, Manajemen, Universitas Amikom, and Kata Kunci. "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS Informatika Universitas Amikom Yogyakarta Abstraksi Keywords : Pendahuluan Tinjauan Pustaka."
- Jtik, Jurnal, Jurnal Teknologi, Rahmat Satria Buana, Windu Gata, Alda Zevana, Putri Widodo, and Hendra Setiawan. 2023. "Analisis Sentimen Pada Komen Twitter Pawang Hujan Mandalika Dengan Support Vector Machine (SVM) Dan Naïve Bayes." 7(2): 0–6.
- Lestari, Sri, and Safira Berliani. 2024. "Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine." 5(3): 951–60.
- Mursyid, Rabhita, and Aries Dwi Indriyanti. 2024. "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)." *Journal of Informatics and Computer Science (JINACS)* 06: 371–83. <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/61322%0Ahttps://ejournal.unesa.ac.id>.
- Octarina, Sisca, Fitri Maya Puspita, Evi Yuliza, and Indrawati Indrawati. 2025. "Pendampingan Penggunaan Google Colab Pada Pembelajaran Python Dan Machine Learning Bagi Dosen Matematika Di Palembang." *Jurnal Pepadu* 6(1): 56–66. doi:10.29303/pepadu.v6i1.6457.
- Permata Aulia, Thalita Meisya, Nur Arifin, and Rini Mayasari. 2021. "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19." *SINTECH (Science and Information Technology) Journal* 4(2): 139–45. doi:10.31598/sintechjournal.v4i2.762.
- Pratama, R A. 2024. "Analisis Sentimen Konsumen Dengan Teknik Text Mining." *Jurnal Dunia Data* 1(6): 1–17. <http://www.cyberarea.id/index.php/duniadata/article/view/109>.
- Ridwansyah, Tengku. 2022. "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier." *KLIK: Kajian Ilmiah Informatika dan Komputer* 2(5): 178–85. doi:10.30865/klik.v2i5.362.

- Riset, Jurnal, William Silalahi, and Adi Hartanto. 2024. "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024 Support Vector Machine Sentiment Classification Based on Optimization Welcoming 2024 Election." 7(2): 245–55. doi:10.30595/jrst.v7i2.18133.
- Riza, Faizal. 2023. "Analisis Sentimen Pengguna Twitter Terhadap Pelaksanaan Implementasi Kurikulum Merdeka." Ismetek 15(2): 38–44.
- Savitri, Ni Luh Putu Chandra, Radya Amirur Rahman, Reyhan Venyutzky, and Nur Aini Rakhmawati. 2021. "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring Pada Twitter Menggunakan Supervised Machine Learning." Jurnal Teknik Informatika dan Sistem Informasi 7(1): 47–58. doi:10.28932/jutisi.v7i1.3216.
- Septiana, Angga, Gifthera Dwilestari, Agus Bahtiar, Teknik Informatika, Sistem Informasi, Data Mining, Kurikulum Merdeka, and Kementerian Pendidikan. 2024. "DALAM ANALISIS SENTIMEN PENGGUNA TWITTER X TERHADAP." 8(1): 323–30.
- Septiani, Dwi, and Ica Isabela. 2022. "Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks." SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia 1(1): 81–88.
- Syahputra, Rizki Agam, and Maulia Rahmi Hanifah. 2024. "Metode Analisis Kesehatan Dengan Mengguakan Mechine Learning Atau Artificial Inteligenci Atau Data Mining Literature Review." Jurnal Industri & Inovasi (INVASI) 1(2): 27–38. <http://jurnal.utu.ac.id/invasi/>.
- Wahyudi, Rizki, and Gilang Kusumawardhana. 2021. "Analisis Sentimen Pada Review Aplikasi Grab Di Google Play Store Menggunakan Support Vector Machine." 8(2): 200–207.
- Yunita, Rani, and Mia Kamayani. 2023. "Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi." Indonesian Journal of Computer Science 12(5): 2879–90. doi:10.33022/ijcs.v12i5.3415.