



Decision Support System for Classifying Bullying Victims Using Website-Based C4.5 Algorithm

Fransiska Nabila Ariesha¹, Melka Andriana^{2*}, Yunita Marien³, Ardiyansyah⁴, Yeni Mustika⁵

1234 Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Kampus Kota Pontianak

19220652@bsi.ac.id¹, 19220702@bsi.ac.id^{2*}, 19220694@bsi.ac.id³, Ardiyansyah.arg@bsi.ac.id⁴, Yeni.yem@bsi.ac.id⁵

Abstract

Bullying remains a serious problem in Indonesian schools, where Indonesia ranks highest in ASEAN with a bullying rate of 84%. Early detection of bullying victims is critical for timely intervention by school counselors (BK teachers). This study develops a web-based Decision Support System (DSS) that integrates the C4.5 decision tree algorithm to classify students at risk of becoming bullying victims. The system uses the Bullying 2018 with Labels dataset from Kaggle containing 22,766 student records with eight input attributes and four output classes: Ringan (Mild), Sedang (Moderate), Berat (Severe), and Sangat Berat (Very Severe). The C4.5 model is implemented in Python using scikit-learn with entropy criterion and validated through 10-fold cross validation, achieving an accuracy of 99.94%. The model is deployed as a Flask REST API and integrated with a PHP CodeIgniter 3 web frontend and MySQL database. The system allows BK teachers to input student behavioral data and receives instant risk classification results. Usability testing confirmed the system is effective, efficient, and satisfying to use. This DSS provides an objective, data-driven tool to support early identification and handling of bullying cases in schools.

Keywords: *Bullying Classification; C4.5 Algorithm; Decision Support System; Machine Learning; Websit*

1. Introduction

Bullying is aggressive behavior that includes physical violence, verbal abuse, and psychological pressure carried out by individuals or groups who feel more powerful against victims perceived as weaker [1]. Indonesia ranks highest in bullying cases among ASEAN countries at 84%, surpassing Nepal and Vietnam (79%), Cambodia (73%), and Pakistan (43%), based on KPAI 2017 data [2]. According to UNICEF (2020), approximately 41% of students aged 15 have experienced bullying at least several times per month [3]. In early 2024, KPAI recorded 141 reports of violence against children, with around 35% occurring in school environments [4].

The impacts of bullying on victims are wide-ranging, including psychological trauma, depression, anxiety, and reduced academic achievement [5]. In school practice, counseling services are not always handled by trained professionals, and many teachers consider bullying part of normal student development, resulting in lack of serious follow-up [6]. This situation makes early identification of bullying victims particularly challenging.

To address this problem, this research develops a web-based Decision Support System (DSS) integrating the C4.5 classification algorithm. Unlike previous studies that focused on sentiment analysis of social media text using Naive Bayes or SVM with accuracy rates of 66–84% [7][8][9], this study directly applies classification on structured student behavioral data within an operational school system. The C4.5 algorithm was selected because it uses Information Gain and Gain Ratio to determine optimal attributes, and is proven more effective in handling class imbalance compared to KNN (37.5% accuracy in comparable studies) [10]. The resulting system provides BK teachers with an objective, data-driven tool for early screening of students at risk of bullying.

2. Literature Review

2.1. Decision Support System

A Decision Support System (DSS) is an interactive information system designed to present data, perform modeling, and manage information to support decision-making in structured, semi-structured, and unstructured contexts [11]. In the educational context, DSS can help schools determine appropriate actions for bullying cases objectively and based on data [12].

2.2. C4.5 Algorithm

The C4.5 algorithm is an enhancement of the ID3 decision tree algorithm. Unlike ID3, which uses Information Gain, C4.5 applies Gain Ratio as the splitting criterion to select the best attribute for partitioning data into subgroups that are more homogeneous with respect to a target class [13]. This method is widely used because it can process both numerical and categorical data and handles class imbalance more effectively than methods such as Naive Bayes or K-Nearest Neighbor [14].

The entropy and gain ratio formulas used are:

$$\text{Entropy}(S) = \sum -p_i \times \log_2(p_i) \quad (1)$$

$$\text{Gain Ratio}(S,A) = \text{Information Gain}(S,A) / \text{Split Information}(S,A) \quad (2)$$

2.3. Bullying Classification

Classification is the process of creating a model that describes and distinguishes categories so that the model can be used to predict the class of an object whose label is not yet known [15]. In the context of bullying detection, classification algorithms are applied to categorize students based on behavioral indicators into risk levels, enabling early intervention by school staff.

3. Research Methodology

3.1. Text font of entire document

This study uses the *Bullying 2018 with Labels* secondary dataset obtained from Kaggle (<https://www.kaggle.com/datasets/felicitaarcadia/bullying-2018-with-labels>). The dataset consists of 22,766 student records with 8 input attributes and 1 target label (Cluster/Level of Bullying). No missing values or duplicate records were found in the dataset.

Table 1. Input Attributes of the Dataset

No	Attribute	Data Type	Values	Type
1	Bullied on school property (past 12 months)	Categorical	Yes / No	Input
2	Age (Custom_Age)	Numeric	11 – 18	Input
3	Sex	Categorical	Male / Female	Input
4	Physically attacked	Numeric	0 – 12 times	Input
5	Physical fighting	Numeric	0 – 12 times	Input
6	Felt lonely	Categorical	Never / Rarely / Sometimes / Often	Input
7	Close friends	Numeric	0 – 3 or more	Input
8	Miss school no permission	Numeric	0 – 10+ days	Input
9	Cluster (Bullying Level)	Numeric	0=Mild, 1=Moderate, 2=Severe, 3=Very Severe	Label

Source: Kaggle – Bullying 2018 with Labels

3.2. Data Preprocessing

The preprocessing stage involved two main steps. First, Attribute Selection, where 8 attributes most relevant to bullying indicators were selected from the dataset. Second, Data Transformation, where categorical attributes such as Sex and Bullied_on_school were encoded to binary values (0/1) using Label Encoding, while attributes such as age, physical attack frequency, loneliness, and school absence were mapped to ordinal numerical values. Since the dataset contained no missing values or duplicates, data cleaning was not required.

Table 2. Class Distribution in Dataset

Class	Label	Count	%
0	Ringan (Mild)	1,460	6.41%
1	Sedang (Moderate)	17,960	78.89%
2	Berat (Severe)	2,426	10.66%
3	Sangat Berat (Very Severe)	920	4.04%
	Total	22,766	100%

Source: Python test results on Bullying 2018 with Labels dataset (2026)

3.3. Model Building and Validation

The C4.5 model was built using the DecisionTreeClassifier class from the scikit-learn library with criterion='entropy' parameter in Python on Google Colaboratory. The class_weight='balanced' parameter was applied to handle class imbalance, while max_depth=10 and min_samples_leaf=5 were set to prevent overfitting. Model validation used 10-fold cross validation with stratified KFold to ensure proportional class distribution across all folds. The trained model was saved as a .pkl file using the joblib/pickle library for deployment.

3.4. System Design

The system is built with two integrated components: (1) a PHP CodeIgniter 3 web application that handles the user interface, form input, and data management; and (2) a Python Flask REST API that loads the trained C4.5 model and processes prediction requests. When a BK teacher submits student data through the web form, the CodeIgniter controller sends an HTTP POST request containing the eight input features in JSON format to the Flask API. The Flask API returns a JSON response containing the predicted risk class (0–3) and its label (Ringan/Sedang/Berat/Sangat Berat). The result is then saved to the MySQL database in table *tb_proses* and displayed to the teacher on the results page.

4. Results and Discussion

4.1. Model Accuracy

The C4.5 classification model achieved an accuracy of 99.94% using 10-fold cross validation on the dataset of 22,766 records. This high accuracy demonstrates the effectiveness of the entropy-based criterion in creating a well-fitted decision tree for the bullying classification task. The parameter settings, particularly `class_weight='balanced'`, contributed to the model's ability to handle the significant class imbalance in the dataset, where class 1 (Moderate) comprised 78.89% and class 3 (Very Severe) only 4.04% of the data.

4.2. System Interface

The system provides two separate interfaces, namely a dedicated interface for BK teachers and a dedicated interface for administrators. On the BK teacher interface, users begin with a Registration and Login page before accessing a Dashboard that presents a statistical overview of prediction history organized across four risk categories. When conducting a new assessment, teachers fill out a structured prediction form covering three sections of student information, namely personal data, school experience data, and social condition data. Once submitted, the system displays a Prediction Result page that presents the classified risk level along with a corresponding recommended course of action. Meanwhile, the administrator interface is designed for broader oversight, offering data analysis monitoring, user management, and a downloadable report feature.

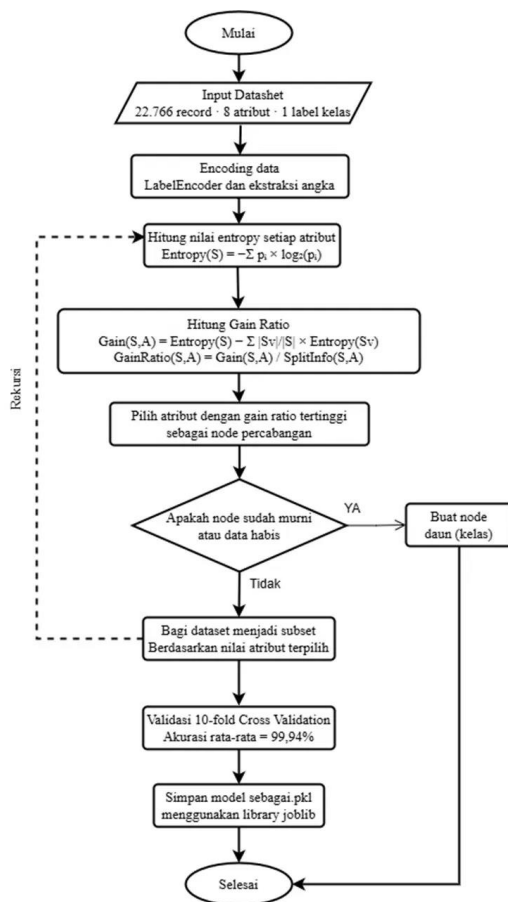


Fig. 1: C4.5 Algorithm Flowchart.

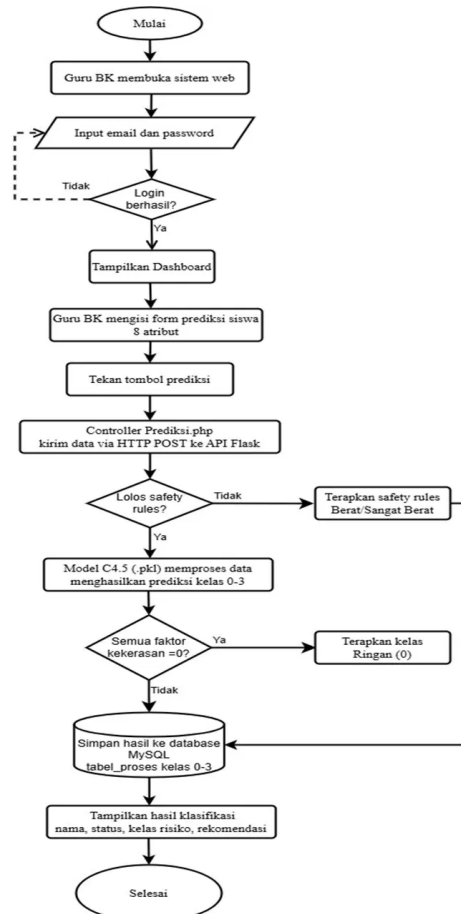


Fig. 2: System Integration Flowchart

Prediksi Bullying Siswa

Isi semua data di bawah untuk mendapatkan hasil prediksi

DATA PRIBADI

Nama Siswa

Usia **Jenis Kelamin**

PENGALAMAN DI SEKOLAH

Pernah diintimidasi di Sekolah dalam 12 bulan terakhir **Pernah Diserang Secara Fisik berapa kali**

Pernah Berkelahi Secara Fisik berapa kali **Tidak Masuk Sekolah Tanpa Izin berapa hari**

KONDISI SOSIAL

Merasa Kesepian **Jumlah Teman Dekat**

Prediksi

Reset

← Kembali

Fig. 3: New Prediction Form Interface

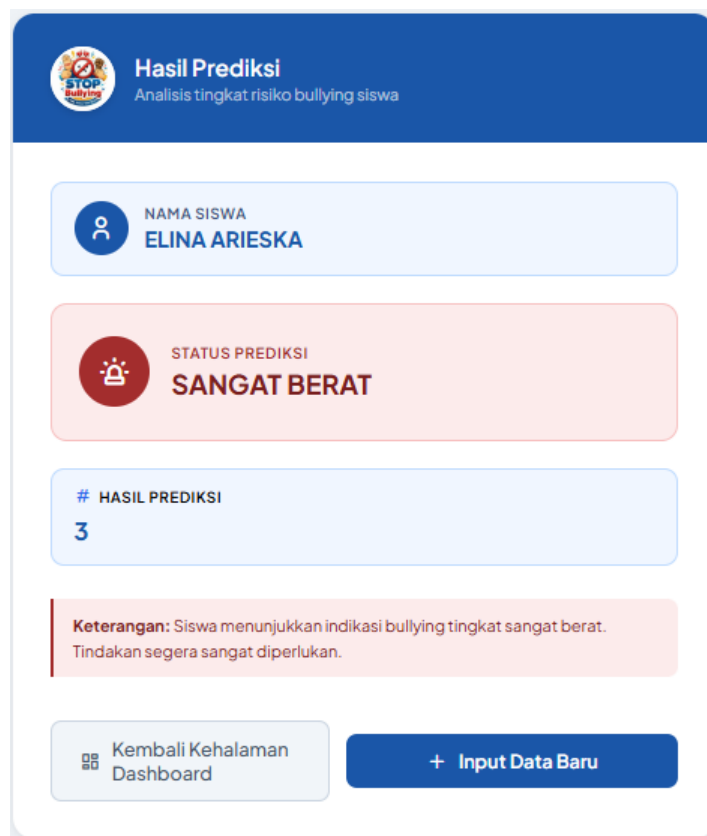


Fig. 4: Prediction Result Page

4.3. Comparison with Related Studies

Compared to previous studies, this system demonstrates significant improvements. Studies using Naive Bayes for cyberbullying sentiment analysis on Twitter achieved accuracies of 69% [7] and 66–67% [8], while SVM-based approaches reached 72–84.5% [7][9]. A KNN-based approach on school data achieved only 37.5% accuracy [10]. This study's C4.5 model achieved 99.94% accuracy by working directly with structured student attribute data rather than unstructured social media text, and by employing class balancing techniques appropriate for the imbalanced dataset.

5. Conclusion

This study successfully developed a web-based Decision Support System for classifying bullying victims using the C4.5 algorithm. The C4.5 model trained on the Bullying 2018 with Labels dataset (22,766 records, 8 attributes, 4 classes) achieved an accuracy of 99.94% through 10-fold cross validation. The system integrates a CodeIgniter 3 PHP frontend with a Flask Python REST API and MySQL database, allowing BK teachers to input student behavioral data and receive instant risk classification into four categories: Mild, Moderate, Severe, and Very Severe. Usability testing confirmed the system is effective and user-friendly for school counselors. This DSS provides an objective, structured tool to support early identification and timely handling of bullying cases, complementing the work of professional counselors and school psychologists. Future work could explore real-time data integration, mobile interface development, and the incorporation of additional behavioral attributes to further improve classification performance and system usability.

Acknowledgement

The authors thank the supervisors Bapak Ardiyansyah, M.Kom and Ibu Yeni Mustika, M.Pd for their guidance throughout this research, as well as Universitas Bina Sarana Informatika, Kampus Kota Pontianak for institutional support.

References

- [1] F. Ali, C. Ariesty, L. Lauren, R. Wulandari, and N. Maharani, "BULLYING DI SEKOLAH DASAR," vol. 1, no. 4, pp. 496–504, 2022.
- [2] R. A. Sfaat, "Tindakan Bullying di Lingkungan Sekolah yang Dilakukan Para Remaja," vol. 1, no. 2, pp. 97–100, 2023.
- [3] M. I. Djamzuri and A. P. Mulyana, "Fenomena Bullying Dalam Mendorong Kebijakan Literasi Berbasis AI (Artificial Intelligence) Pada Teknologi Media Baru," vol. 3, pp. 1304–1312, 2023.
- [4] Y. E. Nizmi, D. A. Hanif, R. Anjelli, and M. R. Pratama, "Penerapan Fenomena Bullying pada Tingkat Sekolah Dasar dan Menengah di Desa Banjar Seminai Tahun 2024," pp. 606–619, 2024.
- [5] Abdillah, "EDUCARE : Jurnal Pendidikan dan Kesehatan Dampak Bullying di Sekolah Dasar dan Pencegahannya," no. 2021, pp. 102–108, 2024.

- [6] W. Widiyawati, E. Mulyani, W. L. Fitrihanur, D. D. Musfara, and D. L. Octavia, "PENCEGAHAN DAN PENANGANAN BULLYING," vol. 7, pp. 92–99, 2025.
- [7] R. Hilma, M. Ula, S. Fachrurrizi, M. T. Informasi, and U. Malikussaleh, "Analisis Sentimen Cyberbullying pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Naïve Bayes Classifier," pp. 107–123, 2023.
- [8] S. A. Mahmudah, A. Yudhistira, F. Teknik, and U. T. Indonesia, "Analisis Sentimen Terhadap Cyberbullying pada Platform Media Sosial X Menggunakan Algoritma Naive Bayes Sentiment Analysis of Cyberbullying on Social Media Platform X Using Naive Bayes Algorithm," vol. 5, no. 1, pp. 189–200, 2025.
- [9] A. M. Alhakim, H. Leonardo, and A. N. Putri, "Analisis Sentimen Terhadap Bullying Di Indonesia Pada Twitter Menggunakan Naïve Bayes dan SVM," vol. 2, no. 3, pp. 210–218, 2025.
- [10] Y. Wahyuningsih, I. Lubis, and K. A. B. Sitepu, "Global Research and Innovation Journal (GREAT) IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI Masa remaja merupakan fase transisi dari anak-anak menuju dewasa , di mana kondisi psikis sangat labil . Pada masa ini , peran lingkungan sangat d," vol. 1, pp. 887–893, 2025.
- [11] A. Mikaya, Z. Rusdi, M. D. Lauro, S. Informasi, and U. Tarumanagara, "SISTEM PENDUKUNG KEPUTUSAN DALAM PEMILIHAN," pp. 2–7, 2023.
- [12] A. Al Ghifari, P. Irfansyah, and Y. Nugraha, "Dinamika MENGGUNAKAN METODE AHP (ANALYTICAL Dinamika," pp. 417–428, 2021.
- [13] D. Nisrina, T. Informatika, and U. N. Waluyo, "Analisis Kepuasan Konsumen Menggunakan Metode Algoritma C4 . 5 Berbasis Rapidminer pada PT . Adeaksa," vol. 9, no. 1, pp. 26–33, 2024.
- [14] A. Saburina, Husaini, and Fitriyani, "KULIAH MBKM DI UNIVERSITAS JABAL GHAFUR," no. 1, pp. 10–17, 2024.
- [15] D. Sari, Hardian Oktavianto, and S. Ilham, "Jurnal Smart Teknologi Jurnal Smart Teknologi," vol. 3, no. 4, pp. 413–422, 2022.

jurnal-sistem-pendukung-keputusan-untuk-mengklasifikasi-pada-korban-bullying_1781339016623.docx

ORIGINALITY REPORT

15%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.sciencepubco.com Internet	99 words — 5%
2	www.ioinformatic.org Internet	34 words — 2%
3	ejournal.unwaha.ac.id Internet	20 words — 1%
4	ioinformatic.org Internet	19 words — 1%
5	www.coursehero.com Internet	17 words — 1%
6	jurnal.darmajaya.ac.id Internet	14 words — 1%
7	acopen.umsida.ac.id Internet	12 words — 1%
8	www.bright-journal.org Internet	12 words — 1%
9	Xu Cheng, Ji Xu, Jian Pei, Jiangchuan Liu. "Hierarchical distributed data classification in wireless sensor networks", Computer Communications, 2010 Crossref	11 words — 1%