

# Analisis Pola Kasus Verbal Bullying menggunakan Algoritma K-Nearest Neighbor (Studi Kasus: Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa)

Abdullah Zaki<sup>1</sup>, Rudy Harijadi Wibowo Pardanus<sup>2</sup>, Johan Reimon Batmetan<sup>3</sup>

<sup>1,2,3</sup> Jurusan Pendidikan Teknologi Informasi dan Komunikasi, Fakultas Teknik, Universitas Negeri Manado

Correspondent Author :

[20208049@unima.ac.id](mailto:20208049@unima.ac.id)

**Abstract** — This study aims to analyze patterns of verbal bullying cases in Senior and Vocational High Schools in Minahasa Regency using the K-Nearest Neighbor (K-NN) algorithm. Primary data were obtained through questionnaires distributed to 521 students, which, after processing, resulted in a dataset of 15,630 entries. The data processing stages were conducted using the SEMMA (Sample, Explore, Modify, Model, Assess) methodology. The K-NN algorithm was applied to classify bullying behavior patterns, with performance evaluation conducted using a confusion matrix. The results showed that peers accounted for the majority of bullying perpetrators (90.70%), compared to teachers (9.30%). The K-NN algorithm demonstrated effectiveness with acceptable accuracy, making it a valuable tool for supporting educational policy decisions to prevent bullying.

**Keyword** — Verbal Bullying, Data Mining, Algoritma K-NN, SEMMA, Confusion Matrix.

**Abstrak** — Penelitian ini bertujuan untuk menganalisis pola kasus verbal bullying di lingkungan Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa menggunakan algoritma K-Nearest Neighbor (K-NN). Data primer diperoleh melalui kuisioner dari 521 siswa, yang setelah diproses menghasilkan dataset sebanyak 15.630 data. Tahapan pengolahan data dilakukan dengan metode SEMMA (Sample, Explore, Modify, Model, Assess). Algoritma K-NN diterapkan untuk klasifikasi pola perilaku bullying dengan evaluasi performa menggunakan confusion matrix. Hasil menunjukkan bahwa teman sebaya merupakan pelaku bullying terbesar (90,70%), dibandingkan guru (9,30%). Algoritma K-NN menunjukkan efektivitas dengan akurasi yang dapat diterima, sehingga dapat digunakan untuk mendukung pengambilan keputusan kebijakan pendidikan terkait pencegahan bullying.

**Kata kunci** — Verbal Bullying, Data Mining, K-NN Algorithm, SEMMA, Confusion Matrix.

## I. PENDAHULUAN

Bullying telah menjadi perhatian global karena dampaknya yang merusak terhadap kesehatan mental dan fisik, terutama pada kalangan remaja. Di Indonesia, fenomena bullying masih sering terjadi, dengan berbagai jenis intimidasi yang dilaporkan, termasuk bullying verbal, yang mencakup penghinaan, gosip, dan ejekan. United Nations International Children's Emergency Fund (UNICEF) melaporkan bahwa Indonesia memiliki tingkat kekerasan anak yang tinggi dibandingkan negara-negara Asia lainnya, seperti Vietnam dan Kamboja.

Menurut Programme for International Students Assessment (PISA) dalam (Oktaviany & Ramadan, 2023), sebanyak 22% anak-anak di Indonesia pernah mengalami penghinaan verbal. Selain itu, dampak psikologis dari bullying mencakup depresi, kecemasan, penurunan kepercayaan diri, bahkan hingga keinginan untuk bunuh diri. Bullying verbal tidak hanya mengancam kesehatan mental tetapi juga memengaruhi prestasi akademik dan kehidupan sosial siswa. Kabupaten Minahasa, dengan 51 Sekolah Menengah Atas dan Kejuruan (Statistik & Minahasa, 2023), juga menghadapi tantangan besar dalam menangani kasus bullying. Sebagai contoh, sebuah kasus di Minahasa pada tahun 2024 menunjukkan bahwa seorang siswa mengalami trauma mendalam akibat bullying verbal yang dilakukan oleh teman sebayanya (seputarsulut.com, 2024). Peneliti juga mengamati beberapa sekolah secara langsung yaitu SMA N 1 Tondano, SMA N 2 Tondano dan SMK N 3 Tondano. Hal ini menyoroti perlunya pendekatan berbasis data untuk memahami pola dan penyebab bullying, sehingga dapat dirumuskan strategi pencegahan yang lebih efektif.

Dalam konteks ini, data mining, khususnya klasifikasi menggunakan algoritma K-Nearest Neighbor (K-NN), menawarkan solusi praktis. Algoritma ini mampu mengolah data dalam jumlah besar untuk mengidentifikasi pola yang tersembunyi (Hakim et al., 2019). Dengan memanfaatkan pendekatan ini, penelitian bertujuan untuk memberikan pandangan baru dalam menangani bullying verbal di kalangan siswa.

Berdasarkan uraian permasalahan diatas, peneliti tertarik untuk melakukan Analisis dengan Algoritma K-Nearest Neighbor (K-NN) dalam menggali dan mengolah data siswa Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa menggunakan software orange data mining. Maka peneliti memilih judul “Analisis Pola Kasus Verbal Bullying Menggunakan Algoritma K-Nearest Neighbor (K-NN) (Studi Kasus: Sekolah Menengah Atas dan Kejuruan Di Kabupaten Minahasa)”.

## II. KAJIAN TEORI

### A. Psikologi

Psikologi menjadi bagian dari upaya memahami rangkaian perkembangan yang dialami manusia. Karena hanya dengan memahami latar belakang tahapan perkembangan dengan baik, informasi ini dapat menjadi

pedoman menentukan langkah terbaik yang sesuai dengan individu masing-masing. Dengan kata lain, setiap orang memiliki kemampuan dan potensi yang berbeda-beda dalam hidupnya. Menemukenali aspek psikologi yang tepat dalam diri seseorang, akan dapat membantu orang tersebut berkembang secara maksimal di dalam lingkungannya. Faktor-faktor psikologis dipandang sebagai cara-cara berfungsinya pikiran siswa dalam pengaruhnya dengan pemahaman bahan pelajaran, sehingga penguasaan terhadap bahan yang disajikan lebih mudah dan efektif. Dengan demikian, proses belajar mengajar itu akan berhasil dengan baik kalau didukung oleh faktor-faktor psikologis dari siswa (Rudy HW Pardanus, Randy A. Pondaag, 2021). Berikut adalah pembahasan lebih komprehensif mengenai teori psikologi yang relevan yang dijadikan acuan bagi peneliti untuk melakukan analisis pola kasus verbal bullying di kalangan siswa SMA/SMK :

1. Teori Harga Diri (Self-Esteem Theory)  
Teori harga diri menyatakan bahwa pandangan individu terhadap diri sendiri memengaruhi respons terhadap lingkungan. Pada remaja, harga diri rentan terhadap interaksi sosial, terutama bullying verbal, yang sering kali merendahkan aspek personal seperti penampilan atau kemampuan akademik. Korban bullying cenderung mengalami penurunan harga diri, kecemasan, dan depresi. Penguatan harga diri melalui intervensi psikologi dapat membantu korban pulih, dengan pihak sekolah berperan menciptakan lingkungan yang positif untuk mendukung mekanisme koping yang lebih baik (Dika, 2022; Prastiti & Anshori, 2023).
2. Teori Lingkungan Sekolah (School Climate Theory)  
Teori ini menyatakan bahwa iklim sekolah yang sehat dan suportif dapat mencegah perilaku bullying dengan memperkuat hubungan siswa-guru dan menciptakan lingkungan inklusif di mana setiap siswa merasa aman dan dihargai (Aziz, 2019; Cut Nurmasyithah, 2022). Strategi pencegahan berbasis iklim sekolah meliputi pelatihan guru untuk menciptakan lingkungan yang adil serta program pengembangan karakter siswa. Implementasi di Indonesia menunjukkan penurunan bullying verbal dan peningkatan keterlibatan siswa dalam aktivitas positif (Asiyah, 2020).
3. Teori Perilaku Sosial dan Pembelajaran Observasional (Social Learning Theory)  
Menurut Albert Bandura, teori pembelajaran sosial menyatakan bahwa individu belajar melalui observasi dan modeling dari lingkungan sosial. Pada siswa SMA/SMK, bullying sering dipengaruhi oleh pengaruh teman sebaya atau tokoh populer. Ketika perilaku bullying mendapat perhatian atau tidak dihukum, siswa cenderung menirunya untuk meningkatkan status sosial. Pendekatan kontrol sosial, seperti penegakan aturan tegas, mentoring, serta pelatihan bagi guru dan orang tua untuk menjadi

panutan positif, dapat membantu mencegah perilaku ini (Prastiti & Anshori, 2023).

#### 4. Teori Dukungan Sosial (Social Support Theory)

Teori dukungan sosial menyatakan bahwa dukungan dari teman, keluarga, dan guru dapat mengurangi dampak negatif bullying dengan meningkatkan ketahanan mental korban dan mendorong mereka mencari bantuan. Studi di Indonesia menunjukkan bahwa dukungan sosial yang kuat membantu siswa mengembangkan mekanisme koping yang lebih baik dan mengurangi risiko masalah kesehatan mental (Cut Nurmasyithah, 2022). Program konseling dan kegiatan positif di sekolah telah berhasil menciptakan lingkungan yang mendukung, sehingga menurunkan risiko bullying verbal dan meningkatkan solidaritas siswa.

#### B. Pengertian Bullying

Kata "bullying" berasal dari Bahasa Inggris, yakni dari kata "bull" yang merujuk pada banteng yang suka mengganggu orang di sekitarnya. Secara etimologis, dalam Bahasa Indonesia, kata "bully" berarti pengertak atau orang yang mengganggu individu yang lebih lemah. Dalam pengertian terminologi, menurut Ken Rigby seperti yang dijelaskan dalam Astuti (2008; 3, dalam Ariesto, 2009 dan dalam (Zakiyah et al., 2017), "bullying" dapat diartikan sebagai suatu keinginan untuk menyakiti, yang kemudian diekspresikan melalui tindakan, menyebabkan penderitaan pada seseorang.

#### C. Data Mining

Data mining menurut Juliansa (2019) dalam (Etriyanti, E., Syamsuar, D., & Kunang, 2020) adalah proses untuk mendapatkan ilmu pengetahuan dari sebuah informasi yang berasal dari gudang basis data. Data mining merupakan proses penggalian informasi dan berguna dari set data besar yang melibatkan konsep interdisipliner yang relatif baru yang melibatkan analisis data dan penemuan pengetahuan dari database dan menggunakan pendekatan multi-sisi yang mencakup analisis statistik, visualisasi data, penemuan pengetahuan, pengenalan pola dan manajemen basis data (Nikmatun, I., & Waspada, 2019).

#### D. Text Mining

Text mining adalah teknik analisis yang mengolah data teks tidak terstruktur menjadi lebih terstruktur untuk menemukan pola dan informasi tersembunyi. Prosesnya meliputi tokenisasi, penghapusan stopwords, dan teknik seperti stemming serta pembobotan TF-IDF. Teknik utama seperti analisis sentimen, klasifikasi, dan pengelompokan membantu mengidentifikasi pola yang sering muncul, seperti dalam kasus bullying verbal di sekolah atau media sosial. Metode ini mempermudah peneliti dalam mengungkap tren perilaku bullying melalui kata-kata atau frasa yang sering digunakan (Teknokrat, 2024; Hutagalung et al., 2021).

#### E. Metode Klasifikasi Data Mining

Data mining adalah analisis otomatis pada data besar untuk menemukan pola. Metode dibagi menjadi Supervised (dengan target variabel) dan Unsupervised (tanpa target variabel). Penelitian ini menggunakan Supervised Methods. Tiga metode utama: Prediction, Association, dan Segmentation. Klasifikasi, termasuk algoritma seperti Decision Trees dan Naïve Bayes, membagi data menjadi training (80%) dan testing (20%) untuk mengukur akurasi dengan confusion matrix. Klasifikasi mengelompokkan data berdasarkan aturan dan label yang mengarahkan pembelajaran untuk mencapai akurasi (Antoni Wibowo, 2017; Angelia Fransiska, 2021; Wikipedia, n.d.).

#### F. Algoritma K-Nearest Neighbor (K-NN)

Menurut Tang, Jing, Li, & Atkinson (2016) dalam (Homepage et al., 2021), klasifikasi berdasarkan sistem informasi memetakan data ke dalam kelas yang sudah ditentukan. Salah satu teknik klasifikasi yang populer adalah K-Nearest Neighbor (KNN), yang digunakan dalam berbagai bidang seperti pengenalan pola dan objek. KNN dianggap efisien dalam menangani data besar. KNN bekerja dengan mengklasifikasikan data baru berdasarkan kemiripan dengan titik data yang sudah ada. Langkah-langkahnya meliputi:

1. Menentukan nilai K (jumlah tetangga terdekat).
2. Menghitung jarak Euclidean setiap objek ke data sampel.
3. Mengurutkan objek berdasarkan jarak terkecil.
4. Mengumpulkan kategori tetangga terdekat.
5. Menentukan klasifikasi berdasarkan mayoritas kategori.

Jarak Euclidean digunakan untuk mengukur kedekatan antar titik dalam ruang dua dimensi. Berikut rumus untuk menghitung jarak Euclidean:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

#### G. SEMMA

Metode SEMMA (Sample, Explore, Modify, Model, Assess) adalah pendekatan dalam data science yang dikembangkan oleh SAS Institute untuk memudahkan prediksi variabel dalam proyek data mining (Kevin Wilson, 2021).

#### H. Cross Validation

Metode validasi yang digunakan adalah Cross Validation, yaitu teknik untuk menilai performa model dengan menguji data pada beberapa subset. Setiap subset secara bergantian berfungsi sebagai data uji, sementara subset lainnya digunakan untuk melatih model. Proses ini diulang beberapa kali.

#### I. Confusion Matrix

Confusion Matrix adalah pengukuran kinerja dalam klasifikasi machine learning dengan dua atau lebih kelas, yang terdiri dari empat kombinasi nilai prediksi dan aktual: True Positif, True Negatif, False Positif, dan False Negatif (Maria Susan Anggreany, S.Kom., 2022).

### III. METODE PENELITIAN

#### Pendekatan SEMMA

1. Sample: Pengumpulan data melalui kuisioner dari 521 siswa. Data ditumpuk untuk menghasilkan 15.630 entri.
2. Explore: Analisis deskriptif dilakukan untuk menemukan pola awal dalam dataset.
3. Modify: Preprocessing data mencakup case folding, tokenisasi, filtering, dan pembobotan TF-IDF.
4. Model: Penerapan algoritma K-NN untuk klasifikasi pola bullying verbal.
5. Assess: Evaluasi model menggunakan confusion matrix untuk mengukur akurasi dan performa klasifikasi.

### IV. HASIL DAN PEMBAHASAN

#### A. Hasil Penelitian

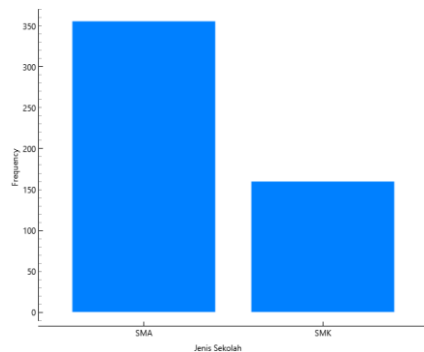
##### 1. Sample (Mengumpulkan Data)

Penelitian ini menggunakan data primer yang diperoleh melalui kuesioner dari 521 siswa SMA/SMK di Kabupaten Minahasa. Data awal mencakup atribut Nama Sekolah, Kelas, Jenis Pelaku, Kalimat Verbal Bullying, dan Kategori. Untuk memperkaya analisis, data ditumpuk berdasarkan kategori, sehingga jumlahnya bertambah menjadi 15.630 data dengan atribut Jenis Pelaku, Kalimat Verbal Bullying, dan Kategori. Berdasarkan kajian teori, penelitian ini berfokus pada tiga variabel utama: Jenis Pelaku, Kalimat Verbal Bullying, dan Kategori.

Tabel 1. Atribut Data

| No. | Atribut                 | Type        | Description                                                     |
|-----|-------------------------|-------------|-----------------------------------------------------------------|
| 1   | Jenis Pelaku            | Target      | Teman / Guru                                                    |
| 2   | Kalimat Verbal Bullying | Text        | Kalimat yang di Lontarkan                                       |
| 3   | Kategori                | Categorical | Diejek, Diredahkan, Difitnah, Digosipkan, Dihina, Dikritik, ... |

## 2. Explore (Deskripsi Data)



Gambar 1 Perbedaan Frekuensi Kasus Verbal Bullying antara Siswa SMA dan SMK

Grafik menunjukkan perbedaan frekuensi verbal bullying antara siswa SMA dan SMK. Kasus di SMA lebih tinggi, sekitar 350 kejadian, sementara di SMK lebih rendah, kurang dari 200 kasus. Berdasarkan 15.630 data yang telah ditumpuk, data dengan kategori jenis pelaku sebanyak 5.109, dapat dilihat pada Tabel 2. kemudian data dengan jenis pelaku tidak ada / tidak pernah sebanyak 10.521 data.

Tabel 2. Jenis Pelaku Verbal Bullying

| No           | Jenis Pelaku | Jumlah       | Persentase % |
|--------------|--------------|--------------|--------------|
| 1            | Guru         | 475          | 9,30 %       |
| 2            | Teman        | 4.634        | 90,70%       |
| <b>Total</b> |              | <b>5.109</b> | <b>100 %</b> |

Data menunjukkan total 5.109 kasus verbal bullying, dengan 90,7% (4.634 kasus) dilakukan oleh teman sebaya dan 9,3% (475 kasus) oleh guru. Hal ini menyoroti perlunya fokus pencegahan pada interaksi antar siswa.

## 3. Modify

Pada tahap Modify, dataset diolah menggunakan Microsoft Excel dan Orange. Di Excel, dilakukan penghapusan duplikasi, deteksi kesalahan data, dan pembersihan menggunakan fitur seperti remove duplicates, conditional formatting, dan find and replace. Selanjutnya, di Orange, dilakukan text preprocessing meliputi transformation, tokenization, filtering, dan stopwords removal untuk membuat dataset lebih terstruktur dan siap dianalisis. Contoh hasil pre-processing data yang dilakukan pada Ms.Excel dapat dilihat pada tabel 3.

Tabel 3. Hasil Preprocessing di Ms. Excel

| No. | Kalimat Verbal Bullying | Kategori    | Jenis Pelaku |
|-----|-------------------------|-------------|--------------|
| 1   | Dianggap bodoh          | Direndahkan | Guru         |
| 2   | Norak banget            | Direndahkan | Teman        |

|      |             |                          |       |
|------|-------------|--------------------------|-------|
| .    | .           | .                        | .     |
| 1310 | Jin dan jun | Diplesetkan              | Teman |
| 1311 | Noni        | Diplesetkan              | Guru  |
| .    | .           | .                        | .     |
| 5108 | Kocok       | Melecehkan secara verbal | Teman |

Selanjutnya, di Orange, dataset menjalani text preprocessing meliputi case folding, tokenization, stopwords filtering, pembagian data (Train-Test Split), dan pembobotan kata (TF-IDF).

### a. Transformation

Transformasi teks mengubah teks mentah menjadi format terstruktur dengan menyeragamkan huruf menjadi kecil (case folding), misalnya "BODOH!" menjadi "bodoh".

Tabel 4. Hasil Case Folding

| Sebelum                                                                                                                | Sesudah                                                                                                                |
|------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Mukamu mirip kertas amplas, badan kayak babi, item hangus mirip batu bara, bau badan lu kayak pembuangan limbah sampah | mukamu mirip kertas amplas, badan kayak babi, item hangus mirip batu bara, bau badan lu kayak pembuangan limbah sampah |

### b. Tokenization

Tokenization adalah proses memecah teks menjadi unit-unit kecil (token), seperti kata atau frasa.

Tabel 5. Hasil Tokenization

| Sebelum                                                                                                                | Sesudah                                                                                                                                                                       |
|------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mukamu mirip kertas amplas, badan kayak babi, item hangus mirip batu bara, bau badan lu kayak pembuangan limbah sampah | ["mukamu", "mirip", "kertas", "amplas", "badan", "kayak", "babi", "item", "hangus", "mirip", "batu", "bara", "bau", "badan", "lu", "kayak", "pembuangan", "limbah", "sampah"] |

### c. Filtering

Pada tahapan filtering, dilakukan penghapusan stopwords, yaitu kata-kata umum seperti "dan", "di", "yang" yang tidak bermakna dalam klasifikasi.

Tabel 6. hasil Filterig (Stopword)

| Sebelum                                                                                                                                                                       | Sesudah                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| ["mukamu", "mirip", "kertas", "amplas", "badan", "kayak", "babi", "item", "hangus", "mirip", "batu", "bara", "bau", "badan", "lu", "kayak", "pembuangan", "limbah", "sampah"] | ["mukamu", "kertas", "amplas", "badan", "babi", "item", "hangus", "batu", "bara", "badan", "pembuangan", "limbah", "sampah"] |

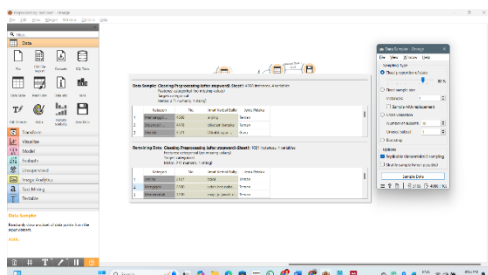
|                                      |
|--------------------------------------|
| “pembuangan”, “limbah”,<br>“sampah”] |
|--------------------------------------|

Setelah preprocessing text di Orange, dilakukan wordcloud untuk mengidentifikasi kata kunci penting dalam kalimat verbal bullying, sehingga peneliti fokus pada kata relevan.



Gambar 2. Hasil Wordcloud

- d. Pembagian Data (Train-Test Split)  
Pembagian data memisahkan 80% untuk training dan 20% untuk testing. Data training digunakan untuk melatih model, sementara data testing untuk mengevaluasi performa model. Pembagian yang tepat penting untuk akurasi dan presisi (Oktafiani et al., 2023). Pembagian ini dilakukan menggunakan widget "Data Sampler" di Orange Data Mining.



Gambar 3. Split Data Menggunakan Widget Data Sampler

Dengan menggunakan Orange Data Mining dan perhitungan manual, dataset 5.109 terbagi menjadi 4.088 data untuk training (80%) dan 1.021 data untuk testing (20%).

Tabel 7. Data Training (80% dari dataset)

| No.  | Kalimat Verbal Bullying              | Kategori                             | Jenis Pelaku |
|------|--------------------------------------|--------------------------------------|--------------|
| 1    | anjing                               | Memanggil dengan sebutan hewan/benda | Teman        |
| ...  |                                      |                                      |              |
| 1080 | kalo gak tau gak usah sok ngejelasin | Meremehkan                           | Guru         |
| 4088 | orang sumatra                        | Dihina                               | Teman        |

Tabel 8. Data Testing (20% dari dataset)

| No.  | Kalimat Verbal Bullying | Kategori                 | Jenis Pelaku |
|------|-------------------------|--------------------------|--------------|
| 1    | tolol                   | Direndahkan              | Teman        |
| ...  |                         |                          |              |
| 1307 | ditampar                | Direndahkan              | Guru         |
| 1021 | matthew jadi mati       | Melecehkan secara verbal | Teman        |

- e. Pembobotan TF-IDF

Setelah stopwords dihapus, data siap untuk Pembobotan Kata TF-IDF, yang mengubah teks menjadi vektor numerik untuk pemodelan Algoritma KNN. TF-IDF menilai signifikansi kata dalam teks (Reimon & Hariguna, 2024). Setiap kata diberi bobot setelah preprocessing.

4. Model (K-Nearest Neighbor)

Setelah pembobotan kata, data yang berjumlah 5.109 kasus diproses menggunakan Algoritma K-Nearest Neighbor (KNN). K-Nearest Neighbor (KNN) mengklasifikasikan data berdasarkan kemiripan dengan tetangga terdekat. Langkah-langkahnya:

- Tentukan nilai K, jumlah tetangga terdekat (optimal dari pelatihan).
- Hitung jarak Euclidean setiap data ke sampel.
- Urutkan data berdasarkan jarak terkecil.
- Identifikasi kategori mayoritas dari K tetangga terdekat.
- Tentukan klasifikasi berdasarkan kategori mayoritas.

Analisis klasifikasi kasus verbal bullying dilakukan menggunakan aplikasi Orange Data Mining untuk mempercepat proses.

1. Dataset Uji (Testing)

Berikut data uji yang akan di klasifikasikan dengan Algoritma K-Nearest Neighbor.

Tabel 9. Data Uji (Testing)

| Kalimat Verbal Bullying                                                               |
|---------------------------------------------------------------------------------------|
| Kamu itu berasal dari suku yang berbeda dengan kami nggak cocok berteman dengan kami. |

2. Hitung Jarak Euclidean

Menghitung jarak data training dan data uji (testing) menggunakan jarak Euclidean, dengan rumus berikut:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

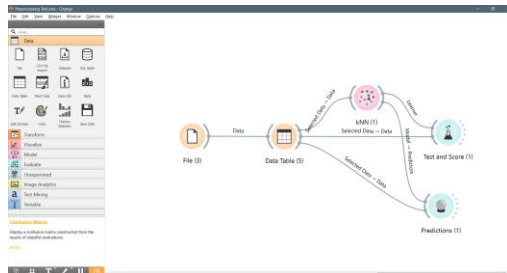
3. Pengujian Model

Model menganalisis faktor-faktor yang memengaruhi pola verbal bullying menggunakan Algoritma K-Nearest Neighbor (KNN). KNN mengklasifikasikan pola berdasarkan atribut terpilih, dengan pengujian beberapa nilai k untuk menemukan yang optimal.

Cross-validation digunakan untuk mencegah overfitting dan meningkatkan generalisasi hasil.

- a. Proses pelatihan model dengan Algoritma K-Nearest Neighbor (KNN)

Pengujian menggunakan 5.109 data hasil angket verbal bullying yang telah diproses. Dilakukan melalui aplikasi Orange Data Mining menggunakan widget KNN, Predictions, dan Test and Score.



Gambar 4. Model Pengujian “Optimaisasi Nilai k dengan Cross Validation”

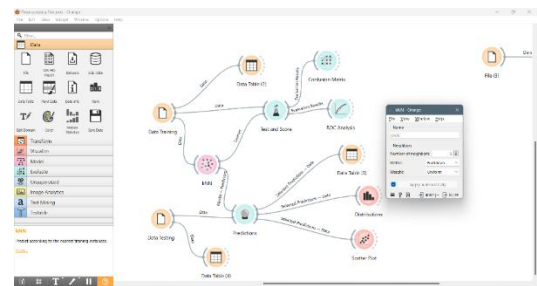
Pengujian menggunakan widget Test and Score dengan teknik cross-validation (10 folds) dan berbagai nilai k (1, 3, 5, 7, 9, 11, 13, 15). Model dilatih dan diuji pada setiap lipatan untuk setiap nilai k.

Tabel 10. Hasil Pengujian Cross Validation

| Nilai <i>k</i> | Score |           |        |       |
|----------------|-------|-----------|--------|-------|
|                | AUC   | Precision | Recall | F1    |
| 1              | 0.605 | 0.865     | 0.857  | 0.861 |
| 3              | 0.695 | 0.874     | 0.889  | 0.881 |
| 5              | 0.725 | 0.877     | 0.905  | 0.882 |
| 7              | 0.765 | 0.877     | 0.905  | 0.883 |
| 9              | 0.779 | 0.875     | 0.905  | 0.880 |
| 11             | 0.791 | 0.879     | 0.907  | 0.883 |
| 13             | 0.806 | 0.874     | 0.905  | 0.879 |
| 15             | 0.821 | 0.876     | 0.906  | 0.881 |

Nilai *k* = 15 memberikan performa terbaik dengan AUC 0,821%, precision 0,876%, recall 0,906%, dan F1-Score 0,881%. Dengan folds = 10, *k* = 15 dianggap optimal, memberikan keseimbangan ideal antara bias dan variansi untuk analisis pola kasus verbal bullying.

- b. Pengujian Semua Atribut dengan Parameter *k* = 15  
Pengujian model KNN dengan *k*=15 menggunakan seluruh atribut dataset hasil angket di Sekolah Menengah Atas dan Kjuruan di Kabupaten Minahasa yang telah diproses. Dilakukan menggunakan Orange Data Mining, pengujian ini mengevaluasi efektivitas model dalam mengklasifikasikan kasus verbal bullying berdasarkan tetangga terdekat.



Gambar 5. Pengujian Semua Atribut dengan Parameter *k*=15

Test and Score

Thu Oct 10 24, 18:31:16

Settings

Sampling type: Stratified 10-fold Cross validation

Target class: None, show average over classes

Scores

| Model | AUC   | CA    | F1    | Prec  | Recall | MCC   |
|-------|-------|-------|-------|-------|--------|-------|
| kNN   | 0.843 | 0.905 | 0.895 | 0.889 | 0.905  | 0.318 |

Gambar 6. Hasil Test and Score Pengujian Semua Atribut

Model dilatih dengan 4.088 data menggunakan cross-validation untuk mencegah bias dan overfitting, lalu diuji pada 1.021 data. Hasil evaluasi menunjukkan AUC 84,3%, precision 88,9%, recall 90,5%, dan F1-score 89,5%. Ini membuktikan KNN dengan *k*=15 memiliki performa baik dalam klasifikasi verbal bullying. Confusion Matrix digunakan untuk evaluasi lanjutan.

## 5. Asses (Evaluasi)

Tahap evaluasi model dilakukan menggunakan confusion matrix untuk menghitung akurasi melalui nilai True Positive, True Negative, False Positive, dan False Negative.

Confusion Matrix Thu Oct 10 24, 18:39:31

Confusion matrix for kNN (showing number of instances)

|        |       |           |       |       |
|--------|-------|-----------|-------|-------|
|        |       | Predicted |       |       |
|        |       | Guru      | Teman | Σ     |
| Actual | Guru  | 107       | 265   | 372   |
|        | Teman | 123       | 3592  | 3,715 |
|        | Σ     | 230       | 3,857 | 4,087 |

Gambar 7. Hasil Confusion Matrix Pengujian Model

Menghitung metrik evaluasi kinerja model :

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{PN} \\
 &= \frac{107 + 3.592}{4.087} \\
 &= \frac{3.699}{4.087} \\
 &= 0.9053 \text{ atau } 90.53\%
 \end{aligned}$$



Akurasi mengukur persentase prediksi yang benar dari total prediksi yang dilakukan.

$$\begin{aligned}\text{Presisi} &= \frac{TP}{TP + FP} \\ &= \frac{107}{107 + 123} \\ &= \frac{107}{230}\end{aligned}$$

$$= 0.4652 \text{ atau } 46.52\%$$

Presisi mengukur seberapa tepat prediksi kelas positif (Guru) dari semua yang diprediksi sebagai kelas positif.

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP + FN} \\ &= \frac{107}{107 + 265} \\ &= \frac{107}{372}\end{aligned}$$

$$= 0.2876 \text{ atau } 28.76\%$$

Recall mengukur seberapa baik model mendeteksi kelas positif (Guru) dari semua yang sebenarnya kelas positif

$$\begin{aligned}F1 &= \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \\ &= \frac{2 \times 0.465 \times 0.18}{0.465 + 0.18} \\ &= \frac{0.167}{0.645}\end{aligned}$$

$$= 0.3556 \text{ atau } 35.56\%$$

F1 Score adalah rata-rata harmonis dari Presisi dan Recall, digunakan untuk menyeimbangkan antara keduanya.

Penggunaan K-Nearest Neighbor untuk klasifikasi kasus verbal bullying di Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa menghasilkan akurasi 90.53%, presisi 46.52%, recall 28.76%, dan F1 score 32.56%. Rendahnya presisi, recall, dan F1 score disebabkan oleh data yang tidak seimbang, di mana sebagian besar kasus adalah "teman," sehingga model cenderung memprediksi "teman" untuk mencapai akurasi tinggi, tetapi kurang efektif mendeteksi kasus "guru."

## B. Pembahasan

Penelitian ini mengkaji pola verbal bullying di Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa menggunakan Algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan pelaku "guru" dan "teman." Proses data mining menghasilkan pola karakteristik setiap kategori dan mengevaluasi kinerja KNN. Perbedaan akurasi antara perhitungan manual (k=3) dan aplikasi Orange (k=15) disebabkan oleh perbedaan parameter, tingkat presisi, dan optimasi otomatis pada Orange.

## 1. Temuan Pola dari Hasil Data Mining.

Tabel 11. Pola Pelaku Guru

| Jenis Pelaku | Pola <i>Bullying</i>                | Kalimat                                                    | Tujuan                                    |
|--------------|-------------------------------------|------------------------------------------------------------|-------------------------------------------|
| Guru         | Perintah Otoritatif                 | "Disuruh bawain buku ke kantor", "Disuruh bersihin toilet" | Menunjukkan kekuasaan, merendahkan siswa  |
|              | Kritikan yang tidak Membangun       | "Bukan begitu caranya", "Penyebab masalah"                 | Menurunkan rasa percaya diri              |
|              | Menyakiti dengan kata-kata Kasar    | "Bodoh", "Pemalas"                                         | Merendahkan langsung harga diri korban    |
|              | Penggunaan Sebutan Hewan atau Benda | "Babi", "Anjing"                                           | Menghina dan menurunkan harga diri korban |

Tabel 12. Pola Pelaku Teman

| Jenis Pelaku | Pola <i>Bullying</i>                        | Kalimat                                          | Tujuan                                              |
|--------------|---------------------------------------------|--------------------------------------------------|-----------------------------------------------------|
| Teman        | Permainan Kata-kata atau Pelesetan          | "Yudi jadi judi", "Spong"                        | Menghina dengan cara halus namun tetap menyakitkan  |
|              | Memfitnah dan Menyebarkan Gosip             | "Ngomongin temenku kalo dia jarang mandi"        | Merusak reputasi sosial korban                      |
|              | Menghina Berdasarkan Fisik atau Kepercayaan | "Rambut indomie", "Kurus kayak cicak"            | Menyerang insekuritas fisik atau kepercayaan korban |
|              | Pelecehan Verbal                            | Menyebut bagian tubuh korban untuk mempermalukan | Membuat korban merasa malu atau cemas               |

## 2. Atribut Khusus dan Pengaruhnya Terhadap Model Atribut Khusus:

- Jenis Pelaku (Target): Variabel target dengan dua kelas, "Guru" dan "Teman," untuk membedakan pelaku bullying. Pengaruhnya: Memengaruhi kemampuan model KNN dalam membedakan tindakan berdasarkan jarak data berlabel.

- b. Kalimat Verbal Bullying (Text): Berisi kalimat pelaku sebagai dasar analisis teks. Pengaruhnya: Digunakan untuk mendeteksi pola bullying melalui transformasi teks, tokenisasi, dan filtering.
- c. Kategori Verbal Bullying: Meliputi kategori seperti "Direndahkan," "Diejek," dll. Pengaruhnya: Membantu mengelompokkan jenis tindakan bullying dan meningkatkan akurasi model KNN.

#### Pengaruh Terhadap Model:

- a. Akurasi: Penggunaan semua atribut menghasilkan akurasi 90,53%. "Jenis Pelaku" sebagai target utama, dengan "Kalimat Verbal Bullying" dan "Kategori" meningkatkan prediksi.
- b. Presisi & Recall: Presisi untuk pelaku "Guru" lebih rendah dibanding "Teman," mencerminkan kesulitan klasifikasi akibat ketidakseimbangan data.

### V. KESIMPULAN

Berdasarkan penelitian ini, penulis menarik kesimpulan sebagai berikut:

1. Pola kasus verbal bullying di Sekolah Menengah Atas dan Kejuruan di Kabupaten Minahasa dapat diidentifikasi menggunakan algoritma K-Nearest Neighbor (K-NN) dengan akurasi yang memadai. Pelaku terbanyak adalah "Teman" (90,70%) dan yang terendah adalah "Guru" (9,30%). Kategori bullying terbanyak adalah "Dipanggil dengan sebutan hewan/benda" (9,57%), "Diejek" (6,82%), "Dihina" (8,41%), "Dipanggil dengan sebutan nama orangtua" (11,84%), "Dipermalukan" (6,54%), dan "Direndahkan" (6,79%). Penggunaan K-NN efektif untuk menganalisis perilaku bullying di sekolah.
2. Penelitian ini menggunakan 4.088 data train dan memperoleh akurasi tertinggi  $k = 15$  pada aplikasi Orange Data Mining (84,3%) dan  $k = 3$  pada perhitungan manual (90,53%). Perbedaan akurasi disebabkan oleh pengaturan parameter yang berbeda, presisi pembulatan, dan optimasi otomatis di Orange.
3. Grafik perbandingan jumlah kasus verbal bullying menunjukkan bahwa insiden bullying lebih banyak terjadi di SMA dibandingkan SMK. Hal ini menunjukkan bahwa faktor-faktor tertentu mungkin mempengaruhi kejadian bullying di SMA.

### DAFTAR ACUAN

- Asiyah, B. (2020). The Bullying Experience in High School Students by the Teacher is reviewed from Psychological Theory. *Proceedings of The ICECRS*, 8, 1–5. <https://doi.org/10.21070/icecrs2020477>
- Cut Nurmasyithah. (2022). Hubungan Antara Iklim Sekolah Dengan Perilaku Bullying Pada Siswa Sma Negeri 1 Glumpang Tiga Kabupaten Pidie.
- Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 1(1), 30–36. <https://doi.org/10.35746/jtim.v1i1.11>
- Homepage, J., Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127.
- Hutagalung, A. S., Negara, A. B. P., & Pratama, E. E. (2021). Aplikasi Pendeteksi Cyberbullying Terhadap Komentar Postingan Media Sosial Instagram dengan Metode Naïve Bayes Classifier Berbasis Website. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 9(3), 364. <https://doi.org/10.26418/justin.v9i3.44843>
- Kevin Wilson. (2021). Proses Data Mining SEMMA. *Proses Data Mining SEMMA*. <https://sis.binus.ac.id/2021/09/30/proses-data-mining-semma/>
- Maria Susan Anggreany, S.Kom., M. K. (2022). Matriks Kebingungan. *Matriks Kebingungan*. <https://socs.binus.ac.id/2020/11/01/confusion-matrix/>
- Nikmatun, I., & Waspada, I. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *e. SIMETRIS*, 422–432.
- Oktafiani, R., Hermawan, A., & Avianto, D. (2023). Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning. *Jurnal Sains Dan Informatika*, 9(April), 19–28. <https://doi.org/10.34128/jsi.v9i1.622>
- Oktaviany, D., & Ramadan, Z. H. (2023). Analisis Dampak Bullying Terhadap Psikologi Siswa Sekolah Dasar. *Jurnal Educatio FKIP UNMA*, 9(3), 1245–1251. <https://doi.org/10.31949/educatio.v9i3.5400>
- Prastiti, J. P., & Anshori, I. (2023). Efek Sosial Dan Psikologis Perilaku Bullying. *Jurnal Sains Sosio Humaniora*, 7(1), 69–77.
- Reimon, J., & Hariguna, T. (2024). Sentimen yang Terungkap : Insentif Kendaraan Listrik di Bawah Lensa Mesin Vektor Dukungan dan Analisis TF-IDF Machine Translated by Google. 5(1), 122–132.
- seputarsulut.com. (2024). Hebat...Mahasiswa FK & Antropologi Unsrat Kaji Bullying. ByRedaksi. <https://www.seputarsulut.com/hebat-mahasiswa-fk-antropologi-unsrat-kaji-bullying/>
- Statistik, B. P., & Minahasa, K. (2023). Jumlah Sekolah, Guru, dan Murid Sekolah Menengah Kejuruan (SMK) di Bawah Kementerian Pendidikan dan Kebudayaan Menurut Kecamatan di Kabupaten Minahasa, 2019/2020. Badan Pusat Statistik. <https://minahasakab.bps.go.id/id/statistics-table/1/MzM3IzE=/jumlah-sekolah-guru-dan-murid-sekolah-menengah-kejuruan-smk-di-bawah->



---

kementerian-pendidikan-dan-kebudayaan-menurut-kecamatan-di-kabupaten-minahasa-2019-2020.html  
Teknokrat, U. I. (2024). Pengertian dan Teknik Text Mining. By FTIK. <https://ftik.teknokrat.ac.id/pengertian-dan-teknik-text-mining/>

Zakiyah, E. Z., Humaedi, S., & Santoso, M. B. (2017). Faktor Yang Mempengaruhi Remaja Dalam Melakukan Bullying. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*, 4(2), 324–330. <https://doi.org/10.24198/jppm.v4i2.14352>.