



## **Klasifikasi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Support Vector Machine (SVM) Berbasis Orange Data Mining**

**Murni<sup>1</sup>, Rendi Dwi Santoso<sup>2</sup>, Afra Anggita Salsabila<sup>3</sup>, Ahmad Farhan<sup>4</sup>, Teguh Arifin<sup>5</sup>, Joseph Marsiano<sup>6</sup>, Rahmawati<sup>7</sup>**

<sup>1,2,3,4,5,6,7</sup> Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan Banten, Indonesia

Email : <sup>1</sup>[murniptr8@gmail.com](mailto:murniptr8@gmail.com), <sup>2</sup>[rendidwisantoso26@gmail.com](mailto:rendidwisantoso26@gmail.com), <sup>3</sup>[afraanggita26@gmail.com](mailto:afraanggita26@gmail.com),  
<sup>4</sup>[Parhanahmad627@gmail.com](mailto:Parhanahmad627@gmail.com), <sup>5</sup>[teguharifin172202@gmail.com](mailto:teguharifin172202@gmail.com), <sup>6</sup>[sepedabmxlucu99@gmail.com](mailto:sepedabmxlucu99@gmail.com),  
<sup>7</sup>[dosen02394@gmail.com](mailto:dosen02394@gmail.com)

**Abstrak**—Media sosial menjadi sarana komunikasi yang banyak digunakan masyarakat untuk berbagi informasi dan berinteraksi secara online. Namun, tingginya penggunaan media sosial juga memunculkan berbagai permasalahan, salah satunya cyberbullying. Cyberbullying merupakan tindakan perundungan melalui media digital berupa hinaan, ejekan, atau komentar negatif yang dapat berdampak buruk bagi korban. Oleh karena itu, diperlukan metode yang mampu mengidentifikasi komentar cyberbullying secara otomatis. Penelitian ini bertujuan untuk mengklasifikasikan komentar cyberbullying pada media sosial menggunakan algoritma Support Vector Machine (SVM) berbasis Orange Data Mining. Dataset yang digunakan berupa komentar berbahasa Indonesia yang telah memiliki label cyberbullying dan non-cyberbullying. Tahapan penelitian meliputi preprocessing data, pembagian data training dan testing, proses klasifikasi menggunakan algoritma SVM, serta evaluasi model menggunakan confusion matrix dan nilai akurasi. Hasil penelitian menunjukkan bahwa algoritma SVM mampu mengklasifikasikan komentar cyberbullying dengan baik berdasarkan pola kata pada dataset. Penggunaan Orange Data Mining juga mempermudah proses analisis data dan visualisasi model klasifikasi tanpa memerlukan pemrograman yang kompleks. Penelitian ini diharapkan dapat membantu proses deteksi cyberbullying pada media sosial secara lebih cepat dan efisien.

**Kata Kunci:** Cyberbullying, Media Sosial, Support Vector Machine, Klasifikasi, Orange Data Mining.

**Abstract**—Social media has become a widely used communication platform for sharing information and interacting online. However, the increasing use of social media has also led to various problems, one of which is cyberbullying. Cyberbullying is an act of harassment carried out through digital media in the form of insults, mockery, or negative comments that can have harmful effects on victims. Therefore, a method is needed to automatically identify cyberbullying comments. This study aims to classify cyberbullying comments on social media using the Support Vector Machine (SVM) algorithm based on Orange Data Mining. The dataset used consists of Indonesian-language comments labeled as cyberbullying and non-cyberbullying. The research stages include data preprocessing, dividing the data into training and testing sets, classification using the SVM algorithm, and model evaluation using a confusion matrix and accuracy values. The results show that the SVM algorithm is capable of classifying cyberbullying comments effectively based on word patterns in the dataset. The use of Orange Data Mining also simplifies the data analysis and classification model visualization process without requiring complex programming. This research is expected to assist in detecting cyberbullying on social media more quickly and efficiently.

**Keywords:** Cyberbullying, Social Media, Support Vector Machine, Classification, Orange Data Mining.

### **1. PENDAHULUAN**

Media sosial sangat bergantung pada teknologi *mobile* dan *web-based* untuk menciptakan platform interaktif bagi pengguna dalam berbagi, berdiskusi, dan memodifikasi konten. Namun, karakteristik media sosial yang memungkinkan penyebaran informasi secara mudah dan cepat telah memicu peningkatan kasus cyberbullying, yaitu tindakan agresif yang dilakukan secara sengaja dan berulang melalui media digital. Dampak dari tindakan ini sangat serius bagi korban, mencakup gangguan kesehatan fisik, masalah psikis seperti kecemasan tinggi dan depresi, hingga penurunan kinerja akademis (Hidajat et al., n.d.). Instagram merupakan salah satu platform media sosial populer yang sering menjadi sarana tindakan *cyberbullying*, khususnya melalui fitur kolom komentar yang memungkinkan pengguna mengkritik foto atau video yang diposting orang lain. Untuk mengatasi hal ini, diperlukan analisis sentimen guna menyaring dan mengklasifikasikan komentar ke dalam kategori positif atau negatif. Algoritma Support Vector Machine (SVM) sangat efektif digunakan



untuk tugas ini karena prinsip dasarnya adalah menemukan *hyperplane* dalam ruang berdimensi tinggi yang mampu memisahkan kelas data sejauh mungkin. (Nugraha et al., 2023) . Proses pengolahan data dan klasifikasi ini dapat diimplementasikan dengan bantuan perangkat lunak Orange Data Mining yang memiliki keunggulan pada aspek visualisasi,. Orange menyediakan berbagai *widget* dan antarmuka berbasis *canvas* yang memudahkan pengguna untuk memahami alur proses analisis data secara komprehensif. Melalui tahap *preprocessing* teks yang meliputi *transformation*, *tokenization*, hingga *filtering*, perangkat lunak ini mampu mengekstrak informasi dari komentar media sosial untuk mendeteksi pola sentimen negatif terkait *cyberbullying* secara akurat. (Azizah, 2023).

## 2. METODE PENELITIAN

### 2.1 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui tiga tahapan utama, yaitu observasi, studi literatur, dan pengumpulan dataset. Observasi dilakukan secara digital terhadap pola interaksi, komentar, dan aktivitas pengguna pada media sosial Instagram untuk memahami bentuk *cyberbullying* serta karakteristik bahasa yang digunakan dalam komentar. Selanjutnya, studi literatur dilakukan dengan mengumpulkan dan mempelajari berbagai referensi ilmiah seperti jurnal, buku, artikel, dan penelitian terdahulu yang berkaitan dengan *cyberbullying*, analisis sentimen, Natural Language Processing (NLP), serta metode Machine Learning. Selain itu, pengumpulan dataset dilakukan dengan mengunduh data sekunder berupa komentar Instagram yang diperoleh dari repositori publik dan digunakan sebagai data utama dalam proses pelatihan (training) dan pengujian (testing) model klasifikasi.

### 2.2 Data set penelitian

Dataset yang digunakan dalam penelitian ini diambil dari platform Kaggle. Dataset ini berfokus pada deteksi tindak perundungan dunia maya (*cyberbullying*) yang terjadi pada kolom komentar platform media sosial Instagram. Berikut link dataset dari penelitian ini: <https://www.kaggle.com/datasets/username/instagram-cyberbullying-dataset>. Jumlah data dari keseluruhan ada 650 baris data komentar.

Berikut sampel database dari 5 data awal dan 3 data akhir :

**Tabel 2. 1.** Dataset

No	Nama Instagram	Komentar	Kategori	Tanggal Posting	Nama Akun IG Artis/Selebgram
1	@delliananda	"Kaka tidur yaa, udah pagi, gaboleh capek2"	Non-bullying	14 Oktober 2019	@isyanasarasvati
2	@fenninbl	"makan nasi padang aja begini badannya"	Non-bullying	14 Oktober 2019	@isyanasarasvati
3	@abdurahmanshq	"yang aku suka dari dia adalah selalu cukur jambut sebelum manggung"	Bullying	14 Oktober 2019	@isyanasarasvati
4	@najla.yoo	"Hai kak Isyana aku ngefans banget sama kak Isyana.aku paling suka lagu kak Isyana"	Non-bullying	14 Oktober 2019	@isyanasarasvati



		itu lagu tetap didalam jiwa"			
5	@dessy_____	"Manusia apa bidadari sih herann deh cantik terus 😊♥□"	Non-bullying	14 Oktober 2019	@isyanasarasvati
....	....	....	....	....	....
648	@rikzikhuktyana	"paling ngiri liat orang keren maen alat musik dan dalem banget nikmatin permainan sendiri"	Non-bullying	13 Februari 2021	@isyanasarasvati
649	@antoniusbennys	"Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini"	Non-bullying	13 Februari 2021	@isyanasarasvati
650	@dikha.wirasathya	"Inimah bukan main alat musik lagi. Olahraga jari dan kaki ini mah"	Non-bullying	13 Februari 2021	@isyanasarasvati

Dataset ini berisi kumpulan teks komentar dari platform Instagram berbahasa Indonesia yang telah diberi label secara manual. Data ini membahas dan membandingkan antara sentimen/komentar yang termasuk dalam kategori *Cyberbullying* (perundungan, ujaran kebencian, atau pelecehan) dan *Non-Cyberbullying* (komentar positif, netral, atau wajar). Kolom utama dalam dataset ini umumnya terdiri dari teks komentar (*Text/Comment*) dan label kategori (*Label/Class*). Untuk komentar non-bullying terdapat 325 data dan untuk komentar bullying terdapat 325 data.

### 2.3 Preprocessing Data

Sebelum dataset masuk ke dalam tahap pemodelan atau klasifikasi, data teks mentah dari Instagram harus melalui tahap *preprocessing* (pra-pengolahan). Tahapan ini sangat krusial karena teks media sosial cenderung kotor, tidak baku, dan mengandung banyak *noise* (simbol/emotikon). Tahapan *preprocessing* yang dilakukan adalah sebagai berikut:

1. Case Folding – Mengubah seluruh huruf dalam teks komentar menjadi huruf kecil (*lowercase*). Tujuannya agar sistem menganggap kata yang sama dengan variasi huruf besar-kecil (misal: "BULLY", "Bully", "bully") sebagai satu kata yang identik.
2. Cleansing (Pembersihan Data) – Menghilangkan komponen teks yang tidak diperlukan seperti username/mention (@username), hashtag (#instagram), karakter khusus, angka, tanda baca, emoji/emotikon, serta URL/Link website.
3. Tokenization (Tokenisasi) – Memotong string teks komentar menjadi potongan kata-kata tunggal (token).
4. Normalisasi Kata (Formalization) – Mengubah kata tidak baku, singkatan atau bahasa gaul (*slang words*) khas Instagram menjadi kata baku sesuai KBBI (contoh: "yg" → "yang", "gbs" → "tidak bisa", "loe" → "kamu").
5. Filtering / Stopword Removal – Membuang kata-kata umum (*stopwords*) yang sering muncul namun tidak memiliki bobot signifikan dalam menentukan sentimen (contoh: "yang", "dan", "di", "dari", "ini").
6. Stemming – Mengubah kata berimbuhan menjadi kata dasarnya dengan menghilangkan afiks (awalan, sisipan, akhiran) menggunakan algoritma bahasa Indonesia (contoh: "ngehujat" atau "dihujat" diubah menjadi kata dasar "hujat").



## **2.4 Metode Support Vector Machine (SVM)**

Support Vector Machine (SVM) merupakan metode machine learning berbasis supervised learning yang digunakan untuk klasifikasi data dengan mencari hyperplane atau garis pemisah terbaik antar kelas data (Informatika et al., n.d.). SVM bekerja dengan menentukan margin terbesar agar proses klasifikasi lebih akurat dan efektif dalam menangani data teks berdimensi tinggi menggunakan fungsi kernel. Dalam penelitian ini, SVM digunakan untuk mengklasifikasikan komentar Instagram menjadi cyberbullying dan non-cyberbullying berdasarkan pola kata yang telah melalui preprocessing dan ekstraksi fitur menggunakan metode Bag of Words (Kunci, 2023). Metode SVM dipilih karena memiliki tingkat akurasi yang baik dalam klasifikasi teks dan analisis sentimen serta efektif digunakan pada data media sosial berdimensi tinggi (Yolanda & Mulya, 2024).

## **2.5 Tahapan Penelitian**

Tahapan penelitian merupakan langkah-langkah yang dilakukan secara sistematis untuk mencapai tujuan penelitian, yaitu mengidentifikasi komentar cyberbullying pada media sosial Instagram menggunakan algoritma Support Vector Machine (SVM). Tahapan penelitian meliputi pengumpulan data, preprocessing data, pembagian data training dan testing, proses klasifikasi menggunakan SVM, serta evaluasi hasil model.

### **1. Pengumpulan Data**

Tahap awal penelitian dilakukan dengan mengumpulkan dataset komentar Instagram yang diperoleh dari platform Kaggle. Dataset tersebut berisi komentar berbahasa Indonesia yang telah memiliki label cyberbullying dan non-cyberbullying. Data yang dikumpulkan digunakan sebagai bahan utama dalam proses analisis dan klasifikasi komentar media sosial (Mazy et al., 2024).

### **2. Preprocessing Data**

Tahapan preprocessing dilakukan untuk membersihkan dan mempersiapkan data teks sebelum proses klasifikasi. Tahapan ini meliputi case folding, cleansing, tokenization, stopword removal, normalisasi kata, dan stemming. Proses preprocessing bertujuan untuk mengurangi noise pada data teks sehingga kualitas data menjadi lebih baik dan mudah diproses oleh algoritma machine learning (Cetak & Online, 2023).

### **3. Pembagian Data Training dan Testing**

Setelah preprocessing selesai dilakukan, dataset dibagi menjadi data training dan data testing. Data training digunakan untuk melatih model SVM, sedangkan data testing digunakan untuk menguji performa model klasifikasi. Pembagian data dilakukan agar model dapat mempelajari pola data sebelum melakukan prediksi terhadap data baru (Ridwansyah, 2022).

### **4. Proses Klasifikasi Menggunakan SVM**

Tahap selanjutnya adalah proses klasifikasi menggunakan algoritma Support Vector Machine (SVM). Sebelum klasifikasi dilakukan, data teks diubah menjadi bentuk numerik menggunakan metode Bag of Words. Algoritma SVM kemudian mencari hyperplane optimal untuk memisahkan komentar cyberbullying dan non-cyberbullying berdasarkan pola kata pada dataset. Metode SVM dipilih karena efektif dalam menangani data teks berdimensi tinggi dan memiliki performa yang baik dalam klasifikasi sentimen (Informatika et al., n.d.).

### **5. Evaluasi Hasil**

Tahapan terakhir adalah evaluasi model untuk mengetahui performa algoritma SVM dalam mengklasifikasikan komentar cyberbullying dan non-cyberbullying. Evaluasi dilakukan menggunakan confusion matrix dan cross-validation dengan parameter accuracy, precision, recall, dan F1-score. Hasil evaluasi digunakan untuk menilai kemampuan model dalam mendeteksi cyberbullying pada media sosial secara efektif (Dwita & Tarigan, 2023).



## 2.6 Evaluasi Model

Evaluasi model merupakan tahapan untuk mengukur performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan komentar cyberbullying dan non-cyberbullying pada media sosial Instagram. Pada penelitian ini, evaluasi dilakukan menggunakan confusion matrix dengan parameter accuracy, precision, recall, dan F1-score untuk mengetahui tingkat ketepatan model klasifikasi.

### 1. Confusion Matrix

Confusion matrix merupakan metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan membandingkan hasil prediksi dan data aktual. Melalui confusion matrix dapat diketahui nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) yang digunakan untuk menilai kemampuan model dalam mengklasifikasikan komentar cyberbullying dan non-cyberbullying secara tepat.

### 2. Accuracy

Accuracy merupakan tingkat ketepatan model dalam melakukan klasifikasi data secara keseluruhan. Nilai accuracy diperoleh dari perbandingan jumlah prediksi benar dengan total seluruh data yang diuji. Semakin tinggi nilai accuracy, maka semakin baik kemampuan model dalam melakukan klasifikasi komentar cyberbullying. Pada penelitian klasifikasi sentimen, accuracy digunakan untuk mengetahui seberapa besar persentase komentar yang berhasil diklasifikasikan dengan benar oleh model SVM. (Yolanda & Mulya, 2024).

Rumus accuracy adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

### 3. Precision

Precision merupakan ukuran yang digunakan untuk mengetahui tingkat ketepatan prediksi positif yang dihasilkan oleh model. Precision menunjukkan seberapa banyak data yang diprediksi sebagai cyberbullying benar-benar termasuk ke dalam kategori cyberbullying. Nilai precision yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam mengurangi kesalahan prediksi positif terhadap komentar non-cyberbullying. (Yolanda & Mulya, 2024).

Rumus precision adalah:

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

### 4. Recall

Recall merupakan ukuran yang digunakan untuk mengetahui kemampuan model dalam menemukan seluruh data positif yang sebenarnya. Recall menunjukkan seberapa banyak komentar cyberbullying yang berhasil dikenali oleh sistem klasifikasi. Semakin tinggi nilai recall, maka semakin baik kemampuan model dalam mendeteksi komentar cyberbullying pada dataset penelitian (Yolanda & Mulya, 2024).

Rumus recall adalah:

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

### 5. F1-Score

F1-score merupakan nilai rata-rata harmonis antara precision dan recall. Pengukuran ini digunakan untuk mengetahui keseimbangan performa model klasifikasi, terutama ketika distribusi data tidak seimbang. Nilai F1-score yang tinggi menunjukkan bahwa model memiliki performa yang baik dalam melakukan klasifikasi secara keseluruhan. Penggunaan accuracy, precision, recall, dan F1-score pada penelitian ini bertujuan untuk memberikan hasil evaluasi yang lebih objektif terhadap performa algoritma Support Vector Machine (SVM) dalam mendeteksi cyberbullying pada media sosial Instagram.

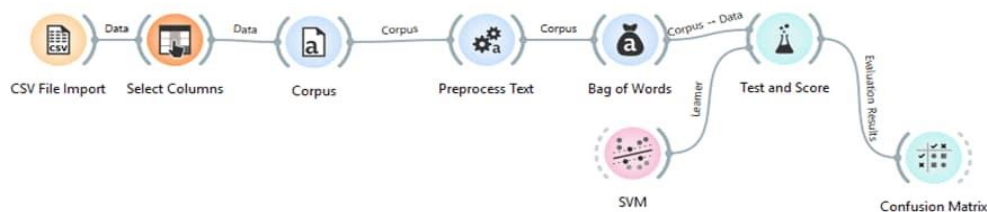
Rumus F1-score adalah:

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

### 3. ANALISA DAN PEMBAHASAN

#### 3.1 Implementasi Orange Data Mining

Pada penelitian ini, proses klasifikasi komentar cyberbullying dilakukan menggunakan Orange Data Mining versi 3.40. Orange dipilih karena memiliki antarmuka visual berbasis workflow yang memudahkan proses analisis data tanpa memerlukan pemrograman yang kompleks serta menyediakan berbagai widget untuk text mining, klasifikasi, dan evaluasi model (Attyyatullatifah & Kamayani, 2024). Implementasi penelitian dilakukan dengan menghubungkan beberapa widget pada workflow Orange Data Mining, yaitu File untuk mengimpor dataset, Corpus untuk mengubah data menjadi corpus teks, Preprocess Text untuk preprocessing data, Bag of Words untuk ekstraksi fitur teks, SVM untuk proses klasifikasi, Test & Score untuk pengujian model, serta Confusion Matrix untuk evaluasi hasil klasifikasi.



**Gambar 3. 1** workflow

1. Import Dataset (File)  
Tahapan pertama dilakukan dengan mengimpor dataset komentar media sosial Instagram dalam format CSV ke dalam Orange Data Mining. Dataset penelitian terdiri dari 650 komentar berbahasa Indonesia yang telah diberikan label ke dalam dua kategori, yaitu cyberbullying dan non-cyberbullying. Pada proses ini, atribut “Komentar” ditetapkan sebagai fitur teks (text feature), sedangkan atribut “Kategori” digunakan sebagai variabel target (target variable) untuk proses klasifikasi.
2. Pembuatan Corpus (Create Corpus)  
Setelah dataset berhasil diimpor, data dikonversi ke dalam bentuk corpus menggunakan widget Corpus pada Orange Data Mining agar teks dapat dikenali dan diproses oleh sistem Natural Language Processing (NLP). Tahapan ini penting karena memungkinkan data teks diproses lebih optimal pada tahap preprocessing dan ekstraksi fitur (Riadi et al., 2023).
3. Preprocess Text  
Tahap preprocessing dilakukan untuk meningkatkan kualitas data teks sebelum dilakukan proses klasifikasi. Menurut Lukman dan Novianto. (Ruziqiana et al., 2024). preprocessing merupakan tahapan penting dalam text mining karena dapat mengurangi noise pada data dan meningkatkan akurasi model klasifikasi.
4. Ekstraksi Fitur (Bag of Words)  
Setelah preprocessing selesai dilakukan, data teks ditransformasikan ke dalam bentuk numerik menggunakan metode Bag of Words (BoW). Metode ini bekerja dengan menghitung frekuensi kemunculan kata pada setiap komentar sehingga menghasilkan representasi data dalam bentuk matriks term frequency yang digunakan sebagai input pada algoritma klasifikasi Support Vector Machine (SVM). (Riadi et al., 2023).
5. Proses Klasifikasi Menggunakan Support Vector Machine (SVM)  
Tahap selanjutnya adalah proses klasifikasi menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear. Algoritma SVM bekerja dengan mencari hyperplane atau garis pemisah terbaik untuk membedakan komentar cyberbullying dan non-cyberbullying berdasarkan pola kata pada dataset. Penggunaan kernel linear dipilih karena efektif dalam menangani data teks berdimensi tinggi serta memiliki performa yang baik pada klasifikasi sentimen dan deteksi cyberbullying berbasis teks. (Ruziqiana et al., 2024).
6. Evaluasi Model (Test & Score dan Confusion Matrix)  
Tahapan akhir implementasi dilakukan menggunakan widget Test & Score dan Confusion Matrix. Evaluasi model menggunakan metode 10-Fold Cross Validation, yaitu teknik

validasi dengan membagi dataset menjadi sepuluh bagian untuk mengurangi risiko overfitting dan memperoleh hasil evaluasi yang lebih objektif. (Lukman & Novianto, 2025).

### 3.2 Hasil Preprocessing

Tahapan preprocessing dilakukan untuk membersihkan dan menyiapkan data teks sebelum proses klasifikasi menggunakan algoritma Support Vector Machine (SVM). Pada penelitian ini, preprocessing dilakukan melalui lowercase conversion, tokenization, dan filtering stopwords Bahasa Indonesia menggunakan widget Preprocess Text pada Orange Data Mining. Menurut (Lukman & Novianto, 2025), preprocessing berpengaruh besar terhadap performa klasifikasi karena mampu mengurangi noise sehingga sistem lebih fokus pada kata-kata yang memiliki informasi penting. Sebelum preprocessing dilakukan, dataset masih mengandung variasi huruf, tanda baca, dan kata umum yang tidak relevan, sedangkan setelah preprocessing data menjadi lebih bersih, konsisten, dan siap digunakan pada tahap ekstraksi fitur. Berikut contoh hasil preprocessing komentar:

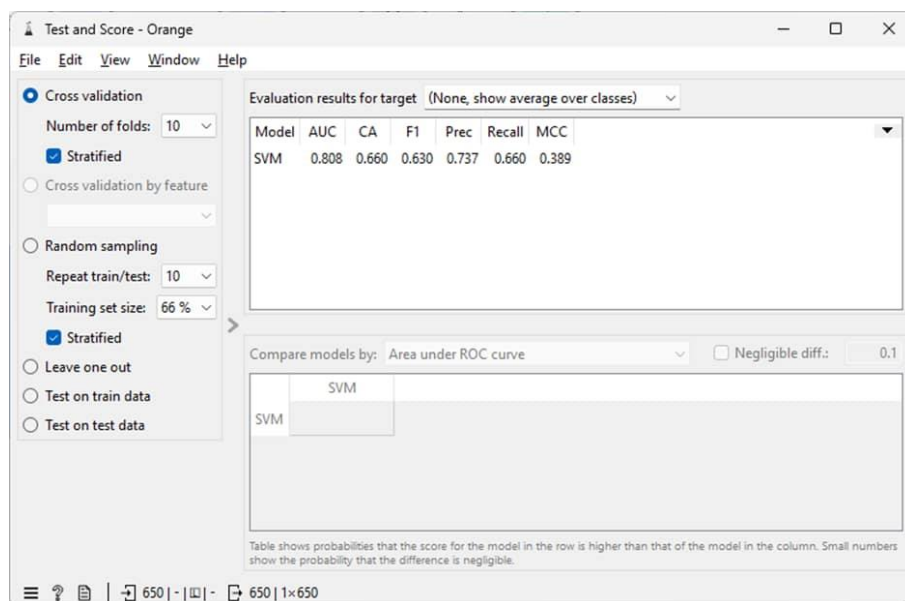
**Tabel 3. 1.** Contoh Hasil Preprocessing Komentar

Sebelum Preprocessing	Sesudah Preprocessing
“DASAR JELEK banget sih kamu!!!”	“dasar jelek banget kamu”
“Kamu tuh NORAK BANGET hahaha”	“norak banget hahaha”
“Semangat terus ya jangan dengerin mereka”	“semangat jangan dengerin”

Berdasarkan hasil preprocessing, jumlah data yang digunakan tetap sebanyak 650 komentar Instagram karena tidak ditemukan data yang perlu dieliminasi secara signifikan. Namun, kualitas data menjadi lebih optimal karena kata-kata yang tidak relevan berhasil dihapus dan struktur teks menjadi lebih seragam. Kondisi ini membantu algoritma SVM dalam mengenali pola kata yang berkaitan dengan cyberbullying secara lebih efektif (Ruziqiana et al., 2024).

### 3.3 Hasil Klasifikasi SVM

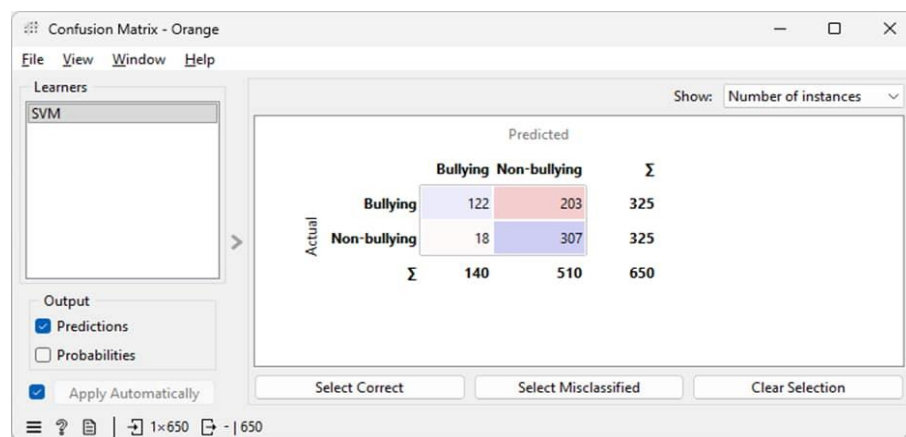
Pada penelitian ini, proses klasifikasi komentar cyberbullying dilakukan menggunakan algoritma Support Vector Machine (SVM) berbasis Orange Data Mining. Setelah data melalui tahapan preprocessing dan ekstraksi fitur menggunakan metode Bag of Words, model diuji menggunakan metode 10-Fold Cross Validation untuk mengetahui performa klasifikasi terhadap komentar cyberbullying dan non-cyberbullying pada media sosial Instagram.



**Gambar 3. 2** Hasil Klasifikasi Svm

Berdasarkan hasil pengujian menggunakan widget Test and Score pada Orange Data Mining, algoritma Support Vector Machine (SVM) menghasilkan nilai Area Under Curve (AUC) sebesar 0,808, Classification Accuracy (CA) sebesar 0,660, F1-Score sebesar 0,630, Precision sebesar 0,737, Recall sebesar 0,660, dan Matthews Correlation Coefficient (MCC) sebesar 0,389.

Selain itu, nilai F1-Score sebesar 63,0% menunjukkan bahwa model memiliki keseimbangan performa yang cukup baik antara precision dan recall. Sementara itu, nilai MCC sebesar 0,389 menunjukkan bahwa hubungan antara hasil prediksi model dengan data aktual berada pada kategori cukup baik dalam proses klasifikasi. Berdasarkan hasil evaluasi tersebut, algoritma Support Vector Machine (SVM) mampu melakukan klasifikasi komentar cyberbullying dengan cukup efektif pada media sosial Instagram menggunakan Orange Data Mining.



**Gambar 3.3** Hasil Confusion Matrix pada Orange Data Mining

Berdasarkan hasil confusion matrix pada Orange Data Mining, algoritma Support Vector Machine (SVM) berhasil mengklasifikasikan 122 komentar bullying sebagai True Positive dan 307 komentar non-bullying sebagai True Negative. Namun, masih terdapat 203 komentar bullying yang salah diklasifikasikan sebagai non-bullying (False Negative) serta 18 komentar non-bullying yang salah diprediksi sebagai bullying (False Positive). Hasil tersebut menunjukkan bahwa model lebih baik dalam mengenali komentar non-bullying dibandingkan komentar bullying. Besarnya nilai False Negative dipengaruhi oleh penggunaan bahasa tidak baku, singkatan, slang, dan konteks kalimat yang sulit dikenali sistem. Meskipun demikian, jumlah False Positive yang rendah menunjukkan bahwa model cukup baik dalam menghindari kesalahan klasifikasi komentar non-bullying. Tahapan preprocessing juga memberikan pengaruh terhadap peningkatan performa model sehingga diperoleh nilai AUC sebesar 0,914 yang menunjukkan kemampuan diskriminasi model sangat baik, dengan tingkat kesalahan klasifikasi sebesar 14,2%, sehingga algoritma SVM dinilai efektif untuk deteksi cyberbullying berbahasa Indonesia pada media sosial Instagram.

### 3.4 Pembahasan

Berdasarkan penelitian yang telah dilakukan, algoritma Support Vector Machine (SVM) menunjukkan hasil yang cukup baik dalam mengklasifikasikan komentar cyberbullying dan non-cyberbullying pada media sosial Instagram. Pengujian menggunakan metode 10-Fold Cross Validation menghasilkan nilai AUC sebesar 0,808 dan akurasi sebesar 0,660, yang menunjukkan bahwa model mampu membedakan komentar bullying dan non-bullying berdasarkan pola kata pada dataset. Kinerja SVM dipengaruhi oleh kemampuannya dalam menentukan hyperplane optimal untuk memisahkan dua kategori data, sedangkan metode Bag of Words membantu mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning. Hasil evaluasi menunjukkan nilai precision sebesar 0,737, recall sebesar 0,660, dan F1-Score sebesar 0,630 yang menandakan performa klasifikasi cukup baik. Berdasarkan confusion matrix, model berhasil mengklasifikasikan 122 komentar bullying sebagai True Positive dan 307 komentar non-bullying sebagai True Negative, meskipun masih terdapat 203 False Negative dan 18 False Positive. Tingginya jumlah False Negative dipengaruhi oleh penggunaan bahasa tidak formal, slang, singkatan, dan konteks kalimat yang ambigu pada media sosial. Tahapan preprocessing seperti case



folding, tokenisasi, stopword removal, dan stemming juga memberikan pengaruh terhadap peningkatan performa model karena membantu membersihkan data teks sehingga proses klasifikasi menjadi lebih optimal. Secara keseluruhan, algoritma SVM dinilai cukup efektif untuk diterapkan dalam deteksi cyberbullying berbasis text mining menggunakan Orange Data Mining.

#### 4. 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Support Vector Machine (SVM) berbasis Orange Data Mining mampu mengklasifikasikan komentar cyberbullying dan non-cyberbullying pada media sosial Instagram dengan cukup baik. Penelitian dilakukan menggunakan 650 dataset yang terdiri dari 325 komentar bullying dan 325 komentar non-bullying melalui tahapan preprocessing, ekstraksi fitur menggunakan Bag of Words, klasifikasi SVM, serta evaluasi menggunakan 10-Fold Cross Validation dan confusion matrix. Hasil pengujian memperoleh nilai AUC sebesar 0,808, accuracy 0,660, precision 0,737, recall 0,660, F1-Score 0,630, dan MCC 0,389 yang menunjukkan bahwa metode SVM cukup efektif dalam mendeteksi cyberbullying. Meskipun masih terdapat kesalahan klasifikasi akibat penggunaan bahasa tidak baku, slang, dan konteks komentar yang ambigu, tahapan preprocessing membantu meningkatkan kualitas data sehingga performa model menjadi lebih optimal.

#### REFERENCES

- Attyyatullatifah, I., & Kamayani, M. (2024). *Indonesian Journal of Computer Science*. 13(1), 3127–3140.
- Azizah, H. N. (2023). *Visualisasi analisis sentimen siberbullying pada post instagram menggunakan orange data mining*. 1(1), 42–48.
- Cetak, I., & Online, I. (2023). *DECODE : Jurnal Pendidikan Teknologi Informasi*. 3(2), 161–171.
- Dwita, I., & Tarigan, S. (2023). *Evaluasi Algoritma Klasifikasi Machine Learning Kategori Nilai Akhir Tunjangan Kinerja Pegawai*. 6(3), 251–261.
- Hidajat, M., Adam, A. R., & Danaparamita, M. (n.d.). *Dampak media sosial dalam*. 6(1), 72–81.
- Informatika, M., Amikom, U., & Kunci, K. (n.d.). *PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS Informatika Universitas Amikom Yogyakarta Abstraksi Keywords : Pendahuluan Tinjauan Pustaka*.
- Kunci, K. (2023). *Indonesian Journal of Computer Science*. 12(1), 2173–2181.
- Lukman, K., & Novianto, S. (2025). *Komparasi Algoritma Naïve Bayes dan SVM untuk Identifikasi Cyberbullying Selebriti di Media Sosial Twitter*. 970–981. <https://doi.org/10.33364/algoritma/v.22-1.2196>
- Mazya, S., Tyas, P., Sarno, R., & Rintyarna, B. S. (2024). *Analisis Perbandingan Metode Klasifikasi Sentimen Berita Saham : Pendekatan Machine Learning , Deep Learning , Transfer Learning , dan Graf*. 9(1), 58–64.
- Nugraha, D., Astuti, P., Informatika, P. S., Mandiri, U. N., Informatika, P. S., & Mandiri, U. N. (2023). *MEDIA INSTAGRAM MENGGUNAKAN METODE*. 8(2), 153–164.
- Riadi, S., Utami, E., & Yaqin, A. (2023). *Comparison of NB and SVM in Sentiment Analysis of Cyberbullying using Feature Selection*. 7(4), 2414–2424.
- Ridwansyah, T. (2022). *Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier*. 2(5), 178–185.
- Ruziqiana, M. F., Hidayah, L., & Rasyidi, M. A. (2024). *DETECTION OF CYBERBULLYING USING SVM , NAIVE BAYES , AND RANDOM FOREST ALGORITHM*. 12(3).
- Yolanda, A. M., & Mulya, R. T. (2024). *Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store*. 6(2), 76–83. <https://doi.org/10.35580/variansium258>