

ANALISIS SENTIMEN ULASAN PELECEHAN DI MEDIA SOSIAL TIKTOK MENGGUNAKAN SVM DAN PENGELOMPOKAN POLA KEJADIAN DENGAN K-MEANS CLUSTERING

Ayu Oktaviyanti^{1*}, Firmansyah², Abdul Rahman Kadafi³

^{1,2,3}Universitas Bina Sarana Informatika; Jl. Margonda No.8, Pondok Cina, Kecamatan Beji, Kota Depok, Jawa Barat 16424; Telp. (021) 78893140

Keywords:

Kekerasan seksual, analisis sentimen, K-Means, Support Vector Machine, media sosial.

Correspondent Email:

ayuoktaviyanti08@gmail.com

Abstrak. Kekerasan seksual merupakan isu sosial kompleks dengan sebaran yang tidak merata di Indonesia. Penelitian ini bertujuan untuk memetakan pola kasus kekerasan seksual per provinsi menggunakan algoritma K-Means Clustering serta menganalisis sentimen masyarakat terhadap isu pelecehan seksual melalui komentar di media sosial TikTok dengan metode Support Vector Machine (SVM). Data kasus diperoleh dari statistik provinsi, sedangkan komentar dikumpulkan melalui proses scraping, kemudian dipraolah dan diberi label sentimen. Hasil menunjukkan bahwa K-Means berhasil mengelompokkan provinsi ke dalam tiga kategori, yaitu Aman, Sedang, dan Rawan. Model SVM untuk klasifikasi sentimen menghasilkan akurasi 90%, dengan distribusi 7,8% negatif, 10,8% positif, dan 81,4% netral. Model juga mampu mendeteksi sentimen negatif dengan precision 0,83 dan recall 0,65. Temuan ini menunjukkan bahwa gabungan K-Means dan SVM efektif dalam mendukung pemetaan wilayah rawan kekerasan serta memahami persepsi publik berbasis data. Penelitian ini memberikan kontribusi terhadap pemanfaatan data mining sebagai alat pendukung pengambilan keputusan pada isu kekerasan seksual.



Copyright © JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. Sexual violence is a complex social issue with uneven distribution across Indonesian provinces. This study aims to map the spatial distribution of sexual violence cases using the K-Means Clustering algorithm and analyze public sentiment toward sexual harassment issues through TikTok comments using the Support Vector Machine (SVM) method. Case data were obtained from provincial statistics, while comments were collected via web scraping, followed by preprocessing and sentiment labeling. The results indicate that K-Means successfully classified provinces into three categories: Safe, Moderate, and Vulnerable. The SVM model achieved 90% accuracy in sentiment classification, with 7.8% negative, 10.8% positive, and 81.4% neutral sentiments. Additionally, the model detected negative comments with a precision of 0.83 and recall of 0.65. These findings demonstrate that combining K-Means and SVM is effective for identifying high-risk regions and understanding public perception in digital spaces. This study contributes to the application of data mining as a decision-support tool for addressing sexual violence.

1. PENDAHULUAN

Kasus kekerasan, baik seksual, fisik, maupun psikis, masih menjadi permasalahan

serius di Indonesia. Data Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KemenPPPA) serta Komnas Perempuan

menunjukkan tren peningkatan setiap tahun, terutama pada kasus kekerasan seksual yang banyak menimpa perempuan dan anak. Hal ini menandakan sistem perlindungan masyarakat belum optimal dalam pencegahan dan penanganan[1].

Media sosial, khususnya TikTok, kini menjadi ruang publik utama dalam menyuarakan pendapat. Komentar yang muncul dapat menggambarkan opini dan emosi kolektif masyarakat terhadap isu kekerasan[2]. Melalui analisis sentimen, kecenderungan opini publik dapat dipetakan untuk memahami dukungan, kecaman, atau sikap netral masyarakat.

Penelitian sebelumnya telah menggunakan *Metode Support Vector Machine* (SVM) dalam analisis sentimen karena kemampuannya menangani data teks yang kompleks dengan tingkat akurasi tinggi[3]. Penelitian lain juga memanfaatkan algoritma K-Means Clustering dalam pemetaan pola kejadian, khususnya pada data spasial. Namun, penelitian spesifik yang menggabungkan SVM dan K-Means untuk menganalisis komentar publik mengenai pelecehan seksual di Indonesia, khususnya di TikTok.

Penelitian ini menggabungkan dua pendekatan. Pertama, analisis sentimen komentar TikTok menggunakan algoritma Support Vector Machine (SVM) yang terbukti efektif dalam klasifikasi teks berdimensi besar[4]. Kedua, pemetaan sebaran kasus kekerasan seksual berdasarkan provinsi menggunakan K-Means Clustering[5].

Dengan pendekatan ini, penelitian diharapkan mampu memberikan gambaran persepsi publik sekaligus memetakan daerah rawan, sehingga dapat menjadi dasar strategi pencegahan dan kebijakan publik yang lebih tepat sasaran.

2. TINJAUAN PUSTAKA

Landasan teori merupakan dasar konseptual dan metodologis yang mendasari pelaksanaan penelitian ini. Penelitian ini terbagi menjadi dua bagian utama: analisis sebaran kasus kekerasan seksual per provinsi menggunakan algoritma K-Means Clustering, dan analisis sentimen komentar masyarakat di media sosial TikTok mengenai pelecehan menggunakan algoritma SVM. Oleh karena itu, dalam bab ini akan diuraikan teori dan konsep mengenai data mining, K-Means, SVM, analisis sentimen,

serta penelitian-penelitian terdahulu yang relevan.

2.1 Analisis Sentimen

Analisis sentimen (*sentiment analysis*) merupakan cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini, emosi dalam suatu teks. Pada umumnya, analisis sentimen dilakukan untuk mengetahui apakah suatu teks bersifat positif, netral, atau negatif[6]. Metode ini sangat berguna untuk mengevaluasi persepsi publik terhadap produk, layanan, maupun isu sosial seperti kontroversi artis di media sosial. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengklasifikasikan komentar TikTok terhadap tingginya kekerasan seksual pada anak.

2.2 Media Sosial

Media sosial merupakan platform interaktif berbasis internet yang memungkinkan pengguna untuk saling berbagi konten, opini, dan informasi dalam waktu nyata. TikTok, sebagai salah satu platform media sosial yang paling banyak digunakan di Indonesia, menyediakan ruang bagi penggunanya untuk berekspresi melalui komentar dan reaksi terhadap unggahan. TikTok termasuk platform media sosial yang memungkinkan pengguna berbagi video dan foto pendek dengan pengikut mereka baik secara publik maupun pribadi untuk memperluas jaringan sosial mereka dengan mengikuti akun orang lain dan juga memperbolehkan pengguna untuk berinteraksi dengan konten yang diunggah melalui berbagai opsi seperti "like", "komentar", "repost", dan sebagainya[7].

2.3 Text Preprocessing

Text preprocessing merupakan tahap awal dimana dalam analisis teks untuk membersihkan (*cleaning*) data untuk mengatasi masalah seperti data hilang atau outlier untuk memahai fitur dari elemen-elemen yang tidak relevan atau mengganggu, seperti simbol, URL, angka, serta mengubah huruf kapital menjadi huruf kecil[8]. Proses ini biasanya mencakup case folding, remove number, tokenization, stopword removal, stemming[9]. Langkah-langkah ini penting agar data komentar yang digunakan dalam analisis sentimen dapat diolah secara akurat oleh model machine learning.

2.4 Support Vector Machine (SVM)

Support Vector Machine adalah algoritma supervised learning yang digunakan untuk

klasifikasi dan regresi. SVM bekerja karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan hyperplane yang optimal untuk klasifikasi. terbaik yang dapat memisahkan data dari dua atau lebih kelas[10]. Dalam analisis sentimen, SVM menunjukkan performa yang baik karena mampu menangani data teks berdimensi tinggi yang direpresentasikan dalam bentuk vektor, misalnya dengan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Dibandingkan metode lain seperti Naïve Bayes, SVM cenderung memberikan hasil yang lebih akurat dalam skenario klasifikasi yang kompleks.

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode transformasi data teks menjadi bentuk numerik untuk digunakan dalam model klasifikasi bertujuan untuk memberikan skor yang lebih tinggi kepada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang muncul dalam dokumen lain di dalam koleksi[11]. *Term Frequency* (TF) mengukur seberapa sering suatu kata muncul dalam dokumen, *sedangkan Inverse Document Frequency* (IDF) menghitung seberapa jarang kata tersebut muncul di seluruh korpus. Kombinasi dari kedua nilai ini memberikan bobot yang lebih tinggi pada kata-kata yang penting dan khas, serta mengurangi pengaruh dari kata-kata umum. TF-IDF sangat efektif untuk merepresentasikan teks komentar dalam bentuk vektor yang dapat dibaca oleh SVM.

2.6 Evaluasi Model Machine Learning

Dalam pengujian kinerja model klasifikasi seperti SVM, evaluasi performa sangat penting untuk memastikan bahwa model bekerja secara optimal dalam mengklasifikasikan data. Metrik yang umum digunakan antara lain adalah:

1. Akurasi: proporsi prediksi yang benar dibandingkan total data.
2. Precision: rasio prediksi positif yang benar dibandingkan semua prediksi positif.
3. Recall: rasio prediksi positif yang benar dibandingkan semua data positif sebenarnya.
4. F1-Score: rata-rata harmonis antara precision dan recall, yang memberikan gambaran seimbang dari performa model terutama saat terdapat ketidakseimbangan kelas.

Penggunaan metrik akurasi dapat menyesatkan jika distribusi kelas tidak seimbang, karena model bisa tampak “baik” hanya karena mengklasifikasikan kelas dominan. Oleh karena itu, evaluasi model yang komprehensif menjadi bagian penting dalam penelitian klasifikasi sentiment[12].

2.7 Data Mining

Data mining merupakan proses ekstraksi pola yang berguna dari sekumpulan data yang besar. Dalam konteks penelitian ini, data mining digunakan untuk mengekstrak pola dari data kekerasan seksual dan komentar media sosial. Data mining memiliki beberapa proses diantaranya forensic data (*association*), mengidentifikasi kelompok tertentu (*clustering*), menemukan data untuk memprediksinya (*forecasting*), menyeter dan menemukan data kedalam kelompok tertentu (*classification*)[13]. Machine learning sebagai cabang dari artificial intelligence memungkinkan komputer untuk belajar dari data tanpa diprogram secara eksplisit.

2.8 Clustering dan Algoritma K-Means

Clustering adalah teknik unsupervised learning yang bertujuan untuk mengelompokkan data ke dalam cluster-cluster berdasarkan kesamaan karakteristik. Objek dalam satu cluster memiliki kesamaan yang tinggi, sedangkan objek dalam cluster yang berbeda memiliki perbedaan yang signifikan.

K-Means adalah algoritma klasterisasi yang membagi data ke dalam sejumlah klaster (*k*) berdasarkan kedekatan jarak antar data. Prosesnya dimulai dengan inisialisasi centroid secara acak, dilanjutkan dengan pengelompokan data berdasarkan jarak Euclidean ke centroid terdekat, dan dilakukan iterasi sampai centroid konvergen[14]. K-Means digunakan dalam penelitian ini untuk mengelompokkan provinsi di Indonesia berdasarkan jumlah kasus kekerasan seksual. Hal ini dilakukan untuk melihat klaster yang paling dominan atau mendominasi dalam dataset. Dengan melihat hasil klasterisasi, diperoleh pemahaman yang lebih baik tentang sebaran data dan dapat mengidentifikasi klaster yang paling signifikan dalam konteks kekerasan seksual.

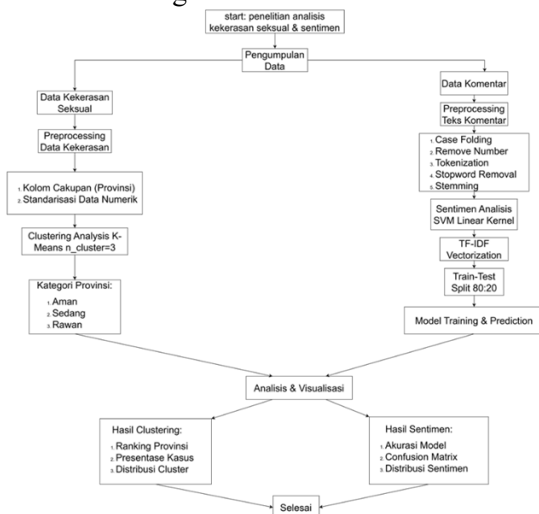
2.9 Google Colab

Google Colab atau Colaboratory sebuah platform berbasis cloud yang memungkinkan pengguna untuk menulis dan mengeksekusi

kode Python di browser tanpa perlu menginstal perangkat lunak apapun[15]. Tujuan utamanya adalah mempermudah pengembangan dan eksperimen dalam bidang data science, machine learning, dan deep learning. Hal ini memudahkan pengguna untuk mengembangkan, menjalankan, dan berbagi kode Python secara interaktif (Ashari et al., 2023).

3. METODE PENELITIAN

3.1 Proses dan langkah Penelitian



Gambar 1. Alur Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan teknik data mining dalam dua bagian utama yaitu analisis sebaran kasus kekerasan seksual antar provinsi menggunakan K-Means Clustering, dan analisis sentimen komentar publik terkait isu pelecehan seksual di TikTok menggunakan SVM. Adapun langkah-langkah dalam penelitian ini dijabarkan dalam beberapa tahapan.

3.2 Analisis Clustering Kekerasan Seksual

Analisis clustering adalah teknik unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Dalam penelitian ini, clustering digunakan untuk mengidentifikasi pola sebaran kasus kekerasan seksual di tingkat provinsi dan mengkategorikan provinsi-provinsi berdasarkan tingkat risiko kekerasan seksual.

1. Pengambilan Data

Dalam penelitian ini, data yang di dapatkan dari laman website KEMENPPA Oktober 2024. Data yang diperoleh mencakup informasi kasus kekerasan di seluruh provinsi Indonesia dengan berbagai jenis kekerasan meliputi kekerasan

seksual, fisik, psikis, eksploitasi, TPPO (Tindak Pidana Perdagangan Orang), dan penelantaran. Data diambil berdasarkan cakupan geografis yaitu provinsi-provinsi di Indonesia, dan data yang digunakan adalah data yang relevan dan lengkap <https://siga.kemenpppa.go.id/>.

2. Preprocessing Data

Proses pengolahan awal data dimulai dari seleksi variabel utama yang akan digunakan untuk analisis clustering. Variabel yang dipilih adalah kolom 'Cakupan' sebagai identifikasi provinsi dan kolom 'Seksual' sebagai data jumlah kasus kekerasan seksual per provinsi. Tahap berikutnya adalah standarisasi data menggunakan StandardScaler untuk menormalisasi nilai-nilai dalam dataset agar memiliki skala yang sama, sehingga algoritma clustering dapat bekerja dengan optimal.

3. Pembangunan Model Clustering

Pada tahapan ini, peneliti menggunakan algoritma K-Means clustering dengan parameter jumlah cluster sebanyak 3 ($k=3$) dan random state 42 untuk memastikan konsistensi hasil. Tujuan dari clustering ini adalah agar model mampu mengelompokkan provinsi-provinsi berdasarkan tingkat risiko kekerasan seksual ke dalam tiga kategori, yaitu kategori Aman (risiko rendah), Sedang (risiko menengah), dan Rawan (risiko tinggi).

4. Analisis Ranking Presentase

Setelah clustering dilakukan, tahap selanjutnya adalah melakukan analisis ranking provinsi berdasarkan jumlah kasus kekerasan seksual dan menghitung persentase kontribusi setiap provinsi terhadap total kasus nasional. Analisis ini mencakup identifikasi provinsi dengan kategori Top 3 (tiga provinsi dengan kasus tertinggi), Bottom 3 (tiga provinsi dengan kasus terendah), dan Middle 3 (tiga provinsi dengan kasus rata-rata).

3.3 Analisis Sentimen

Analisis Sentimen, atau sering disebut Opinion Mining, adalah studi yang mengkaji opini, sikap, dan emosi individu terhadap suatu subjek, seperti isu kekerasan seksual dalam konteks penelitian ini. Proses ini melibatkan pengidentifikasian pendapat seseorang tentang sesuatu dan kemudian mengambil keputusan berdasarkan pendapat tersebut. Dalam penelitian analisis sentimen komentar publik terkait isu kekerasan seksual ini, algoritma SVM akan digunakan untuk menganalisis data.

Adapun beberapa tahapan analisis data tersebut meliputi:

$$\text{Presentase}_i = \left(\frac{\text{Jumlah Data pada Kelas}_i}{\text{Jumlah Total Data}} \right) \times 100$$

1. Pengambilan Data

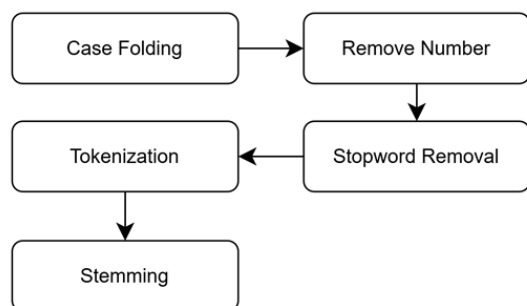
Dalam penelitian ini, data komentar publik terkait isu kekerasan seksual dengan cara scrapping data komentar TikTok. Data yang diperoleh mencakup kolom 'Comment Text' yang berisi teks komentar asli, kolom 'Sentimen' yang berisi label sentimen (Positif, Negatif, Netral), dan kolom 'Clean Text' yang berisi teks yang telah dibersihkan. Data yang digunakan adalah komentar yang relevan dengan topik kekerasan seksual.

2. Pelabelan

Data komentar yang telah terkumpul sudah dikategorikan menjadi tiga kategori yaitu positif, negatif, dan netral. Pelabelan ini dilakukan berdasarkan analisis konten komentar terhadap isu kekerasan seksual, dimana komentar yang mendukung upaya pencegahan dan penanganan kekerasan seksual dikategorikan sebagai positif, komentar yang mengandung unsur negatif atau tidak mendukung dikategorikan sebagai negatif, sedangkan komentar yang bersifat informatif tanpa opini tertentu dikategorikan sebagai netral.

3. Data Preprocessing

Tahap praproses ini dilakukan agar dalam pengelompokan dapat diproses dengan baik. Selain itu, preprocessing dalam penelitian ini juga bertujuan untuk menghilangkan noise pada data yang akan diproses yang dapat mempengaruhi hasil clustering.



Gambar 2. Alur Preprocessing Data

Proses pembersihan data dimulai dari melakukan tahap case folding untuk mengubah huruf kapital menjadi huruf kecil. Tahap berikutnya adalah penghapusan angka (remove numbers) untuk menghilangkan karakter

numerik yang tidak diperlukan. Kemudian dilakukan penghapusan tanda baca (remove punctuation) untuk menghilangkan karakter khusus dan tanda baca. Tahap berikutnya adalah tokenizing untuk memisahkan teks menjadi bagian-bagian kecil berupa kata-kata individual. Selanjutnya dilakukan stopwords removal yang digunakan untuk menghilangkan kata-kata umum yang tidak diperlukan dalam analisis sentimen menggunakan stopwords bahasa Indonesia. Tahapan terakhir adalah melakukan stemming menggunakan Sastrawi Stemmer untuk mengubah kata-kata menjadi bentuk dasar.

4. Pembobotan TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode yang mengubah teks menjadi representasi numerik, menunjukkan seberapa penting sebuah kata dalam dokumen. TF menghitung seberapa sering kata itu muncul di satu dokumen, sementara IDF mengurangi bobot kata-kata yang terlalu umum dan muncul di banyak dokumen.

5. Pembangunan Model

Pada tahapan ini, peneliti menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear, yang dikenal sebagai metode machine learning yang efektif untuk klasifikasi teks. Tujuan dari pelatihan ini adalah agar model mampu mengenali pola-pola yang terdapat dalam data teks, sehingga dapat secara otomatis mengklasifikasikan sentimen komentar publik terkait isu kekerasan seksual ke dalam tiga kategori, yaitu sentimen positif, negatif, dan netral.

6. Evaluasi

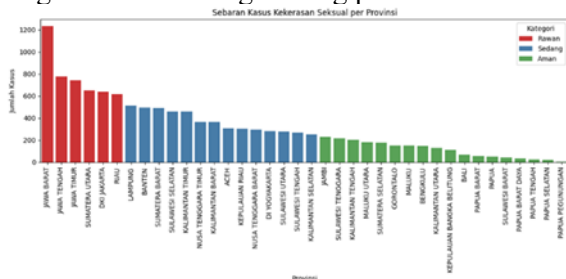
Untuk menilai kinerja model klasifikasi, peneliti akan menganalisis dan mengevaluasi efektivitas model klasifikasi SVM. Pada tahap ini, digunakan confusion matrix, sebuah format matriks yang memvisualisasikan performa model dengan membandingkan hasil prediksinya dengan data aktual. Penilaian kinerja model mencakup berbagai aspek, seperti akurasi (accuracy), daya ingat (recall), presisi (precision), dan f1-score.

4. HASIL DAN PEMBAHASAN

4.1 Distribusi Kasus kekerasan Seksual per Provinsi

Dalam analisis pertama, dilakukan clustering terhadap provinsi berdasarkan jumlah kasus

kekerasan seksual. Hasil clustering mengelompokkan provinsi ke dalam tiga kategori risiko: Aman, Sedang, dan Rawan. Visualisasi yang dihasilkan menunjukkan sebaran kasus kekerasan seksual di berbagai provinsi dengan warna yang menggambarkan tingkat risiko masing-masing provinsi.



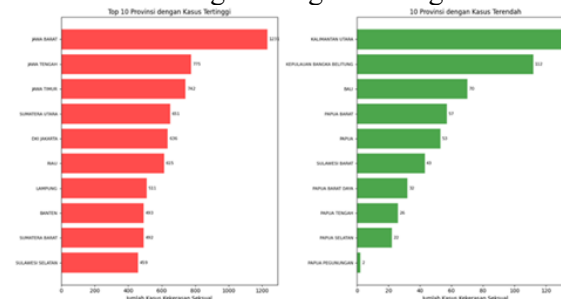
Gambar 3. Hasil VisualisasiSebarab Kasus Kekerasan Seksual Per Provinsi

Hasil visualisasi menunjukkan sebaran jumlah kasus kekerasan seksual per provinsi di Indonesia. Berdasarkan visualisasi, Provinsi Jawa Barat menempati urutan pertama dengan jumlah kasus tertinggi, diikuti Jawa Tengah, Jawa Timur, Sumatera Utara, dan DKI Jakarta. Pewarnaan batang dibedakan berdasarkan kategori tingkat risiko, yaitu kategori "Rawan" ditandai warna merah, "Sedang" warna biru, dan "Aman" warna hijau. Hasil ini menunjukkan adanya konsentrasi kasus pada wilayah-wilayah dengan kepadatan penduduk tinggi, sementara provinsi di kawasan timur Indonesia umumnya berada pada kategori aman. Visualisasi ini dapat digunakan sebagai dasar pengambilan kebijakan untuk prioritas pencegahan kekerasan seksual di daerah-daerah yang rawan.

4.2 Ranking Provinsi Berdasarkan Kasus Kekerasan Seksual

Berdasarkan data yang telah diolah, dilakukan ranking provinsi berdasarkan jumlah kasus kekerasan seksual. Provinsi dengan jumlah kasus kekerasan seksual tertinggi adalah Jawa Barat (1231 kasus atau 10,46%), Jawa Tengah (775 kasus atau 6,58%), dan Jawa Timur (742 kasus atau 6,30%), yang semuanya masuk dalam kategori rawan. Sebaliknya, tiga provinsi dengan jumlah kasus terendah adalah Papua Tengah, Papua Selatan, dan Papua Pegunungan, yang masing-masing memiliki kurang dari 30 kasus dan dikategorikan aman. Sementara itu, Sulawesi Tengah, Kalimantan Selatan, dan

Jambi memiliki jumlah kasus mendekati rata-rata nasional dengan kategori sedang dan aman.



Gambar 4. Analisis Presentase Kasus Kekerasa Seksual

Melalui analisis persentase, dihitung proporsi setiap provinsi terhadap total kasus kekerasan seksual nasional. Berikut adalah total kasus kekerasan seksual secara nasional tercatat sebanyak 11.771 kasus dengan rata-rata per provinsi sebesar 309,8 kasus. Jawa Barat menjadi provinsi dengan jumlah kasus tertinggi, yaitu 1.231 kasus, sementara Papua Tengah memiliki jumlah kasus terendah, yaitu 26 kasus. Dari segi kategori risiko, terdapat 18 provinsi dalam kategori Aman (47,4%) dengan rata-rata 111,1 kasus, 14 provinsi dalam kategori Sedang (36,8%) dengan rata-rata 365,8 kasus, dan 6 provinsi dalam kategori Rawan (15,8%) dengan rata-rata 775 kasus per provinsi. Hasil ini menunjukkan distribusi yang tidak merata dan adanya wilayah dengan tingkat kejadian yang memerlukan perhatian lebih serius.

```

=== RINGKASAN ANALISIS ===
Total kasus kekerasan seksual nasional: 11,771
Rata-rata per provinsi: 309.8
Provinsi dengan kasus tertinggi: JAWA BARAT (1231 kasus)
Provinsi dengan kasus terendah: PAPUA TENGAH (26 kasus)

Distribusi Cluster:
- Aman: 18 provinsi (47.4%), rata-rata 111.1 kasus
- Sedang: 14 provinsi (36.8%), rata-rata 365.8 kasus
- Rawan: 6 provinsi (15.8%), rata-rata 775.0 kasus

```

Gambar 5. Hasil Presentase Kasus kekerasan Seksual

4.3 Distribusi Kelas Sentimen Komentar

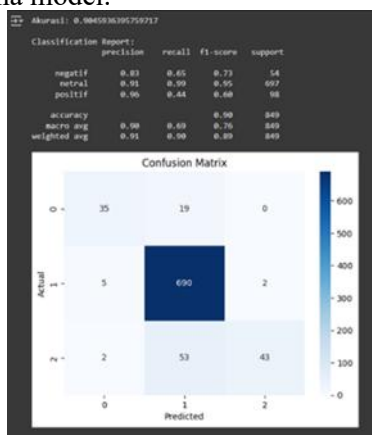
Melakukan analisis terhadap komentar yang diberikan, dilakukan dengan pengolahan teks, termasuk Case Folding, Tokenization, Stopword Removal, dan Stemming. Dari analisis ini, diperoleh distribusi kelas sentimen yang terdiri dari Positif, Negatif, dan Netral.

Tabel 1. Hasil Preprocessing Data

Teks	Case Folding	Remove Number	Tokenization	Stopword Removal	Stemming
#œkita tidak butuh feminisme lagi# lalu siapa yang akan bersuara untuk wanita di afganistan?	#œkita tidak butuh feminisme lagi# lalu siapa yang akan bersuara untuk wanita di afganistan?	#œkita tidak butuh feminisme lagi# lalu siapa yang akan bersuara untuk wanita di afganistan?	#œkita tidak butuh feminisme lagi# lalu siapa yang akan bersuara untuk wanita di afganistan?	#œkita butuh feminisme lagi# bersuara wanita afganistan	kita butuh feminisme lagi suara wanita afganistan
tau gasi kebanyakan cewe keluar malem itu yang ditakutin bukan hantu tapi cowok#	tau gasi kebanyakan cewe keluar malem itu yang ditakutin bukan hantu tapi cowok#	tau gasi kebanyakan cewe keluar malem itu yang ditakutin bukan hantu tapi cowok#	tau gasi kebanyakan cewe keluar malem itu yang ditakutin bukan hantu tapi cowok#	tau gasi kebanyakan cewe malem ditakutin hantu cowok#	tau gas banyak cewe malem ditakutin hantu cowok
baru kejadian kemaren malem, 1 gerbong kreta full laki laki cuma ada 1 perempuan yg kemudian bilang ke aku #œkak kamu di sini aja ya aku takut, soalnya cowo semua#	baru kejadian kemaren malem, 1 gerbong kreta full laki laki cuma ada 1 perempuan yg kemudian bilang ke aku #œkak kamu di sini aja ya aku takut, soalnya cowo semua#	baru kejadian kemaren malem, 1 gerbong kreta full laki laki cuma ada 1 perempuan yg kemudian bilang ke aku #œkak kamu di sini aja ya aku takut, soalnya cowo semua#	baru kejadian kemaren malem, 1 gerbong kreta full laki laki cuma ada 1 perempuan yg kemudian bilang ke aku #œkak kamu di sini aja ya aku takut, soalnya cowo semua#	kejadian kemaren malem gerbong kreta full laki laki perempuan yg bilang #œkak aja ya takut cowo semua#	jadi kemaren malem gerbong kreta full laki laki perempuan yg bilang kak aja ya takut cowo semua

4.4 Hasil Classification Report dan Confution Matrix

Menampilkan matriks yang menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas. Ini akan memberikan gambaran lebih jelas tentang kinerja model pada masing-masing kelas. Visualisasi Confusion Matrix menunjukkan perbandingan antara label yang diprediksi dan label yang sebenarnya. Setiap kotak menunjukkan jumlah data untuk kombinasi kelas yang diprediksi dan kelas yang sebenarnya. Warna yang lebih gelap menunjukkan jumlah yang lebih besar, memberikan gambaran yang jelas tentang performa model.

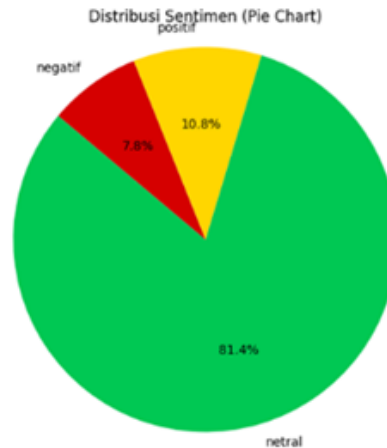


Gambar 6. Hasil Akurasi dan Confusion Matrix

4.5 Hasil Distribusi Sentimen (Pie Chart)

Pie Chart menunjukkan distribusi persentase masing-masing kelas sentimen (Positif, Negatif, dan Netral) dalam dataset. Setiap irisan dalam diagram ini mewakili proporsi data untuk setiap

kelas sentimen, dengan label yang menunjukkan nilai persentase dari total data. Netral 81.4%, Negatif 7.8%, Positif 10.8% hasil dari distribusi dari berbagai macam komentar.



Gambar 7. Hasil Distribusi Pie Chart

5. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian, dapat disimpulkan bahwa algoritma K-Means Clustering mampu mengklasifikasikan provinsi di Indonesia ke dalam tiga kategori tingkat kasus kekerasan seksual, yaitu Aman (47,4%), Sedang (36,8%), dan Rawan (15,8%), sehingga terlihat adanya pola spasial signifikan dalam distribusi kasus di tiap wilayah. Sementara itu, analisis sentimen menggunakan algoritma Support Vector Machine (SVM) dengan TF-IDF berhasil mengklasifikasikan komentar masyarakat di TikTok ke dalam tiga kelas, yakni positif (10,8%), netral (81,4%), dan negatif (7,8%) dengan akurasi 90%, yang menunjukkan efektivitas metode ini dalam memahami persepsi publik terhadap isu pelecehan seksual. Evaluasi model juga memperlihatkan kinerja precision dan recall yang baik, khususnya pada kelas negatif yang relevan dengan fokus penelitian, serta memperkuat temuan bahwa K-Means Clustering mampu mengidentifikasi wilayah dengan tingkat kerawanan tinggi. Dengan demikian, kedua pendekatan ini terbukti memberikan kontribusi penting dalam pemetaan masalah sosial serta memahami persepsi masyarakat terhadap isu kekerasan seksual.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] W. Adiyanto, "Pemanfaatan Media Sosial Instagram Sebagai Ruang Diskusi Upaya Pencegahan Pelecehan Seksual Di Lingkungan Akademis," *Jurnal Ilmiah Pangabdhi*, Vol. 6, No. 2, Pp. 78–83, Oct. 2020, Doi: 10.21107/Pangabdhi.V6i2.7594.
- [2] A. Medvi And M. Syahminan, "Strategi Komunikasi Dan Penanggulangan Pelecehan Seksual Dalam Media Sosial Tiktok," *Jurnal Sains Sosio Humaniora*, Vol. 8, No. 1, Pp. 85–97, Aug. 2024, Doi: 10.22437/Jssh.V8i1.36526.
- [3] N. Dinanti, S. Lestanti, And S. N. Budiman, "Penerapan Algoritma K-Means Untuk Pengelompokkan Tindak Kejahatan Di Wilayah Hukum Polres Blitar Kota," Oct. 2023.
- [4] A. Muhariya, "Pengelompokkan Komentar Pada Media Sosial Instagram Menggunakan Metode K-Means Clustering Untuk Identifikasi Awal Cyberbullying," 2022.
- [5] D. Nugraha And P. Astuti, "Analisis Sentimen Cyberbullying Pada Sosial Media Instagram Menggunakan Metode Support Vector Machine," *Information System For Educators And Professionals*, Vol. 8, No. 2, Pp. 152–164, Dec. 2023.
- [6] Z. Alhaq, A. Mustopa, And J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," 2021.
- [7] D. Nugraha And P. Astuti, "Analisis Sentimen Cyberbullying Pada Sosial Media Instagram Menggunakan Metode Support Vector Machine," *Information System For Educators And Professionals*, Vol. 8, No. 2, Pp. 152–164, 2023.
- [8] N. H. Alfajr, G. Garno, And D. Yusup, "Studi Komparasi Algoritma Random Forest Classifier Dan Support Vector Machine Dalam Prediksi Penyakit Jantung," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 13, No. 3, Jul. 2025, Doi: 10.23960/Jitet.V13i3.6569.
- [9] I. P. Darmawan, G. Pradnyana, And I. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *Science And Information Technology*, Vol. 6, 2023, [Online]. Available: <https://doi.org/10.31598/https://doi.org/10.31598>
- [10] M. Rahman And M. K. F. Fili, "Analisis Hasil Rekapitulasi Pilkada Daerah Khusus Jakarta (Dkj) 2024 Menggunakan Metode Support Vector Machine," 2025.
- [11] T. A. S. Tampubolon And Devega Mariza, "Analisis Sentimen Cyberbullying Pada Komentar Instagram Menggunakan Metode Klasifikasi Support Vector Machine(Svm)," 2025. [Online]. Available: <https://console.apify.com/>
- [12] F. Andriansyah And P. Astuti, "Analisis Sentimen Komentar Konten Edukatif Di Instagram Dengan Metode Naïve Bayes Dan Support Vector Machine," 2025. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/imtechno>
- [13] A. Muhariya, "Pengelompokkan Komentar Pada Media Sosial Instagram Menggunakan Metode K-Means Clustering Untuk Identifikasi Awal Cyberbullying," Nov. 2022.
- [14] N. Dinanti *Et Al.*, "Penerapan Algoritma K-Means Untuk Pengelompokkan Tindak Kejahatan Di Wilayah Hukum Polres Blitar Kota," Oct. 2023.
- [15] R. Andarsyah And A. Yanuar, "Sentimen Analisis Aplikasi Posaja Pada Google Playstore Untuk Peningkatan Pospay Superapp Menggunakan Support Vector Meachine," Apr. 2024.