

# Analisis Sentimen Publik Terkait Kekerasan Seksual di Indonesia dengan Algoritma Naïve Bayes dan SVM

Ni Made Naila Nalista\*, Ni Made Ika Marini Mandenni, I Made Agus Dwi Suarjaya

Fakultas Teknik, Teknologi Informasi, Universitas Udayana, Bali, Indonesia

Email: <sup>1</sup>nailanalista@student.unud.ac.id, <sup>2</sup>made\_ikamarini@unud.ac.id, <sup>3</sup>agussuarjaya@it.unud.ac.id

Email Penulis Korespondensi: nailanalista@student.unud.ac.id

Submitted 24-04-2025; Accepted 28-05-2025; Published 30-06-2025

## Abstrak

Semakin meningkatnya kasus kekerasan seksual yang terjadi di Indonesia, dan media sosial merupakan ruang bagi masyarakat Indonesia untuk mengekspresikan pendapat. Penelitian dilakukan dengan maksud untuk melakukan klasifikasi terhadap berbagai jenis sentimen publik yang tertuang dalam *X (Twitter)* dan juga komentar *Instagram* menerapkan dua algoritma sebagai perbandingan yakni *Naïve Bayes* dan *SVM*. Beberapa proses dilaksanakan meliputi pengumpulan data dari media sosial, *preprocessing* data, pelabelan manual, serta penerapan kedua algoritma tersebut pada dataset yang telah diproses. Sumber data dalam penelitian ini berasal dari platform *X (Twitter)* dan juga *Instagram* yang ditulis dalam Bahasa Indonesia dan berfokus pada isu kekerasan seksual yang terjadi di Indonesia. Hasil analisis sentimen digolongkan ke dalam tiga klasifikasi utama: positif, negatif, dan netral. Temuan yang dihasilkan mengindikasikan, *SVM* menghasilkan akurasi 82,17% pada pembagian data 80:20 tanpa menggunakan *GridSearch* untuk mengoptimalkan hasil, hasil akurasi *SVM* lebih unggul jika dibandingkan dengan *Naïve Bayes* yang memiliki tingkat akurasi sebesar 78,92%. Berdasarkan analisis tersebut, dapat disimpulkan *SVM* terbukti lebih optimal dalam melakukan analisis terhadap sentimen publik terkait kekerasan seksual di Indonesia dibandingkan dengan *Naïve Bayes*. Hasil analisis sentimen di media sosial terhadap kekerasan seksual di Indonesia menunjukkan bahwa mayoritas bersentimen netral, dengan dataset yang didominasi oleh informasi, laporan kasus tanpa ekspresi emosional, dan komentar di luar topik.

**Kata Kunci:** Analisis Sentimen; Instagram; Kekerasan Seksual; Naïve Bayes; Support Vector Machine (SVM); X (Twitter).

## Abstract

The increasing number of sexual violence cases in Indonesia, along with the role of social media as a space for the public to express their opinions, forms the basis for this research. The study aims to classify various types of public sentiment expressed on X (formerly Twitter) and Instagram comments by applying two algorithms for comparison: Naïve Bayes and SVM. Several processes carried out, including data collection from social media, data preprocessing, manual labeling, and the implementation of both algorithms on the processed dataset. The data sources utilized are posts written in Indonesian on X (Twitter) and Instagram, focusing on issues of sexual violence in Indonesia. The sentiment analysis results were grouped into three main categories: positive, negative, and neutral. The outcomes show that SVM achieved an accuracy of 82.17% using an 80:20 data split without applying GridSearch for optimization. The SVM results outperformed those of Naïve Bayes, which achieved an accuracy of 78.92%. This investigation leads to the conclusion that SVM is more optimal in analyzing public sentiment related to sexual violence in Indonesia compared to Naïve Bayes. The sentiment analysis results from social media regarding sexual violence in Indonesia show that the majority of sentiments are neutral, with the dataset being dominated by informative content, case reports without emotional expression, and off-topic comments.

**Keywords:** Sentiment Analysis; Instagram; Sexual Violence; Naïve Bayes; Support Vector Machine (SVM); X (Twitter).

## 1. PENDAHULUAN

Kekerasan seksual merupakan salah satu isu sosial yang terus mendapat perhatian luas dari masyarakat di Indonesia, mengingat dampaknya yang sangat kompleks dan multidimensional [1]. Informasi yang didapat lewat sistem berbasis *web* dalam mendukung perlindungan perempuan dan anak [2] tahun 2024 mencatat adanya 27.303 kasus kekerasan, dengan mayoritas korban adalah perempuan, yakni sebanyak 23.709 orang. Angka ini mencerminkan bahwa kekerasan terhadap perempuan masih menjadi permasalahan serius yang belum tertangani secara optimal. Bentuk kekerasan seksual yang terjadi sangat beragam, mulai dari pelecehan verbal, eksploitasi seksual, hingga pemerkosaan, yang semuanya meninggalkan dampak mendalam terhadap kualitas hidup korban [3]. Dalam beberapa tahun terakhir, kesadaran publik terhadap isu ini meningkat secara signifikan, seiring dengan munculnya gerakan sosial, serta keberanian para korban yang mulai berbicara secara terbuka di ruang publik. Media sosial memainkan peran vital dalam dinamika ini, tidak hanya sebagai ruang berbagi informasi dan pengalaman, tetapi juga sebagai alat untuk membentuk opini publik dan mendorong perubahan sosial.

Platform seperti *X (Twitter)* dan *Instagram* menjadi kanal yang paling umum dipakai oleh masyarakat dalam menyampaikan pandangan dan berpartisipasi dalam diskusi daring terkait kekerasan seksual. Di Indonesia, *X (Twitter)* termasuk platform jejaring sosial terpopuler dengan jumlah pengguna mencapai 78 juta dari total populasi di negara Indonesia [4]. Media sosial juga berfungsi sebagai alat pemantau sosial yang efektif, memberikan tekanan kepada instansi pemerintah, lembaga hukum, maupun pelaku kekerasan agar bertanggung jawab atas tindakan mereka [5]. Dalam konteks inilah, analisis terhadap sentimen publik di media sosial menjadi sangat bernilai, karena mampu memberikan gambaran yang jelas mengenai bagaimana isu ini dipersepsikan dan direspons oleh masyarakat dari berbagai latar belakang. Sentimen yang muncul, baik positif, negatif, maupun netral, dapat dijadikan indikator sosial yang berguna bagi penyusunan kebijakan publik, strategi kampanye edukatif, maupun upaya pemberdayaan korban kekerasan seksual. Selain itu, analisis semacam ini juga membuka peluang untuk mengidentifikasi celah komunikasi antara masyarakat dan pihak berwenang, serta mendeteksi kecenderungan narasi yang berkembang di ruang digital.

Dalam upaya untuk menganalisis sentimen publik tersebut, diperlukan metode yang andal dalam mengolah data tidak terstruktur dan berjumlah besar seperti teks media sosial. *Machine learning* menjadi solusi yang banyak digunakan dalam penelitian-penelitian terkini, terutama teknik *Naïve Bayes* serta SVM dikenal efektif melakukan pengelompokan teks dan menganalisis sentimen. Studi-studi sebelumnya menunjukkan hasil yang bervariasi. Penelitian oleh [6] mengenai sentimen terhadap *BTS* di *X (Twitter)* menunjukkan keunggulan algoritma SVM dengan akurasi 81%, lebih tinggi dibandingkan *Naïve Bayes* yang mencapai 79%. Penelitian [7] tentang sentimen terhadap kebijakan penanganan kekerasan seksual di kampus menunjukkan bahwa SVM unggul dengan akurasi 92%, jauh lebih tinggi dibandingkan *Naïve Bayes* (64%), menggunakan *TF-IDF* dan *K-Fold Cross Validation*. Penelitian [8] tentang sentimen terhadap mobil listrik menunjukkan keunggulan SVM dalam hal akurasi, sementara [9] dalam kajian perekrutan PPPK menemukan *Naïve Bayes* sebagai algoritma yang lebih akurat. Adapun studi [10] mengenai vaksinasi *Covid-19* menunjukkan bahwa meskipun SVM sangat akurat, *Naïve Bayes* menunjukkan kinerja yang lebih seimbang. Perbedaan hasil ini menunjukkan bahwa efektivitas algoritma sangat bergantung pada konteks dan karakteristik dataset yang digunakan. Dalam penelitian mengenai kekerasan seksual dan berbahasa Indonesia sebagai bahasa utama menjadi faktor pembeda utama dari studi-studi sebelumnya. Oleh karena itu, perbandingan dua algoritma ini menjadi penting untuk menilai mana yang lebih sesuai dalam klasifikasi sentimen publik terhadap kekerasan seksual di media sosial.

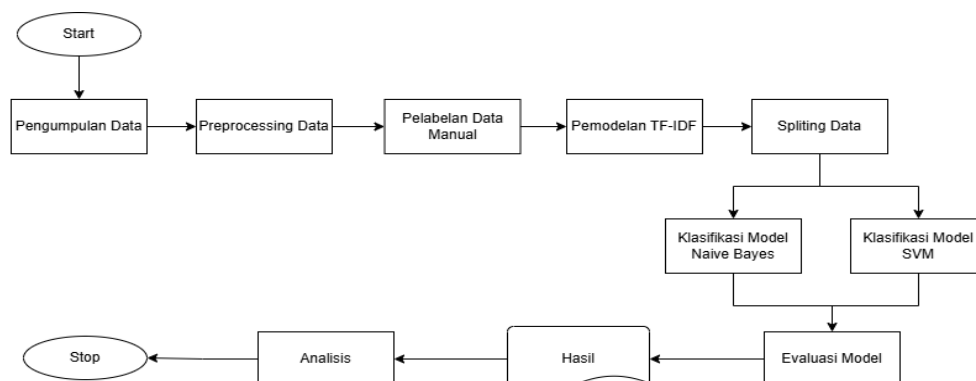
Dalam pengolahan analisis sentimen, teknik *machine learning* seperti *Naïve Bayes* dan SVM terbukti efektif untuk menangani data teks dalam jumlah besar yang tidak terstruktur. *Naïve Bayes* merupakan algoritma probabilistik yang sederhana tetapi memiliki daya kerja yang kuat untuk klasifikasi teks karena kemampuannya mengelola fitur dalam jumlah besar dengan efisien [11]. Sebaliknya, SVM dikenal karena kemampuannya dalam menemukan *hyperplane* terbaik yang memisahkan data secara optimal, sehingga sering kali menghasilkan akurasi yang tinggi dalam klasifikasi data [12]. Menggabungkan kedua algoritma dalam satu penelitian memungkinkan peneliti untuk membandingkan secara langsung kinerja tiap-tiap model terkait penanganan data sentimen yang rumit dan sensitif. Dengan kata lain, pemilihan model yang tepat bukan hanya bergantung pada akurasi, tetapi juga pada kestabilan performa, kecepatan pemrosesan, dan kemudahan dalam interpretasi hasil klasifikasi. Oleh karena itu, penelitian ini menguji dan membandingkan kedua algoritma tersebut dalam konteks yang lebih spesifik dan relevan secara sosial, yaitu sentimen publik terhadap kekerasan seksual.

Penelitian-penelitian sebelumnya memiliki keterkaitan dengan penelitian ini. Metodologi yang digunakan serupa, yaitu analisis sentimen dengan algoritma *Naïve Bayes* dan SVM berbasis data dari media sosial. Meskipun demikian, penelitian ini lebih terfokus pada isu kekerasan seksual dengan pendekatan analisis yang lebih kontekstual. Perbedaan juga terletak pada pelabelan sentimen yang digunakan, di mana dalam penelitian ini bersifat teknis dan berbasis leksikal, berdasarkan keberadaan kata-kata positif atau negatif, bukan berdasarkan sikap pro atau kontra terhadap isu tersebut karena hampir tidak ditemukan pernyataan yang secara eksplisit mendukung kekerasan seksual, sehingga pelabelan sentimen tidak didasarkan pada dukungan atau penolakan, melainkan murni pada nuansa linguistik yang terkandung dalam teks.

Tujuan utama dari penelitian adalah untuk mengeksplorasi persepsi, sikap, dan reaksi masyarakat Indonesia terhadap isu kekerasan seksual dengan pendekatan analisis sentimen berbasis *machine learning*. Dengan menggali opini yang tersebar di media sosial, studi ini diharapkan dapat memberikan dampak yang nyata dalam memahami psikologi sosial masyarakat, serta mendukung pemangku kebijakan serta lembaga yang fokus pada perlindungan perempuan dan anak untuk merancang intervensi yang lebih efektif. Temuan dari penelitian juga dapat dijadikan dasar bagi kampanye edukatif, pembentukan opini publik yang lebih sehat, serta penguatan regulasi yang berpihak pada korban kekerasan seksual.

## 2. METODOLOGI PENELITIAN

Tahap metodologi yang diterapkan di sini mencakup pengumpulan data, *preprocessing*, pelabelan data, pengujian model, dan visualisasi data.



**Gambar 1.** Alur Proses Sistem

Penelitian dilakukan dengan menggunakan salah satu bahasa *coding* populer, yaitu *Python*. Guido van Rossum mengembangkan *Python* di *Centrum Wiskunde & Informatica* (CWI) yang berlokasi di Belanda, dan paling awal diperkenalkan pada 1991 [13]. Penelitian dimulai dengan mengumpulkan data dari *X* (*Twitter*) dan *Instagram* melalui *crawling*, kemudian membersihkan data dengan *preprocessing*. Sebanyak 6.000 data yang telah dibersihkan dilabeli manual ke dalam kelompok sentimen negatif, positif, dan netral, lalu diproses dengan *TF-IDF*. Data dipecah ke dalam dua bagian, *training* dan *testing*, dan diklasifikasikan mengaplikasikan algoritma *Naive Bayes* dan *SVM*. Model diperiksa dengan menggunakan indikator kinerja seperti akurasi dan presisi, dan digunakan untuk menilai sentimen pada data tidak berlabel. Hasilnya dianalisis untuk memahami pola sentimen publik terkait kekerasan seksual dan membandingkan kinerja kedua algoritma.

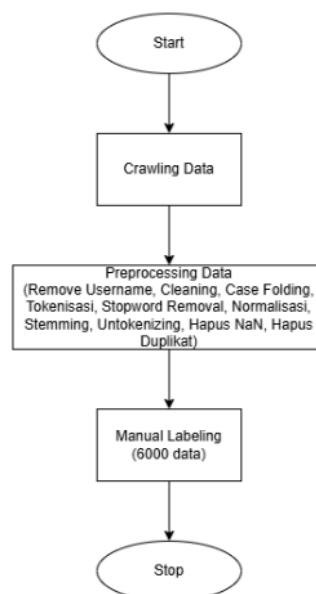
## 2.1 Pengumpulan Data

Dataset dikumpulkan melalui *crawling* di platform *X* dan *Instagram*, dengan total 1.545.184 data. Proses *crawling* data bertujuan untuk mengumpulkan dataset yang diperlukan, sehingga dapat digunakan dalam analisis lebih lanjut [14]. Pengumpulan data *X* dilakukan menggunakan library *tweet-harvest* di *Python*. Pengumpulan data menggunakan *Tweet-Harvest* dapat dikonfigurasi dengan menentukan rentang waktu pengambilan data, kata kunci yang digunakan, serta jumlah maksimum data yang dikumpulkan [15].

Data *Instagram* diperoleh melalui *Instaloader* dan *Data Miner*, dengan menggunakan *URL* postingan yang relevan dengan kekerasan seksual. *Instaloader* adalah sebuah library yang dirancang untuk memudahkan pengunduhan berbagai jenis konten dan data dari platform *Instagram*. Alat *Instaloader* memungkinkan untuk mengunduh berbagai macam media, seperti foto, video, cerita *Instagram* (*instastory*), serta informasi profil, tagar, dan lokasi [16]. Sedangkan *Data Miner* adalah sebuah alat untuk *web scraping* dan *crawling* yang memudahkan pengguna dalam mengekstrak data dari berbagai halaman *web*. Data yang diperoleh dapat disimpan dengan format *CSV* atau *Excel*. Tool *Data Miner* tersedia dalam bentuk ekstensi untuk browser *Google Chrome* dan *Microsoft Edge* [17], membuatnya mudah diakses bagi pengguna yang ingin mengotomatisasi pengambilan data dari internet.

## 2.2 Preprocessing Data

*Preprocessing* data dilangsungkan guna membersihkan dan memproses data mentah agar siap dianalisis. Tujuan utama dari *preprocessing* adalah untuk menjamin data digunakan analisis sudah dalam kondisi bersih, terorganisir, dan siap untuk diolah atau diproses lebih lanjut [18]. Proses ini mencakup beberapa langkah sistematis, seperti penghilangan data yang tidak relevan (misalnya simbol, angka, dan *stopwords*), normalisasi teks, serta penghapusan duplikat. *Preprocessing* untuk mengonversi data teks menjadi format yang lebih mudah dimengerti, proses *preprocessing* membantu menemukan hubungan antar data dan mengekstraksi fitur sehingga data siap untuk diproses dan dianalisis [19]. Langkah-langkah tersebut bertujuan untuk memastikan kualitas data yang optimal, sehingga dapat digunakan untuk pelatihan model analisis sentimen yang lebih efektif.



Gambar 2. Alur Preprocessing Data

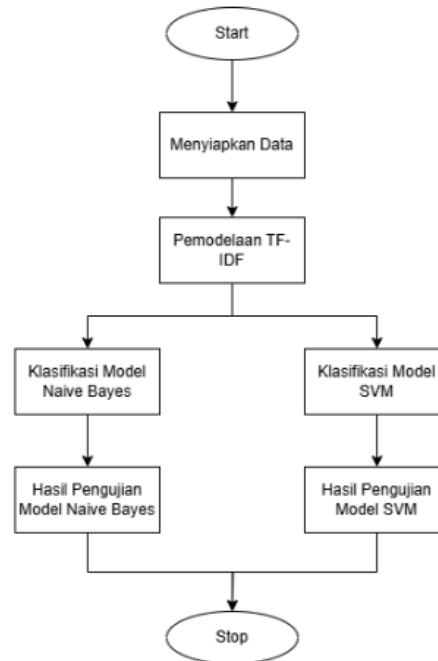
Gambar 2 merupakan alur dari proses *preprocessing*. Proses *preprocessing* dimulai dengan penghapusan informasi yang tidak sesuai, contohnya *username*, tanda baca, simbol, serta *stopwords*. Selanjutnya, dilakukan *case folding* untuk mengubah teks agar menjadi huruf kecil, tokenisasi untuk memisahkan kata-kata, dan normalisasi untuk menyamakan bentuk kata. *Stemming* digunakan untuk mengembalikan kata ke bentuk dasarnya, diikuti dengan *untokenizing* untuk menggabungkan kembali kata-kata menjadi teks utuh. Data yang mengandung nilai *NaN* atau duplikat juga dihapus.

### 2.3 Pelabelan Data

Pada tahap pelabelan, dua orang melakukan pelabelan manual terhadap 6.000 data, dibagi dalam tiga kelompok sentimen: negatif, positif, netral, masing-masing terdiri dari 2.000 data. Label ditentukan berdasarkan kata-kata negatif atau positif dalam teks, teks yang mengandung kata negatif diberi label negatif, sementara teks tanpa kata negatif diberi label positif. Teks yang tidak relevan atau berisi kalimat berita diberi label netral.

### 2.4 Pengujian Model

Proses pengujian model bertujuan untuk menilai performa dan efektivitas model dalam menjalankan tugas dengan akurasi dan keandalan yang optimal.



Gambar 3. Alur Pengujian Model

Gambar 3 menggambarkan alur analisis sentimen mempergunakan algoritma *Naive Bayes* dan SVM. Proses dimulai melalui pengumpulan dan pembersihan data, lalu dikonversi menjadi bentuk numerik dengan metode *TF-IDF*. Data tersebut kemudian diuji melalui dua jalur klasifikasi: *Naive Bayes* dan SVM, untuk perbandingan kinerja yang objektif. Setelah klasifikasi, model dievaluasi guna menghitung kinerja, mempergunakan metrik evaluasi misalnya akurasi, presisi, *recall*, serta *F1 score* dalam mendeteksi sentimen secara akurat dan andal. Berikut beberapa *metric evaluation* yang digunakan.

Akurasi adalah rasio antara jumlah prediksi yang tepat (*True Positif* dan *True Negatif*) terhadap seluruh prediksi yang dilakukan, dengan rumus sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Rumus akurasi 1 dipakai guna mengukur kinerja model klasifikasi dalam mengelompokkan data secara tepat. Nilai akurasi diukur melalui pembagian jumlah prediksi yang benar (TP dan TN) melalui total prediksi yang dilakukan (TP, TN, FP, FN), yang menunjukkan persentase data yang diklasifikasikan dengan benar. Semakin mendekati 1, semakin baik performa model. Namun, untuk penilaian yang lebih menyeluruh, khususnya pada data dengan ketidakseimbangan kelas, akurasi perlu dikaji bersama metrik lain seperti *precision*, *recall*, dan *F1-score* [20].

*Precision* merupakan metrik yang mengukur keakuratan model dalam meramalkan kategori positif, dihitung dari proporsi prediksi positif yang akurat dibanding semua prediksi yang tergolong positif [21]. Nilai presisi yang tinggi menunjukkan sedikitnya false positive yang dihasilkan oleh model.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Rumus presisi 2 digunakan untuk mengukur ketepatan prediksi positif, yaitu rasio antara *True Positive* (TP) dengan jumlah TP dan *False Positive* (FP). Nilai presisi yang tinggi menunjukkan efektivitas model dalam mengidentifikasi data positif secara tepat.

*Recall*, atau sensitivitas, mengukur kemampuan model dalam mendeteksi seluruh prediksi positif akurat, dengan membagi *True Positive* (TP) dengan total kasus positif sebenarnya (TP + FN). Nilai *recall* yang tinggi menunjukkan bahwa model efektif dalam meminimalkan *false negative*.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Rumus 3 merupakan rumus *recall* yang mengukur proporsi data positif yang berhasil dikenali model, dihitung dengan membagi *True Positive* (TP) di mana seluruh ramalan yang tergolong positif aktual (*True Positive + False Negative*).

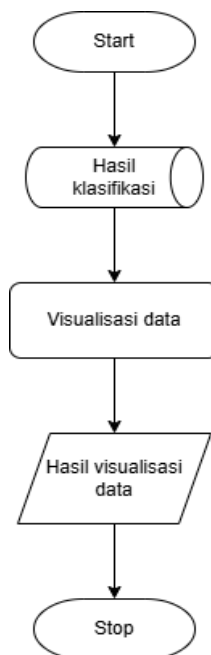
*F1-Score*, yang merupakan gabungan berdasarkan indikator presisi dan daya ingat (*recall*). *Precision* mengindikasikan keakuratan hasil positif, sedangkan *Recall* mengukur cakupan deteksi kasus positif oleh model [20].

$$F1\ Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (4)$$

Rumus *F1 Score* pada rumus 4 mengukur kinerja model klasifikasi dengan menyeimbangkan *precision* dan *recall* melalui rata-rata harmonik. Nilai ini diperoleh dari rasio antara TP dan gabungan TP, FP, serta FN. *F1 Score* sangat berguna saat data tidak seimbang atau ketika ingin meminimalkan kesalahan FN dan FP secara bersamaan.

## 2.5 Visualisasi Data

Tahap akhir penelitian adalah visualisasi data, yang bertujuan menyajikan hasil analisis klasifikasi dalam bentuk grafis agar lebih mudah dipahami secara intuitif. Alur visualisasi ditampilkan pada Gambar 4.



**Gambar 4.** Alur Visualisasi Data

Gambar 4 merupakan alur dari visualisasi data. Visualisasi data merupakan tahap akhir dalam penelitian yang bertujuan menyajikan hasil analisis secara informatif dan mudah dipahami. Melalui alat bantu visualisasi seperti Tableau, data hasil klasifikasi ditampilkan dalam bentuk grafik atau diagram yang menarik dan intuitif.

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan serta bahasan terkait proses pengumpulan dan pra-pemrosesan data, pelabelan, pengujian model, dan visualisasi data. Penelitian dilakukan melalui beberapa tahapan, dimulai dari pengumpulan data menggunakan teknik *crawling* dari media sosial *X (Twitter)* dan *Instagram* untuk mendapatkan komentar terkait kekerasan seksual. Data yang terkumpul kemudian diproses melalui tahap *preprocessing* untuk membersihkan unsur yang tidak relevan, dilanjutkan dengan pelabelan sentimen secara manual oleh dua orang terhadap 6.000 data menjadi tiga kategori: positif, negatif, dan netral. Setelah itu, data diubah menjadi representasi numerik menggunakan metode *TF-IDF*, lalu dibagi menjadi data latih dan data uji dengan pembagian data 60:40, 70:30, 80:20, dan dengan menggunakan *GridSearch*. Dua algoritma klasifikasi, yaitu *Naive Bayes* dan *SVM*, diterapkan untuk mengklasifikasikan sentimen, dan performanya dievaluasi menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Model yang telah dilatih kemudian digunakan untuk memberi label pada data tidak berlabel, yang hasilnya dianalisis guna memahami pola sentimen publik serta membandingkan performa kedua algoritma tersebut.



### 3.1 Pengumpulan Data

Pengumpulan datanya melalui platform media sosial, salah satunya adalah *X (Twitter)* dan *Instagram*. Pengumpulan data dilakukan dengan menggunakan *Python* yang dijalankan melalui *platform Google Colaboratory*. Untuk platform *X*, data diperoleh menggunakan *library tweet-harvest*. Proses *crawling Instagram* dilakukan dengan menggunakan *instaloader* dan dengan menggunakan *tools* bernama *Data Miner*.

|    | A          | B          | C          | D                                                                                     | E        | F          | G          | H     | I           | J         | K         | L         |
|----|------------|------------|------------|---------------------------------------------------------------------------------------|----------|------------|------------|-------|-------------|-----------|-----------|-----------|
| 1  | conversati | created_a  | favorite_c | full_text                                                                             | id_str   | image_url  | in_reply_t | lang  | location    | quote_cou | reply_cou | retweet_c |
| 2  | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes ya Allah..... spices                                                     | 1.67E+18 | tanyarlif  | es in      |       | 0-1.        | 0         | 0         | 0 h       |
| 3  | 1.67E+18   | Sun Jun 18 | 30         | @tanyarlifes anak bajingan                                                            | 1.67E+18 | tanyarlif  | es in      |       | not rps fre | 0         | 0         | 1 h       |
| 4  | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes setan aja minder sama kelakuan manusia                                   | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 0         | 0 h       |
| 5  | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes Badjangan                                                                | 1.67E+18 | tanyarlif  | es eu      |       |             | 0         | 0         | 0 h       |
| 6  | 1.67E+18   | Sat Jun 17 | 1          | @tanyarlifes manusia lagi berebut singgasana setan.                                   | 1.67E+18 | tanyarlif  | es in      |       | jatim.      | 0         | 0         | 1 h       |
| 7  | 1.67E+18   | Sat Jun 17 | 4          | @tanyarlifes anjirr cok horor bgt anak aman sekarang                                  | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 1         | 0 h       |
| 8  | 1.67E+18   | Sat Jun 17 | 4          | @tanyarlifes Dari dulu dah sering serem2 nder. Karena sosial media lebih bka terti    | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 0         | 0 h       |
| 9  | 1.67E+18   | Sat Jun 17 | 2          | @tanyarlifes Setan pun bingung kenapa kelakuan manusia malah melebihi setan yang      | 1.67E+18 | https://pt | tanyarlif  | es in | Bogor koti  | 0         | 0         | 0 h       |
| 10 | 1.67E+18   | Sat Jun 17 | 3          | @tanyarlifes Yang katanya kesel karena ditagih duit kas terus bunuh temennya itu?     | 1.67E+18 | tanyarlif  | es in      |       | Thailand    | 0         | 0         | 0 h       |
| 11 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes Tadi siang di berita juga ada 3 bocah SD nyiksa odgi sampe meninggal ter | 1.67E+18 | tanyarlif  | es in      |       | Hacienda    | 0         | 2         | 0 h       |
| 12 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes Gila...                                                                  | 1.67E+18 | tanyarlif  | es eu      |       | āeāā...āā.  | 0         | 0         | 0 h       |
| 13 | 1.67E+18   | Sat Jun 17 | 11         | @tanyarlifes hukuman anak dibawah umur bisa nggkak dibikin biar kapok makin lama      | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 1         | 1 h       |
| 14 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes sekarang berita mudah kesebar apa emang orang pada kek dakjal?           | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 0         | 0 h       |
| 15 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes Dajjal                                                                   | 1.67E+18 | tanyarlif  | es         |       | Taichung (  | 0         | 0         | 0 h       |
| 16 | 1.67E+18   | Sat Jun 17 | 2445       | @tanyarlifes set4n aja minder liat kelakuan manusia jaman sekarang mana diil3c3hin    | 1.67E+18 | https://pt | tanyarlif  | es in |             | 2         | 0         | 355 h     |
| 17 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes Makanya klo ada kasus keji kayak gini gak usah masih di bawah umur itu   | 1.67E+18 | tanyarlif  | es in      |       |             | 0         | 0         | 0 h       |
| 18 | 1.67E+18   | Sat Jun 17 | 0          | @tanyarlifes @atpvov makin bgt kan cwk                                                | 1.67E+18 | tanvarlif  | es in      |       | dreamies    | 0         | 1         | 0 h       |

**Gambar 5.** Hasil Pengumpulan Data Dari X (Twitter)

Gambar 5 menunjukkan hasil pengumpulan data dari *Twitter* melalui teknik *crawling* menggunakan *Python* dan *library tweet-harvest*. Data dikumpulkan dengan kata kunci terkait kekerasan seksual, seperti “pelecehan”, “perkosaan”, dan “rudapaksa”, serta URL postingan yang relevan. Sebanyak 444.195 data mentah terkumpul antara 6 September 2024 hingga 11 Desember 2024.

| A  | B                    | C                                                                                                                                                                            | D | E | F | G | H | I |
|----|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|---|---|
| 36 | aris_kurnia_hikmawan | "Spill pondoknya donk"                                                                                                                                                       |   |   |   |   |   |   |
| 37 | panritasbr           | "Payah jg nih kiyai"                                                                                                                                                         |   |   |   |   |   |   |
| 38 | slulupaly            | "harusnya nasehatin anaknya pak"                                                                                                                                             |   |   |   |   |   |   |
| 39 | menghademengcapek    | "Najis"                                                                                                                                                                      |   |   |   |   |   |   |
| 40 | untungwalluyo        | "Masalah keluarga?"                                                                                                                                                          |   |   |   |   |   |   |
| 41 | echahasbi            | "Yaa Allah jadi kepikiran ntipin anak di pesantren...ðŸ™ˆc"                                                                                                                  |   |   |   |   |   |   |
| 42 | onandio_             | "DPO ujung ujungnya lolosðŸ™ˆ,"                                                                                                                                              |   |   |   |   |   |   |
| 43 | muhammadz_17         | "Kiyai kiyaii. Hmmm"                                                                                                                                                         |   |   |   |   |   |   |
| 44 | bank_thaher          | "Kiai kenthu ...."                                                                                                                                                           |   |   |   |   |   |   |
| 45 | axldien_             | "Respect sama pak polisinya yang sabar dengerin walaupun dalam tekanan"                                                                                                      |   |   |   |   |   |   |
| 46 | luluspuji4           | "Gak salah kok lari lawack"                                                                                                                                                  |   |   |   |   |   |   |
| 47 | alter00ego           | "Keliatan banget pak polisi dengerin omongan si kiyai dengan ekspresi kesal yang meluber ðŸ™ˆE sama pak saya pun bakal gitu kalo kiyai-nya macam gitu. Belum apa2 udah bilan |   |   |   |   |   |   |
| 48 | risa_sunarya         | "Masalah kel itu klo korban & pelaku sama2 dlm ikatan keluarga. ðŸ™ˆcðŸ™ˆc"                                                                                                  |   |   |   |   |   |   |
| 49 | yudie_kons           | "Kinerja polisi dipertanyakan sih klo kek gini. Polisi itu penegak keadilan. Bukan tunduk dan patuh gitu aja"                                                                |   |   |   |   |   |   |
| 50 | airaakbar8           | "Lah gimana sih? Ikuti hukum dunk pa kiai"                                                                                                                                   |   |   |   |   |   |   |
| 51 | dimas_rio32          | "@divisihumaspolri jangan sampe negara kalah ðŸ™ˆðŸ™ˆðŸ™ˆE"                                                                                                                  |   |   |   |   |   |   |
| 52 | mona_rhania          | "Speclez"                                                                                                                                                                    |   |   |   |   |   |   |
| 53 | 1124chanx            | "Untuk kejayaan Indonesia pala lu ðŸ™ˆ. netien indo jadi heboh niih mbah. Gimana donk"                                                                                       |   |   |   |   |   |   |

**Gambar 6.** Hasil Pengumpulan Data Dari Instagram

Gambar 6 menunjukkan hasil pengumpulan data dari *Instagram* menggunakan teknik *crawling* dengan *Python*, *library Instaloader* dan dengan *tool Data Miner*. Data dikumpulkan melalui *URL* postingan terkait kekerasan seksual, terdiri dari kolom “username” dan “komentar”. Sebanyak 340.531 data diperoleh dari *Instaloader* dan 760.458 data dari *Data Miner* antara 10 November 2024 hingga 18 Februari 2025.

### 3.2 Preprocessing Data

*Preprocessing* termasuk tahap krusial dalam pembersihan dan pengubahan data supaya siap diolah/analisis, dengan tujuan meningkatkan kualitas data. Contoh hasil *preprocessing* tersaji dalam Tabel 1.

**Tabel 1.** Hasil Preprocessing Data

| Komentar                                                                                 | Tahapan Preprocessing | Hasil Preprocessing                                                                                   |
|------------------------------------------------------------------------------------------|-----------------------|-------------------------------------------------------------------------------------------------------|
| @folkshittmedia Hukum mati aja, kasihan masa depan anaknya di penuhi dengan rasa trauma☹ | Removed               | Hukum mati aja, kasihan masa depan anaknya di penuhi dengan rasa trauma☹                              |
|                                                                                          | Username              |                                                                                                       |
|                                                                                          | Cleaning              | Hukum mati aja kasihan masa depan anaknya penuhi dengan rasa trauma                                   |
|                                                                                          | Case Folding          | hukum mati aja kasihan masa depan anaknya penuhi dengan rasa trauma                                   |
|                                                                                          | Tokenisasi            | ['hukum', 'mati', 'aja', 'kasihan', 'masa', 'depan', 'anaknya', 'penuhi', 'dengan', 'rasa', 'trauma'] |

|                  |                                                                    |
|------------------|--------------------------------------------------------------------|
| Stopword Removal | ['hukum', 'mati', 'aja', 'kasihan', 'anaknya', 'penuhi', 'trauma'] |
| Normalisasi      | hukum mati saja kasihan anaknya penuhi trauma                      |
| Stemming         | ['hukum', 'mati', 'saja', 'kasihan', 'anak', 'penuhi', 'trauma']   |
| Untokenizing     | hukum mati saja kasihan anak penuh trauma                          |

Setelah dilakukan *preprocessing*, termasuk penghapusan nilai *NaN* dan duplikat, jumlah data yang tersisa adalah 922.973. Penghapusan duplikat dilakukan dengan mengurutkan kata dalam teks secara alfabetis terlebih dahulu, guna menghindari redundansi yang dapat memengaruhi hasil analisis. Hal ini sesuai dengan pendekatan *TF-IDF* yang digunakan, di mana bobot kata dihitung berdasarkan frekuensi kemunculannya tanpa mempertimbangkan urutan kata dalam kalimat.

### 3.3 Labeling Manual

Labeling manual dilakukan oleh dua orang untuk menentukan kategori sentimen secara akurat. Kategori sentimen didasarkan pada identifikasi kata-kata positif atau negatif guna menetapkan apakah teks sifatnya positif, negatif, ataukah netral.

**Tabel 2.** Hasil Labeling Manual

| Komentar                                                                                                                                                      | Sentimen |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| alhamdulillah sesuai harap netijen                                                                                                                            | Positif  |
| binatang kalah brutal anjing biadab banget laku                                                                                                               | Negatif  |
| plis sampai takut banget jogging daerah jalan dekat rumah dilecehin orang kenal lari pagi teman pas pulang bokong pegang ber laki motor pegang spontan teriak | Netral   |
| kasihan periksa psikologis                                                                                                                                    | Positif  |
| bangsat banget manusia manusia sekarang                                                                                                                       | Negatif  |
| breaking news laku begal payudara resah sragen tangkap bocah ingus                                                                                            | Netral   |

Tabel 2 menampilkan contoh hasil pelabelan manual terhadap 6.000 data hasil praproses, masing-masing berjumlah 2.000 berlabel positif, negatif, netral. Label ditentukan berdasarkan keberadaan kata negatif, sementara komentar tidak relevan atau berupa berita diberi label netral.

### 3.4 Pengujian Model

Hasil pengujian model menunjukkan performa kedua algoritma dalam mengklasifikasikan sentimen publik terhadap kekerasan seksual. Berikut merupakan tabel perbandingan perolehan pengujian model *Naïve Bayes* serta SVM dalam mengkaji opini masyarakat mengenai kekerasan seksual di media sosial pada Tabel 3.

**Tabel 3.** Rasio Perolehan Pengujian Model Naive Bayes dan SVM

| Train:Test | Naïve Bayes |           |        |          | Support Vector Machine |            |            |            |
|------------|-------------|-----------|--------|----------|------------------------|------------|------------|------------|
|            | Accuracy    | Precision | Recall | F1 Score | Accuracy               | Precision  | Recall     | F1 Score   |
| 60:40      | 78,7%       | 79%       | 78%    | 78%      | 80,42%                 | 82%        | 80%        | 81%        |
| 70:30      | 76,6%       | 77%       | 76%    | 76%      | 80,28%                 | 81%        | 80%        | 80%        |
| 80:20      | 76,25%      | 76%       | 76%    | 76%      | <b>82,17%</b>          | <b>83%</b> | <b>82%</b> | <b>82%</b> |
| GridSearch | 78,92%      | 79%       | 79%    | 78%      | 81,5%                  | 82%        | 81%        | 81%        |

Tabel 3 menunjukkan bahwa SVM unggul dibandingkan *Naïve Bayes* dalam semua skenario pembagian data (60:40, 70:30, 80:20) serta saat menggunakan *GridSearch*. Akurasi tertinggi dicapai oleh SVM pada pembagian 80:20 sebesar 82,17%, mengungguli *Naïve Bayes* yang hanya 76,25%. Di samping hal tersebut, SVM juga mencatat *precision*, *recall*, serta *F1 score* yang lebih baik, menandakan performa klasifikasi yang lebih optimal.

|    | untokenized                                                                                                                                       | label   | Predicted_Label_NB |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------------------|
| 1  |                                                                                                                                                   |         |                    |
| 2  | pedofil bangsat                                                                                                                                   | Negatif | Negatif            |
| 3  | anjing musnah orang sel                                                                                                                           | Negatif | Negatif            |
| 4  | jera mui dorong laku keras seksual hukum maksimal                                                                                                 | Netral  | Netral             |
| 5  | astaga ketua laku keras seksual anak umur bocah laki laki korban                                                                                  | Netral  | Netral             |
| 6  | nauzubillah moga allah jauh putra putri buruk moga allah senantiasa lindung anak anak jalan amin                                                  | Positif | Positif            |
| 7  | bangsat ayah jahanam                                                                                                                              | Negatif | Negatif            |
| 8  | astaga bejat                                                                                                                                      | Negatif | Negatif            |
| 9  | innalillahi innailaihi roji kasihan anak moga anak sehat kuat iman amin allah                                                                     | Positif | Positif            |
| 10 | manusia akhir zaman                                                                                                                               | Positif | Negatif            |
| 11 | pria difabel sangka duga keras seksual bas perempuan nusa tenggara barat ntb wayan agus suatama agus resmi tahan jaksa negeri kejar mataram Kamis | Netral  | Netral             |
| 12 | kontol bosok penjara                                                                                                                              | Negatif | Negatif            |
| 13 | moga keluarga hindar jadi                                                                                                                         | Positif | Positif            |
| 14 | dapat laku lecah seksual pk1 salah staf kantor sampai berani speak takut                                                                          | Netral  | Netral             |
| 15 | astaghfirullah yaallahh                                                                                                                           | Positif | Positif            |
| 16 | kebiri saja tunggu hukum tahi anjing                                                                                                              | Negatif | Negatif            |

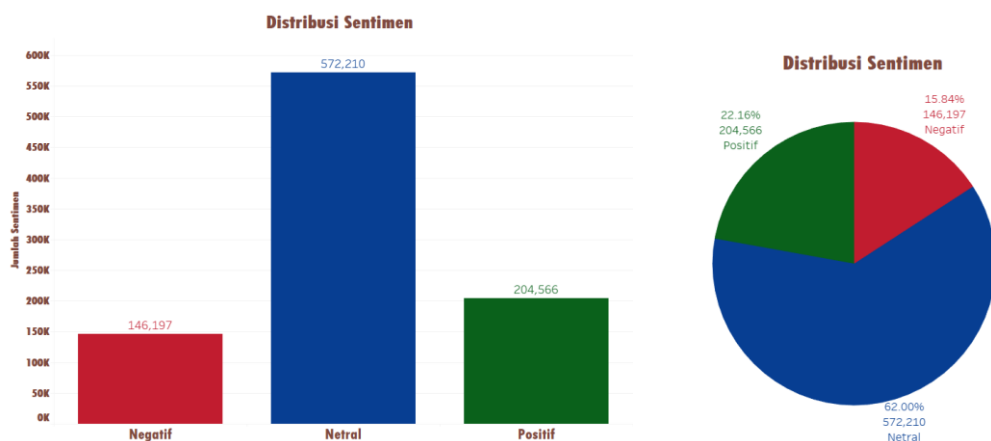
| 1  | untokenized                                                                                                                                       | label   | Predicted_Label_SVM |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------|---------|---------------------|
| 2  | pedofil bangsat                                                                                                                                   | Negatif | Negatif             |
| 3  | anjing musnah orang sel                                                                                                                           | Negatif | Negatif             |
| 4  | jera mui dorong laku keras seksual hukum maksimal                                                                                                 | Netral  | Netral              |
| 5  | astaga ketua laku keras seksual anak umur bocah laki laki korban                                                                                  | Netral  | Netral              |
| 6  | nauzubillah moga allah jauh putra putri buruk moga allah senantiasa lindung anak anak jalan amin                                                  | Positif | Positif             |
| 7  | bangsat ayah jahanam                                                                                                                              | Negatif | Negatif             |
| 8  | astaga bejat                                                                                                                                      | Negatif | Negatif             |
| 9  | innalillahi innailaihi roji kasihan anak moga anak sehat kuat iman amin allah                                                                     | Positif | Positif             |
| 10 | manusia akhir zaman                                                                                                                               | Positif | Netral              |
| 11 | pria difabel sangka duga keras seksual bas perempuan nusa tenggara barat ntb wayan agus suartama agus resmi tahan jaksa negeri kejar mataram kams | Netral  | Netral              |
| 12 | kontol bosok penjara                                                                                                                              | Negatif | Negatif             |
| 13 | moga keluarga hindar jadi                                                                                                                         | Positif | Positif             |
| 14 | dapat laku lecah seksual pkl salah staf kantor sampai berani speak takut                                                                          | Netral  | Netral              |
| 15 | astaghfirullah yaallahh                                                                                                                           | Positif | Positif             |
| 16 | kebiri saja tunggu hukum tahi anjing                                                                                                              | Negatif | Negatif             |

**Gambar 7.** Evaluasi Hasil Prediksi Dengan Data Baru

Gambar 7 merupakan hasil evaluasi terhadap 15 data baru pada yang menunjukkan perbandingan antara pelabelan manual dan label hasil prediksi pada model SVM dan *Naive Bayes* yang mencapai tingkat akurasi tertinggi. Model SVM menunjukkan kinerja 14 dari 15 data yang terklasifikasi dengan benar, sementara *Naive Bayes* juga sama, berhasil mengklasifikasikan 14 data dengan benar dan 1 salah. Model SVM maupun *Naive Bayes* sama-sama mengalami kesalahan dalam 1 dari 15 data yang diuji. Hal ini sejalan dengan hasil akurasi dari kedua model yang memang tidak jauh berbeda, walaupun SVM mengindikasikan akurasi yang sedikit lebih baik jika disandingkan *Naive Bayes*.

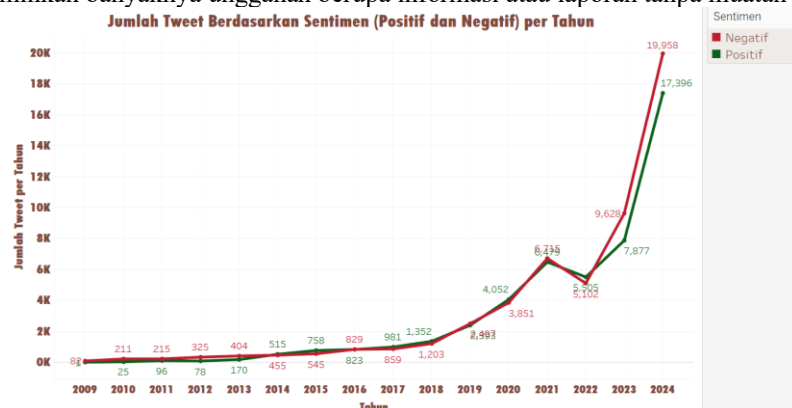
### 3.5 Visualisasi Data

Visualisasi data digunakan untuk memahami sentimen masyarakat terhadap isu kekerasan seksual di media sosial. Visualisasi dibuat berdasarkan hasil klasifikasi sentimen yang dihasilkan oleh metode SVM, karena model mencapai akurasi lebih baik dibandingkan model *Naive Bayes*.



**Gambar 8.** Distribusi Sentimen Pada Seluruh Data

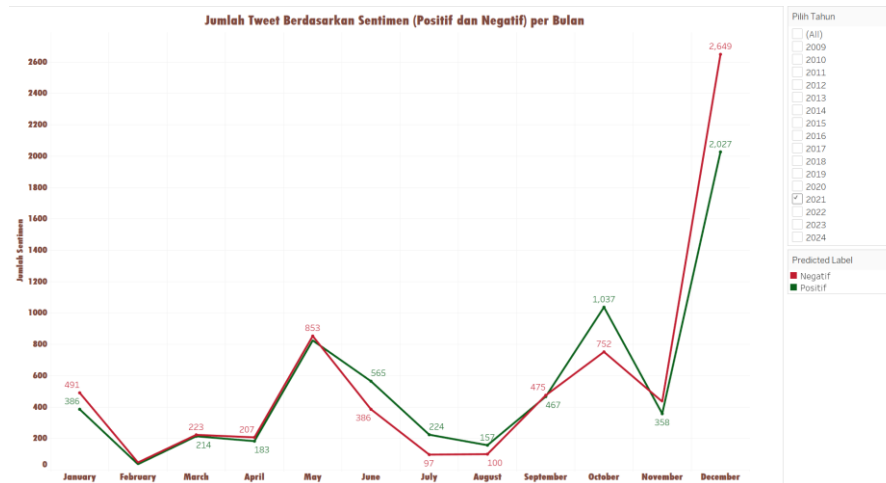
Gambar 8 menampilkan visualisasi distribusi sentimen pada seluruh data hasil klasifikasi. Grafik batang di sebelah kiri menunjukkan jumlah masing-masing kategori sentimen, dengan sentimen netral mendominasi sebanyak 572.210 data, diikuti oleh sentimen positif sejumlah 204.566 data, serta negatif sejumlah 146.197 data. Sementara itu, diagram lingkaran di sebelah kanan memperlihatkan persentase dari masing-masing sentimen: 62,00% netral, 22,16% positif, dan 15,84% negatif. Visualisasi ini menunjukkan bahwa mayoritas respons publik di media sosial terhadap isu kekerasan seksual bersifat netral, mencerminkan banyaknya unggahan berupa informasi atau laporan tanpa muatan emosi yang jelas.



**Gambar 9.** Tren Sentimen Positif & Negatif per Tahun Pada Data X (Twitter)

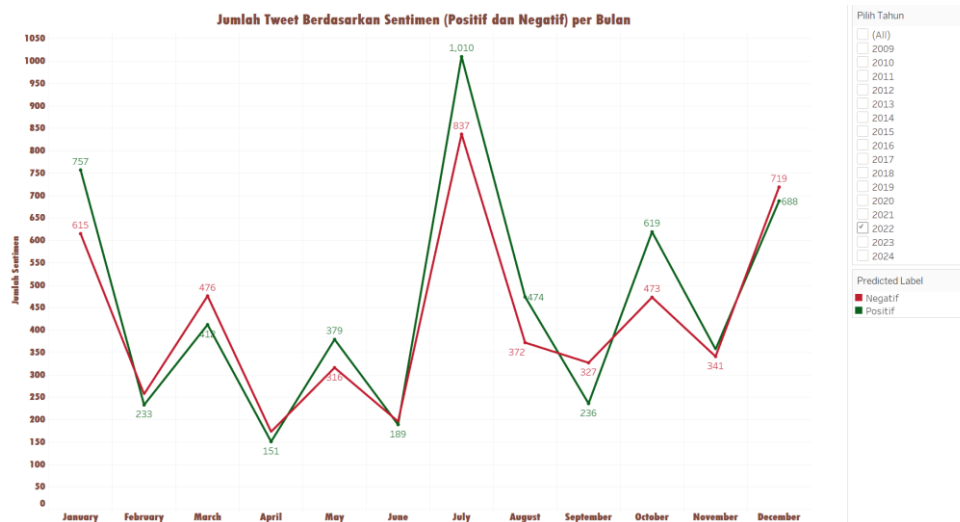


Grafik pada Gambar 9 menunjukkan tren sentimen positif dan negatif dari tahun ke tahun. Grafik yang ditampilkan hanya merepresentasikan data yang diperoleh dari *platform X (Twitter)*, karena data dari *Instagram* tidak menyertakan informasi tahun atau tanggal. Oleh karena itu, analisis tren sentimen positif dan negatif per tahun dalam grafik sepenuhnya berdasarkan data dari *platform X (Twitter)*. Dari hasil analisis, terlihat bahwa jumlah sentimen positif dan negatif mengalami peningkatan yang signifikan setelah tahun 2018, dengan lonjakan drastis pada tahun 2023 dan 2024. Meskipun sentimen negatif lebih tinggi pada tahun-tahun tertentu, tren keseluruhan menunjukkan bahwa jumlah sentimen positif juga meningkat secara signifikan.



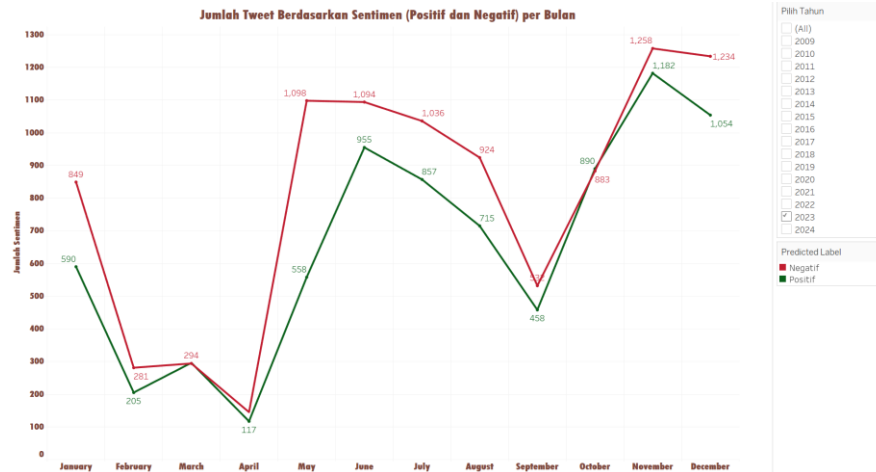
**Gambar 10.** Tren Sentimen Positif dan Negatif per Bulan Tahun 2021

Grafik pada Gambar 10 menunjukkan tren sentimen positif dan negatif per bulan tahun 2021. Grafik yang ditampilkan hanya merepresentasikan data yang diperoleh dari *platform X (Twitter)*, karena data dari *Instagram* tidak menyertakan informasi tahun atau tanggal. Tahun 2021 menunjukkan fluktuasi signifikan dalam jumlah *tweet* bersentimen positif dan negatif setiap bulannya. Bulan Desember menjadi puncak aktivitas tertinggi dengan lonjakan drastis, terutama pada sentimen negatif yang mencapai 2.649 *tweet*, sementara sentimen positif tercatat sebanyak 2.027 *tweet*. Periode Mei juga mencatat jumlah tinggi untuk kedua sentimen, menandakan adanya isu penting pada bulan tersebut. Jumlah *tweet* paling rendah terlihat pada bulan Juli, mengindikasikan menurunnya intensitas diskusi di media sosial.



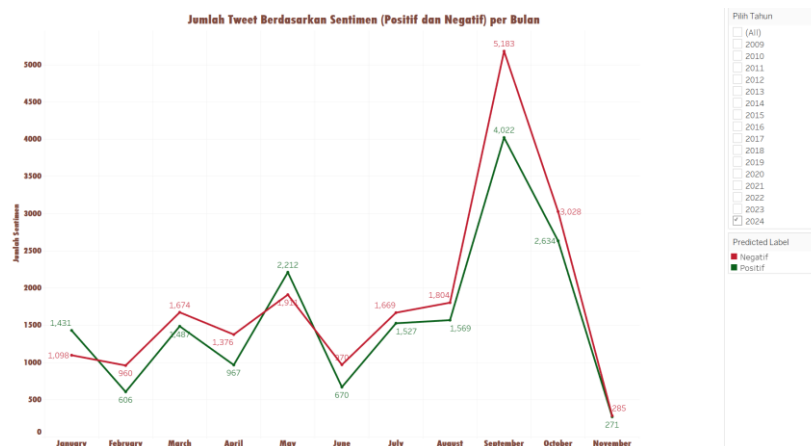
**Gambar 11.** Tren Sentimen Positif dan Negatif per Bulan Tahun 2022

Grafik pada Gambar 11 menunjukkan tren sentimen positif dan negatif per bulan tahun 2022. Distribusi sentimen pada tahun 2022 terlihat lebih seimbang dan tidak terlalu ekstrem dibandingkan tahun sebelumnya. Jumlah *tweet* tertinggi tercatat pada bulan Juli dengan 1.010 *tweet* positif dan 837 *tweet* negatif, menunjukkan meningkatnya perhatian publik pada isu tertentu. Bulan April menjadi titik terendah dalam aktivitas *tweet* untuk kedua jenis sentimen. Dominasi sentimen positif terjadi di beberapa bulan, mencerminkan suasana publik yang relatif lebih optimis.



**Gambar 12.** Tren Sentimen Positif dan Negatif per Bulan Tahun 2023

Grafik pada Gambar 12 menunjukkan tren sentimen positif dan negatif per bulan tahun 2023. Jumlah *tweet* dengan dominasi sentimen negatif yang lebih besar dibandingkan sentimen positif pada sebagian besar bulan. Lonjakan tertinggi tercatat pada bulan November dengan 1.258 *tweet* negatif dan 1.182 *tweet* positif, mencerminkan adanya isu besar yang ramai diperbincangkan. Bulan April mencatat aktivitas *tweet* paling rendah, baik untuk sentimen positif maupun negatif. Sentimen negatif terus mendominasi hingga akhir tahun.



**Gambar 13.** Tren Sentimen Positif dan Negatif per Bulan Tahun 2024

Grafik pada Gambar 13 menunjukkan tren sentimen positif dan negatif per bulan tahun 2024. Jumlah *tweet* meningkat tajam pada pertengahan hingga akhir tahun. Puncak aktivitas terjadi pada bulan September, dengan 5.183 *tweet* negatif dan 4.022 *tweet* positif, mencerminkan reaksi besar terhadap suatu isu atau peristiwa penting. Tren ini terus berlanjut hingga Oktober, sebelum menurun drastis di bulan November. Dominasi sentimen negatif terlihat lebih kuat sepanjang tahun, meskipun beberapa bulan seperti Mei menunjukkan kecenderungan positif yang lebih tinggi.

Jumlah data positif: 284566

Jumlah data negatif: 146197



**Gambar 14.** Wordcloud Sentimen Positif dan Negatif Seluruh Data

Gambar 14 merupakan *wordcloud* yang menampilkan hasil analisis sentimen berdasarkan kata yang banyak muncul pada seluruh dataset. Terdapat dua kategori sentimen yang ditampilkan, yaitu positif dan negatif. Kategori sentimen positif menggunakan warna biru. Kata-kata yang dominan dalam kategori ini adalah “orang tua”, “polisi”, “korban”, “anak”, “sudah”, dll. Menunjukkan bahwa komentar dengan sentimen positif cenderung menyoroti upaya perlindungan dan dukungan terhadap korban kekerasan seksual, termasuk peran aparat hukum dan keluarga dalam menangani kasus tersebut. Kategori sentimen negatif menggunakan warna merah. Kata-kata yang paling menonjol dalam kategori ini adalah “anjing”, “laku”, “hukum mati”, “gila”, dan “goblok”. Menunjukkan ekspresi emosi yang kuat, terutama dalam menuntut hukuman berat bagi pelaku. Dominasi kata-kata tersebut mengindikasikan adanya emosi kuat seperti kemarahan atau ketidakpuasan dalam pembahasan.

Jumlah data positif: 2232  
Jumlah data negatif: 1911



**Gambar 15.** Wordcloud Bulan Mei 2024 Data X (Twitter)

Gambar 15 menunjukkan *wordcloud* berdasarkan data dari platform X (Twitter) pada bulan Mei 2024, yang mencatat peningkatan signifikan dalam *tweet* dengan sentimen positif. Peningkatan ini mencerminkan empati dan dukungan publik terhadap korban, terutama anak-anak, dengan kata-kata seperti “anak”, “korban”, “percaya”, “tolong”, dan “orang tua”. Sentimen negatif didominasi oleh kata-kata seperti “laku”, “perkosa”, “anjing”, dan “penjara”, mencerminkan kemarahan terhadap pelaku dan desakan hukuman. Kenaikan sentimen positif ini menandakan respons lebih suportif dari masyarakat terhadap isu kekerasan seksual yang sedang dibahas. Contoh kekerasan seksual yang terjadi di masyarakat adalah peristiwa kekerasan yang menimpa seorang siswi SMP di Lampung pada Mei 2024 yang mengejutkan publik. Korban yang masih berada di tingkat SMP menjadi korban kekerasan seksual oleh sepuluh orang pelaku. Peristiwa ini mendapat kepedulian besar dari Adde Rosi Khoerunnisa, merupakan perwakilan Komisi III di Dewan Perwakilan Rakyat Republik Indonesia, yang mendorong aparat pihak berwenang agar segera menangkap pelaku yang sempat buron [22].

Jumlah data positif: 379  
Jumlah data negatif: 316



**Gambar 16.** Wordcloud Bulan Mei 2022 Data X (Twitter)

Gambar 16 menampilkan *wordcloud* sentimen positif dan negatif pada Mei 2022. Sentimen positif menunjukkan peningkatan, ditandai dengan kata-kata seperti “korban”, “anak”, “keluarga”, dan “adil”, mencerminkan empati dan dukungan publik. Sentimen negatif seperti “pelaku”, “perkosa”, dan “hukum”, yang menggambarkan kemarahan terhadap kasus kekerasan seksual. Peningkatan pembahasan positif mengindikasikan adanya dorongan solidaritas dan perhatian publik yang lebih dalam terhadap korban apalagi anak, sekaligus tetap menyuarakan kemarahan terhadap pelaku dalam narasi yang emosional. Salah satu peristiwa kekerasan seksual yang dilaporkan pada saat itu adalah pelecehan seksual terhadap anak disabilitas di Tamansari, Jakarta Barat, oleh pelaku berinisial D alias Bobby. Pelaku melakukan perbuatan asusila terhadap tetangganya yang berkebutuhan khusus. Setelah melihat pelaku, korban merasa ketakutan dan melaporkan kejadian tersebut kepada orangtuanya, yang segera melakukan visum dan melaporkan kejadian ke polisi. Unit PPA Polres Jakarta Barat berkolaborasi bersama Pusat Terpadu Pelayanan Perempuan dan Anak DKI Jakarta agar mendampingi korban. Pelaku diancam pidana penjara 5 sampai 15 tahun [23].





#### 4. KESIMPULAN

## REFERENCES

- Copyright © 2025 The Ni Made Naila Nalista, Page 183  
This Journal is licensed under a Creative Commons Attribution 4.0 International License

- [5] Kompasiana, "Media Sosial sebagai Wadah dalam Upaya Meningkatkan Kesadaran akan Isu Lingkungan," *www.kompasiana.com*, 2024. <https://www.kompasiana.com/nafraalyssasyahbaniah2727/66f6ce2f34777c5c4a0fea32/media-sosial-sebagai-wadah-dalam-upaya-meningkatkan-kesadaran-akan-isu-lingkungan>
- [6] Rina Noviana and Isram Rasal, "Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter," *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [7] M. Derryl Qinand, A. Nilogiri, and T. Timur, "Sentimen Pada Komentar Youtube Tentang Pencegahan Dan Penanganan Kekerasan Seksual Pada Permendikbud Berbasis Naive Bayes Dan Support Vector Machine," vol. 7, no. 2, pp. 114–121, 2022, [Online]. Available: <http://jurnal.unmuhjembar.ac.id/index.php/JUSTINDO>
- [8] W. Ningsih, B. Alfianda, R. Rahmadden, and D. Wulandari, "Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [9] F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.
- [10] D. Alita and R. A. Shodiqin, "Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.58602/jaiti.v1i1.20.
- [11] A. Pamungkas, "Memahami Algoritma Naive Bayes: Konsep dan Penerapan," *pemrogramanmatlab.com*, 2024. <https://pemrogramanmatlab.com/2023/08/09/memahami-algoritma-naive-bayes-konsep-dan-penerapan/>
- [12] IBM, "Apa itu support vector machines (SVM)?," *www.ibm.com*, 2023. <https://www.ibm.com/id-id/topics/support-vector-machine#:~:text=Support vector machines %28SVM%29 adalah algoritma machine learning,memaksimalkan jarak antara setiap kelas dalam ruang N-dimensi.>
- [13] N. N. Pandika Pinata, I. M. Sukarsa, and N. K. Dwi Rusjyanthi, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 3, p. 188, 2020, doi: 10.24843/jim.2020.v08.i03.p04.
- [14] N. S. Marga, A. R. Isnain, and D. Alita, "Terhadap Kasus Corona Menggunakan Metode Naive Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, pp. 453–463, 2020.
- [15] P. A. Setiawati, I. M. A. D. Suarjaya, and I. N. P. Trisna, "Sentiment Analysis of Unemployment in Indonesia During and Post COVID-19 on X (Twitter) Using Naive Bayes and Support Vector Machine," *J. Inf. Syst. Informatics*, vol. 6, no. 2, pp. 662–675, 2024, doi: 10.51519/journalisi.v6i2.713.
- [16] M. F. Zulfikar *et al.*, "Sistem Monitoring Keadaan Darurat Berdasarkan Caption Instagram Menggunakan Naive Bayes Dengan Jarak Levenshtein," vol. 8, pp. 1–12, 2025.
- [17] D. Miner, "Data Miner is the most powerful web scraping tool for professional data miners," *dataminer.io*, 2021. <https://dataminer.io/>
- [18] S. K. A. Firdausi, "Data Preprocessing Adalah: Pengertian, Manfaat, & Tahapannya," *dibimbing.id*, 2024. <https://dibimbing.id/blog/detail/data-preprocessing-adalah-pengertian-manfaat-tahapannya> (accessed Sep. 10, 2024).
- [19] G. A. M. Kukuh Jaluwana, Gusti Made Arya Sasmita, and I Made Agus Dwi Suarjaya, "Analysis of Public Sentiment Towards Government Efforts to Break the Chain of Covid-19 Transmission in Indonesia Using CNN and Bidirectional LSTM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 511–520, 2022, doi: 10.29207/resti.v6i4.4055.
- [20] W. I. Maulana, "Mengurai Metrics: Apa Itu Presisi, Recall, dan Bagaimana Menginterpretasikannya.," *medium.com*, 2023. [https://medium.com/himit-pens/mengurai-metrics-apa-itu-presisi-recall-dan-bagaimana-menginterpretasikannya-a15e7f90411e#:~:text=Presisi atau biasa disebut Precision adalah metrik yang,positif yang benar dengan jumlah total prediksi positif \(accessed Jan. 18, 2024\).](https://medium.com/himit-pens/mengurai-metrics-apa-itu-presisi-recall-dan-bagaimana-menginterpretasikannya-a15e7f90411e#:~:text=Presisi atau biasa disebut Precision adalah metrik yang,positif yang benar dengan jumlah total prediksi positif (accessed Jan. 18, 2024).)
- [21] D. N. Avianty, P. I. G. P. S. Wijaya, and F. Bimantoro, "The Comparison of SVM and ANN Classifier for COVID-19 Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 2, p. 128, 2022, doi: 10.24843/lkjiti.2022.v13.i02.p06.
- [22] Em. D. RI, "Siswi SMP di Lampung Alami Kekerasan Seksual, Adde Rosi: Tangkap Tiga Pelaku Buron!," *emedia.dpr.go.id*, 2024. <https://emedia.dpr.go.id/2024/05/05/siswi-smp-di-lampung-alami-kekerasan-seksual-adde-rosi-tangkap-tiga-pelaku-buron/> (accessed Apr. 16, 2025).
- [23] KOMPASTV, "Pelaku Pelecehan Seksual Terhadap Anak Disabilitas di Tamansari Terancam 15 Tahun Penjara!," *www.kompas.tv*, 2022. <https://www.kompas.tv/amp/video/290575/pelaku-pelecehan-seksual-terhadap-anak-disabilitas-di-tamansari-terancam-15-tahun-penjara> (accessed Apr. 17, 2025).
- [24] KumparanNEWS, "Cabuli dan Aniaya Balita, Pria di Bali Divonis Penjara 13 Tahun dan Denda Rp 5 M," *kumparan.com*, 2022. [https://kumparan.com/kumparannews/1zItoLg4tta/full?utm\\_source=Mobile&utm\\_medium=tw&shareID=xOtPwmTDY3j](https://kumparan.com/kumparannews/1zItoLg4tta/full?utm_source=Mobile&utm_medium=tw&shareID=xOtPwmTDY3j) (accessed Apr. 17, 2025).