

Analisis Sentimen Ujaran Kebencian Pada Kolom Komentar Di Instagram

Evip Athirah Salsabila¹, Mellysa Pratama², Putri Wahyuni³, Vina Purwaningrum⁴,
Mufid Faruq Aziz⁵

Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu

Email: ¹evipathirahsalsabila@gmail.com, ²Pratamamelisa831@gmail.com,
³ptri wahyuni123@gmail.com, ⁴vinapurwanigrum60@gmail.com,
⁵Mufidfaruq.a@gmail.com

ABSTRAK

Instagram merupakan salah satu media sosial paling populer di Indonesia yang digunakan untuk berbagai keperluan, mulai dari berbagi informasi hingga mengekspresikan pendapat. Fitur kolom komentar di Instagram memungkinkan pengguna untuk berinteraksi secara langsung, namun kebebasan ini sering kali disalahgunakan untuk menyebarkan ujaran kebencian yang dapat memicu konflik sosial serta berdampak pada kesehatan mental. Penelitian ini bertujuan untuk menganalisis sentimen pada komentar Instagram yang mengandung ujaran kebencian dengan pendekatan Natural Language Processing (NLP) menggunakan algoritma Support Vector Machine (SVM). Data dikumpulkan dari kolom komentar publik akun Instagram yang memiliki interaksi tinggi. Hasil penelitian menunjukkan dominasi komentar negatif, sedangkan komentar positif dan netral memiliki proporsi lebih sedikit. Diharapkan hasil penelitian ini dapat menjadi masukan untuk pengembangan kebijakan moderasi konten dan meningkatkan literasi digital agar tercipta lingkungan media sosial yang lebih sehat.

Kata kunci: Instagram, analisis sentimen, ujaran kebencian, NLP, SVM

ABSTRACT

Instagram is one of the most popular social media platforms in Indonesia, used for various purposes ranging from sharing information to expressing opinions. The comment feature on Instagram allows users to interact directly; however, this freedom is often misused to spread hate speech, which can trigger social conflicts and negatively impact mental health. This study aims to analyze the sentiment of Instagram comments

containing hate speech using a Natural Language Processing (NLP) approach with the Support Vector Machine (SVM) algorithm. Data were collected from public comments on high-interaction Instagram accounts. The results show a dominance of negative comments, while positive and neutral comments appear in smaller proportions. It is hoped that the findings of this study can provide input for the development of content moderation policies and improve digital literacy to create a healthier social media environment.

Keywords: Instagram, sentiment analysis, hate speech, NLP, SVM

PENDAHULUAN

Teknologi pada saat ini tidak terlepas dari kehidupan manusia sebagai alat komunikasi yang mudah dan cepat banyak akses yang digunakan untuk bersosialisasi. Seiring bertambahnya waktu cara berkomunikasi tidak hanya menggunakan sms dan telepon, melainkan juga menggunakan internet. Banyak aplikasi yang ditawarkan melalui internet, satu diantaranya adalah akun jejaring sosial. Jejaring sosial merupakan sebuah situs berbasis pelayanan yang memungkinkan untuk membuat profil, melihat daftar pengguna yang tersedia, serta mengundang atau menerima teman untuk bergabung dalam situs tersebut (Firmansyah, 2010). Tampilan dasar situs jejaring sosial ini menampilkan halaman profil pengguna, yang di dalamnya terdiri dari identitas diri dan foto pengguna. Jejaring sosial tersebut seperti facebook, twitter dan instagram.

Instagram merupakan sebuah aplikasi dari smartphone yang khusus untuk media sosial yang merupakan salah satu dari media digital yang mempunyai fungsi hampir sama dengan twitter, namun perbedaannya terletak pada pengambilan foto dalam bentuk atau tempat untuk berbagi informasi terhadap penggunanya. Instagram juga dapat memberikan inspirasi bagi penggunanya dan juga dapat meningkatkan kreativitas, karena instagram mempunyai fitur yang dapat membuat foto lebih indah, lebih artistik dan menjadi lebih bagus (Atmoko, 2012:10).

Instagram diciptakan oleh Kevin Systrom dan Mike Krieger, dua sarjana dari Stanford University di Amerika Serikat. Mereka berdua meluncurkan instagram pada bulan Oktober 2010. Layanan instagram yang tadinya masih berupa aplikasi smartphone ini mendapatkan popularitas yang tinggi dalam waktu cepat, dengan lebih dari 100 juta pengguna yang terdaftar (dan sekitar 90 juta pengguna aktif bulanan) per Januari 2013. Ini berarti hanya dalam kurun waktu 3 tahun saja, jumlah pengguna instagram sudah

mencapai ratusan juta. (Kevin Systrom dan Mike Krieger, “Sejarah Instagram” (Online)).

Seiring kemajuan teknologi dan informasi, Perkembangan internet telah membawa perubahan besar termasuk dalam komunikasi dan pertukaran informasi. Jumlah pengguna internet terus meningkat setiap tahun, terutama di Indonesia, mencapai sekitar 77% dari total jumlah penduduk per Januari 2023[1]. Salah satu fenomena yang muncul seiring dengan perkembangan internet adalah penggunaan media sosial yang semakin meluas. Media sosial seperti Facebook, Twitter, dan Instagram telah menjadi platform populer bagi pengguna internet untuk berinteraksi, berbagi konten, dan membangun jejaring sosial[2]. Penggunaan media sosial telah mengubah cara kita berkomunikasi, berbagi informasi, dan membentuk hubungan sosial dalam kehidupan sehari-hari. Berdasarkan laporan terkini dari We Are Social dan Hootsuite, jumlah pengguna sosial media secara global mencapai 4,76 miliar. Yaitu setara dengan 59,4% dari keseluruhan penduduk dunia saat ini. Laporan tersebut juga mencatat bahwa di Indonesia, jumlah penggunanya mencapai 60,4% dari total populasi negara pada bulan yang sama (databoks, Januari 2023).

Dalam konteks ini, Instagram menjadi salah satu sosial media yang sangat populer bagi pengguna internet terutama di kalangan remaja dan orang dewasa. Instagram menawarkan beragam fitur menarik seperti unggahan foto dan video, fitur cerita (stories), penggunaan hashtag untuk mencari dan mengikuti topik-topik tertentu[3]. Data dari databoks menyebutkan bahwa India menjadi negara dengan jumlah pengguna terbanyak, dengan 326 juta pengguna, diikuti Amerika Serikat dengan 168 juta pengguna, Brasil dengan 132 juta pengguna, dan Indonesia dengan 106 juta pengguna (databoks, April 2023). Namun, di balik popularitasnya terdapat masalah, Instagram juga memiliki dampak buruk, fitur komentar memungkinkan pengguna memberikan tanggapan pada suatu postingan. Meskipun sebagian besar komentar di Instagram dapat berisi dukungan, apresiasi atau kritik sayangnya platform ini juga memberi ruang bagi komentar negatif, ejekan atau bisa disebut juga cyberbullying[4].

Cyberbullying adalah tindakan yang merujuk pada penggunaan teknologi digital untuk sengaja menyakiti, merendahkan, atau membuli orang lain secara daring[5]. Bentuk-bentuk cyberbullying sangat beragam, mencakup pengiriman pesan kebencian, pelecehan verbal, penyebaran informasi pribadi yang tidak sah, dan ancaman fisik atau psikologis[6]. Masalah yang terjadi ketika ada banyak pengguna yang tidak setuju atau tidak menyukai suatu postingan akan bersatu dan memberikan komentar negatif untuk

menyerang pemilik postingan, tidak jarang pula terjadi perdebatan karena perbedaan pendapat pada kolom komentar bahkan menggunakan kata-kata negatif dan hinaan. Fenomena ini sering kali menimbulkan dampak buruk pada korban, berupa gangguan kesehatan mental seperti kehilangan rasa percaya diri, stres, dan depresi, dalam beberapa kasus yang tragis, cyberbullying bahkan dapat menyebabkan berakhirnya kehidupan seseorang[7].

Penelitian sebelumnya oleh Talita Cintiana Adisti et al. (2025) menganalisis ujaran kebencian pada Twitter dengan pendekatan analisis sentimen menggunakan algoritma machine learning, sedangkan penelitian Desti Lestari et al. (2023) menganalisis ujaran kebencian pada Instagram dari sudut pandang linguistik forensik. Berdasarkan penelitian-penelitian tersebut, penelitian ini memadukan analisis sentimen berbasis NLP dengan kajian linguistik untuk memberikan gambaran yang lebih komprehensif.

METODE PENELITIAN

Data Penelitian

Data yang digunakan dalam penelitian ini berupa komentar publik yang dikumpulkan dari beberapa akun Instagram populer yang memiliki tingkat interaksi tinggi, seperti akun berita, selebritas, dan tokoh publik. Pengumpulan data dilakukan dengan metode web scraping selama periode Juni 2024 hingga Januari 2025.

Jumlah total komentar yang dikumpulkan sebanyak 500 komentar, dengan rincian:

- a. Komentar berpotensi ujaran kebencian: 320
- b. Komentar netral: 100
- c. Komentar positif: 80

Setiap komentar diberi label secara manual sebagai data latih, terdiri dari kategori positif, negatif, dan netral.

Teknik Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik scraping dengan memperhatikan etika penelitian. Hanya komentar yang bersifat publik dan dapat diakses secara umum yang digunakan dalam penelitian ini.

Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk membersihkan data sebelum analisis, meliputi:

1. Case folding: mengubah semua huruf menjadi huruf kecil.
2. Tokenizing: memecah teks menjadi token kata.
3. Stopword removal: menghapus kata-kata umum yang tidak memiliki makna penting.
4. Stemming: mengembalikan kata ke bentuk dasarnya.

Klasifikasi Sentimen

Komentar yang telah diproses selanjutnya diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Algoritma Support Vector Machine (SVM) digunakan karena terbukti memiliki kinerja yang baik dalam klasifikasi teks.

Validasi dan Evaluasi

Model dievaluasi menggunakan cross-validation serta metrik akurasi, precision, recall, dan F1-score untuk memastikan keandalan model dalam mendekripsi ujaran kebencian.

Implementasi Kode Python

Berikut potongan kode Python yang digunakan dalam penelitian ini.

1. Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import string
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
```

2. Preprocessing Data

```
nltk.download('stopwords')
stop_words = set(stopwords.words('indonesian'))
stemmer = PorterStemmer()

def clean_text(text):
    text = text.lower()
    text = re.sub(r'\d+', '', text)
    text = text.translate(str.maketrans('', '', string.punctuation))
    text = text.strip()
    tokens = text.split()
    tokens = [word for word in tokens if word not in stop_words]
    tokens = [stemmer.stem(word) for word in tokens]
    return ' '.join(tokens)

df = pd.read_csv('komentar_instagram.csv')
df['cleaned_comment'] = df['comment'].apply(clean_text)
```

3. Ekstraksi Fitur dan Split Data

```
tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df['cleaned_comment']).toarray()
y = df['label']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

4. Training dan Prediksi

```
svm = SVC(kernel='linear')
svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)
```

5. Evaluasi

```
print("Akurasi:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

6. Visualisasi Word Cloud

```
from wordcloud import WordCloud
all_words = ' '.join([text for text in df['cleaned_comment']])
wordcloud = WordCloud(width=800, height=500, max_font_size=110, collocations=False).generate(all_words)

plt.figure(figsize=(10, 7))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

HASIL DAN PEMBAHASAN

Berdasarkan hasil analisis terhadap 500 komentar yang dikumpulkan dari kolom komentar Instagram publik, diperoleh hasil bahwa komentar dengan sentimen negatif

memiliki dominasi tertinggi. Sebanyak 65% komentar termasuk dalam kategori negatif, diikuti oleh komentar netral sebanyak 20%, dan komentar positif sebanyak 15%.

Hasil ini menunjukkan bahwa ruang komentar di Instagram cenderung dipenuhi oleh ekspresi negatif yang berpotensi mengarah pada ujaran kebencian. Komentar-komentar tersebut sering kali mencerminkan ketidakpuasan, kemarahan, serta kecenderungan untuk merendahkan atau menyerang pihak tertentu secara langsung maupun tidak langsung.

Algoritma Support Vector Machine (SVM) digunakan untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen: positif, negatif, dan netral. Model ini berhasil mencapai tingkat akurasi sebesar 85%. Nilai precision dan recall yang tinggi, khususnya pada kelas komentar negatif, mengindikasikan bahwa model cukup efektif dalam mengenali komentar dengan kecenderungan ujaran kebencian.

Selain pengujian kuantitatif, penelitian ini juga memanfaatkan visualisasi word cloud sebagai pendukung analisis. Word cloud menampilkan kata-kata yang paling sering muncul dalam komentar yang dikategorikan negatif. Beberapa kata atau frasa yang dominan di antaranya adalah "**bodoh**", "**ga punya otak**", "**dasar idiot**", "**mental miskin**", dan "**gendut mati aja**". Kata-kata ini sering digunakan oleh netizen sebagai bentuk penghinaan atau serangan verbal terhadap individu maupun kelompok.

Untuk memperkuat hasil dari analisis berbasis machine learning, dilakukan pula pengodean manual secara linguistik. Pengodean ini bertujuan untuk mengidentifikasi jenis ujaran kebencian berdasarkan bentuk bahasa dan konteks penggunaannya. Adapun klasifikasi pengodean yang digunakan dalam penelitian ini mencakup:

Kode	Keterangan
UK	Ujaran Kebencian
PH	Penghinaan
PNB	Pencemaran Nama Baik
PNT	Penistaan
MBB	Menyebarluaskan Berita Bohong
MHST	Menghasut

Sementara untuk bentuk ujaran yang diidentifikasi secara struktural, digunakan pengkodean:

Kode	Keterangan
Ka	Kata
F	Frasa
Kl	Klausa
Kt	Kalimat

Berdasarkan analisis linguistik, berikut beberapa kata yang sering muncul dalam komentar kebencian:

No	Kata atau Frasa	Kategori
1	Bodoh	Penghinaan
2	Ga punya otak	Penghinaan
3	Dasar idiot	Penghinaan
4	Gendut mati aja	Serangan fisik
5	Mental miskin	Serangan personal

Berikut beberapa contoh pengodean komentar yang ditemukan:

- UK1 (15/06/2025) - PH - "Bodoh banget adminnya, ga punya otak!" (Ka, F)
UK2 (16/06/2025) - PH - "Gendut mati aja lo!" (Kt)
UK3 (17/06/2025) - PNB - "Dasar kampus ini sampah semua!" (Kt)
UK4 (17/06/2025) - MHST - "Ayo kita ramai-ramai report akunnya!" (Kt)

Dari pengodean tersebut, terlihat bahwa sebagian besar komentar bermuatan negatif termasuk dalam kategori **penghinaan**, sementara sebagian kecil lainnya termasuk **menghasut** dan **pencemaran nama baik**. Komentar-komentar tersebut umumnya disampaikan dalam bentuk kalimat utuh, namun ada juga yang berupa frasa atau bahkan satu kata saja, tetapi maknanya tetap menyinggung atau menyerang.

Secara keseluruhan, hasil ini memperlihatkan bahwa ujaran kebencian masih menjadi bagian yang cukup besar dalam interaksi pengguna di media sosial, khususnya di Instagram. Oleh karena itu, dibutuhkan kolaborasi antara pengguna, pihak platform, dan pemerintah untuk memperkuat literasi digital serta membangun ekosistem komunikasi yang sehat dan saling menghargai di ruang maya.

KESIMPULAN

Penelitian ini berhasil menganalisis sentimen komentar pada kolom komentar Instagram dengan pendekatan NLP dan algoritma SVM. Hasil menunjukkan dominasi komentar negatif yang mengandung ujaran kebencian, mengindikasikan masih rendahnya kesadaran pengguna akan etika berkomunikasi di media sosial. Model SVM menunjukkan kinerja yang baik dengan akurasi mencapai 85%. Diharapkan penelitian ini dapat menjadi acuan dalam pengembangan kebijakan moderasi konten, edukasi literasi digital, serta teknologi deteksi otomatis guna menekan penyebaran ujaran kebencian.

DAFTAR PUSTAKA

- We Are Social & Hootsuite. (2023). Digital 2023: Indonesia.
- Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE).
- Talita Cintiana Adisti, et al. (2025). Analisis Sentimen Ujaran Kebencian pada Tweet di Twitter. JATI: Jurnal Mahasiswa Teknik Informatika.
- Desti Lestari, et al. (2023). Ujaran Kebencian Netizen Pada Kolom Komentar di Instagram BEM KBM Untirta Tahun 2022 (Kajian Linguistik Forensik). NUSRA: Jurnal Penelitian dan Ilmu Pendidikan.
- Firmansyah, Aditya. 2010. "Situs Jejaring Sosial Menggunakan Elgg" Makalah tidak diterbitkan. Sekolah Teknik Elektro dan Informatika. ITB. Bandung.
- Atmoko Dwi, Bambang. 2012. Instagram Handbook Tips Fotografi Ponsel. Jakarta: Media kita.
- R. Gunawan et al., "ADIKSI MEDIA SOSIAL DAN GADGET BAGI PENGGUNA INTERNET DI INDONESIA," Jurnal Techno-Socio Ekonomika, vol. 14, no. 1, 2020.
- P. Studi Ilmu Komunikasi, "TREN PENGGUNAAN MEDIA SOSIAL SELAMA PANDEMI DI INDONESIA," 2020.
- A. Rachmayanti and Y. Candrasari, "PERILAKU CYBERBULLYING DI INSTAGRAM."
- M. Jubaidi and N. Fadilla, "PENGARUH FENOMENA CYBERBULLYING SEBAGAI CYBER-CRIME DI INSTAGRAM DAN DAMPAK NEGATIFNYA," Shaut Al-Maktabah : Jurnal Perpustakaan, Arsip dan Dokumentasi, vol. 12, no. 2, pp. 117–134, Dec. 2020, doi: 10.37108/shaut.v12i2.327.
- F. Syifa Mutma, "DESKRIPSI PEMAHAMAN CYBERBULLYING DI MEDIA SOSIAL PADA MAHASISWA," Jurnal Common |, vol. 4, no. 1, 2020, doi: 10.34010/common.
- J. Penelitian et al., "Agustus Tahun 2021 | Hal," vol. 1, no. 8, pp. 257–263, 2021, doi: 10.56393/decive.v1i8.298.
- T. Sutini et al., "DIRI PADA REMAJA : LITERATUR REVIEW."