

ANALISIS SENTIMEN KOMENTAR CYBERBULLYING DI MEDIA SOSIAL INSTAGRAM MENGUNAKAN PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN NAIVE BAYES

Rismayanti¹, Muhammad Zulfadhilah², Nurhaeni³, Evi Lestari Pratiwi⁴
Teknologi Informasi, Universitas Sari Mulia, Banjarmasin

email: *rismayhantie@gmail.com¹, zulfadhilah@unism.ac.id², nurhaeni030@gmail.com³,
evi.pratiwi@poliban.ac.id⁴

ABSTRAK

Instagram merupakan salah satu platform media sosial yang sangat populer di Indonesia. Walaupun media sosial memudahkan interaksi dan berbagi informasi, ada sisi negatif yang tidak bisa diabaikan, yaitu cyberbullying. Cyberbullying dapat merugikan kesehatan mental seseorang, memperburuk kondisi depresi dan kecemasan, serta menimbulkan perasaan takut dan marah pada korban. Penelitian ini bertujuan untuk mengevaluasi sentimen komentar tentang cyberbullying di Instagram dengan membandingkan metode Support Vector Machine (SVM) dan Naive Bayes dalam mengklasifikasikan sentimen menjadi positif, negatif, atau netral. Metode SVM yang digunakan mencapai akurasi 68%, sementara metode Naive Bayes memperoleh akurasi 79%. Hasil ini mengindikasikan bahwa kedua metode tersebut dapat mengklasifikasikan sentimen pada data uji dengan efektif. Selain itu, proporsi sentimen menunjukkan bahwa 31,2% komentar berlabel positif, 60,1% berlabel netral, dan hanya sekitar 8,8% berlabel negatif. Meskipun demikian, perhatian tetap diperlukan terhadap komentar-komentar negatif karena dapat memiliki dampak yang signifikan, terutama jika berkaitan dengan masalah seperti cyberbullying.

Kata kunci

analisis sentimen, cyberbullying, instagram, naive bayes, support vector machine

ABSTRACT

Instagram is one of the most popular social media platforms in Indonesia. Although social media facilitates interaction and information sharing, there is an undeniable negative aspect: cyberbullying. Cyberbullying can harm an individual's mental health, worsen conditions like depression and anxiety, and cause feelings of fear and anger in victims. This study aims to evaluate the sentiment of comments about cyberbullying on Instagram by comparing the Support Vector Machine (SVM) and Naive Bayes methods for classifying sentiment as positive, negative, or neutral. The SVM method achieved an accuracy of 68%, while the Naive Bayes method reached an accuracy of 79%. These results indicate that both methods can effectively classify sentiment in test data. Additionally, the sentiment distribution shows that 31.2% of comments are labeled as positive, 60.1% as neutral, and only about 8.8% as negative. Nevertheless, attention is still needed for negative comments as they can have significant impacts, especially when related to issues such as cyberbullying.

Keywords

sentiment analysis, cyberbullying, instagram, naive bayes, support vector machine

1. PENDAHULUAN

Media sosial adalah platform digital yang memungkinkan pengguna untuk berbagi informasi, berkomunikasi, dan berinteraksi secara online. Pada tahun 2023, laporan *We Are Social* mencatat bahwa Indonesia memiliki 167 juta pengguna aktif media sosial, dengan 153 juta di antaranya berusia di atas 18 tahun, mencakup 79,5% dari populasi. Namun, terjadi penurunan 12,57% dari 191 juta pengguna aktif pada tahun sebelumnya (Widi, 2023).

Salah satu platform media sosial yang sedang populer pada tahun 2020 adalah aplikasi Instagram. Instagram merupakan suatu platform media sosial yang fokus pada berbagi foto dan video, yang dimiliki oleh perusahaan Amerika, Meta Platform. Aplikasi ini memungkinkan pengguna untuk mengunggah media yang dapat diubah dengan bantuan filter atau disusun dengan menggunakan tagar dan penandaan geografis (Sari & Basit, 2020). Meskipun media sosial memudahkan interaksi dan berbagi informasi, platform ini juga memiliki sisi gelap, seperti cyberbullying. Berdasarkan penelitian dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), hampir setengah dari 5.900 responden melaporkan telah mengalami perundungan di dunia maya. Sementara itu, 47,2% responden mengatakan belum pernah mengalaminya, dan 2,8% lainnya tidak memberikan jawaban (Dewi, 2023).

Cyberbullying adalah jenis perundungan atau pelecehan yang terjadi secara online melalui teknologi digital, seperti media sosial, dengan tujuan merendahkan, menghina, mengancam, atau mengejek seseorang secara virtual (Yi & Zubiaga, 2023). Beberapa tindakan cyberbullying seperti menyebarkan kebohongan atau mengunggah foto atau video memalukan seseorang dan mengomentari postingan orang di media sosial, juga berupa mengirimkan pesan, gambar, atau video yang menyakitkan, kasar, atau mengancam melalui platform media sosial (Fadli, 2023). Cyberbullying dapat berdampak serius pada kesehatan mental seseorang, meningkatkan risiko depresi dan kecemasan, serta menimbulkan rasa takut dan marah pada korban. Berdasarkan hal-hal tersebut, maka penting untuk melakukan analisis sentimen pada komentar di akun media sosial seperti Instagram public figure dengan tujuan untuk mengidentifikasi apakah ada kecenderungan untuk terjadinya cyberbullying pada komentar yang diberikan.

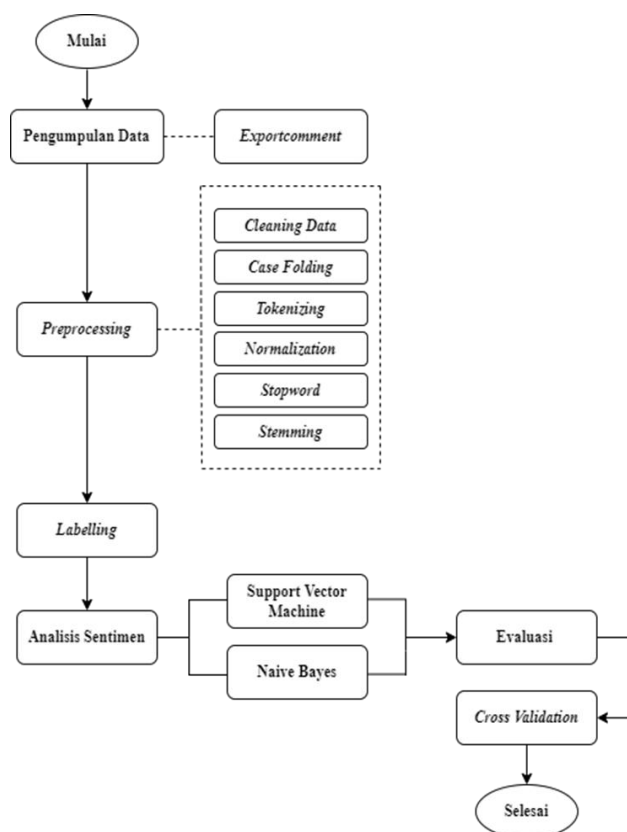
Analisis sentimen atau dikenal juga sebagai Opinion Mining, adalah bidang studi yang bertujuan untuk menganalisis opini, sentimen, penilaian, evaluasi, sikap, dan perasaan masyarakat terhadap suatu entitas seperti produk, layanan, masalah, organisasi, peristiwa, atau fitur tertentu. Dalam melakukan analisis sentimen diperlukan metode yang menunjang proses klasifikasi, seperti metode Support Vector Machine (SVM) dan Naive Bayes (Imam & Santoso, 2023).

SVM adalah metode machine learning dalam supervised learning yang digunakan untuk klasifikasi dan regresi. SVM memiliki konsep yang lebih matang dan jelas secara matematis dibandingkan teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linier maupun nonlinier (Asana & Yanti, 2023). Sedangkan algoritma Naive Bayes adalah metode klasifikasi yang menggunakan pendekatan probabilitas dan statistik. Algoritma ini berfungsi untuk mengklasifikasikan data ke dalam kategori atau kelas tertentu berdasarkan fitur-fitur atau atribut yang terdapat pada data (Apriyani & Kurniati, 2020).

2. METODE PENELITIAN

Lokasi penelitian dilaksanakan di Banjarmasin secara daring melalui media sosial Instagram. Sasaran penelitiannya yaitu pengguna media sosial Instagram, khususnya yang

berkomentar pada postingan public figure di Instagram. Penelitian fokus pada pengembangan dan pemahaman tentang opini, sentimen, dan perilaku yang terkandung dalam komentar-komentar yang ada pada media sosial Instagram, untuk mengklasifikasikan apakah komentar tersebut merupakan bentuk cyberbullying atau tidak. Jenis penelitian yang dilakukan adalah penelitian eksperimental. Penelitian eksperimental adalah suatu bentuk penelitian yang bertujuan untuk mengidentifikasi dampak atau hubungan antara satu variabel dengan variabel lainnya dalam suatu konteks yang telah ditentukan. Alur rancangan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1: Rancangan Penelitian

Data yang digunakan adalah data sekunder yang di dikumpulkan melalui komentar netizen pada konten public figure di media sosial Instagram. teknik pengumpulan data dilakukan dengan cara mengumpulkan 10 konten public figure di Instagram yang terkait dengan topik cyberbullying. Dataset yang digunakan adalah data komentar Instagram yang dikumpulkan dengan cara crawling data menggunakan web Export Comments yang berfungsi untuk mengeksport atau menyimpan komentar dari suatu platform, salahsatunya platform media sosial Instagram yang dapat disimpan ke dalam format CSV (Comma-Separated Values), Excel, atau file teks.

Jumlah data komentar yang digunakan seban yak 1000 komentar (data awal). Pelabelan setiap data dilakukan secara otomatis menggunakan Textblob, label yang diberikan ada 3 kelas, yaitu kelas positif, negatif, dan netral. Kelas negatif berarti komentar yang mengandung elemen cyberbullying, kelas positif berarti komentar mengandung unsur motivasi atau dukungan, dan kelas netral adalah komentar yang tidak mengandung elemen cyberbullying.

Analisis data dimulai dengan pengumpulan komentar dari platform Instagram menggunakan alat web Export Comments. Setelah proses crawling, diperoleh 1088

komentar dari 10 akun publik figure. Data yang diperoleh berupa data mentah yang perlu dibersihkan terlebih dahulu. Proses pembersihan data meliputi beberapa tahap text preprocessing, yaitu cleaning data, case folding, tokenizing, normalization, stopword removal, dan stemming. Setelah tahap preprocessing, data akan diberi label secara otomatis sebelum diterapkan pada model algoritma SVM dan Naive Bayes untuk analisis sentimen. Setelah model dilatih, tahap evaluasi dilakukan untuk mengukur kinerja model dalam mengklasifikasikan sentimen dengan menentukan akurasi, presisi, recall, dan F1-score. Selanjutnya, dilakukan cross-validation untuk menilai performa metode SVM dan Naive Bayes dengan membagi data secara acak dan mengelompokkan data berdasarkan nilai K k-fold. Cross-validation dilakukan sebanyak K yang ditentukan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Dataset yang digunakan terdiri dari komentar-komentar netizen pada postingan Instagram dari public figure, yang dikumpulkan melalui proses crawling menggunakan alat web Export Comments. Didapat sebanyak 1088 komentar netizen dari 10 akun Instagram. Hasil crawling data yang telah dikumpulkan dapat dilihat pada Tabel 1.

Tabel 1. Hasil Crawling Data

| No | Komentar |
|------|--|
| 1 | bibirnya dientup tawon kah? 🙄 |
| ... | |
| 1087 | Berbakti pada ortu sbg laki ² tentu harus tp tidak dg melupakan nafkah utama pada istri krn anda sudah menikah. Belajar lagi yang bener abang mantan teller gausah playing victim 🙄 |

Setelah menyelesaikan proses crawling data, langkah selanjutnya adalah tahap text preprocessing yang diterapkan sebagai bagian dari proses analisis data, di mana text preprocessing merupakan langkah krusial dalam mempersiapkan data mentah sebelum dilakukan analisis lebih lanjut (Suryati et al., 2023). Preprocessing dilakukan menggunakan Google Colab dan melibatkan berbagai teknik analisis, seperti cleaning data, case folding, tokenizing, normalization, stopword, dan stemming.

Cleaning data digunakan untuk menghapus karakter-karakter yang tidak relevan atau tidak diinginkan dari komentar. Case folding bertujuan mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Normalization berfungsi untuk mengubah kata-kata non-standar, seperti bahasa gaul, kata tidak baku, atau singkatan, menjadi bentuk standar. Stopword digunakan untuk menghapus kata-kata umum atau kata dasar yang tidak memberikan makna penting dalam analisis teks. Terakhir, stemming adalah proses untuk menghilangkan imbuhan dari kata. Hasil dari proses preprocessing data dapat dilihat pada Tabel 2

Tabel 2. Hasil Preprocessing Data

| No | Komentar | komentar_english | Label Sentimen |
|-----|----------------------|--------------------------------|----------------|
| 1 | bibir disengat tawon | wasp stung on the lips | Netral |
| 2 | gak berat apa kak | it doesn't weigh anything, bro | Netral |
| ... | ... | ... | ... |

| | | | |
|----|---|--|---------|
| 93 | bakti orang tua | devotee as a man but forget the | Positif |
| 4 | sebagai laki tapi lupa nafkah utama istri karna meni ajar bener abang mantan teller gausah playing victim | main livelihood of the wife because meni teach the real brother of the former teller gausah playing victim | |
| 93 | siapa yg | who invited the aunt here to | Negatif |
| 3 | membenarkan si tante kan sini bahas cincin | discuss the ring | |

Setelah dilakukan proses preprocessing data, terdapat beberapa data yang ditolak karena dianggap tidak akurat dan kurang terstruktur. Dari 1.088 data point awal, hanya tersisa 978 data setelah proses cleaning selesai.

langkah selanjutnya adalah pelabelan data. Sebelum melakukan pelabelan, data terlebih dahulu diterjemahkan ke dalam bahasa Inggris menggunakan library translate. Library tersebut memiliki fungsi yang disebut dengan `convert_eng` untuk menerjemahkan teks dari bahasa Indonesia ke bahasa Inggris. Hasil dari translate dapat dilihat pada Gambar 1.

| | Komentar | komentar_english |
|-----|---|---|
| 0 | bibir disengat lawon | wasp stung on the lips |
| 1 | niat nya mau lihat seksi kok gini ya jijik deh... | The intention was to look sexy like this, but ... |
| 2 | gak berat apa kak | it doesn't weigh anything, bro |
| 3 | mamah takut nak | mamah scared son |
| 4 | astaghfirullah haladzim banyak komennya bikin ... | astaghfirullah haladzim bnyak comments make a ... |
| ... | ... | ... |
| 929 | ah masa sih tiap hari w pantengin lgs bukan dr... | ah the time is every day w pantengin lgs not d... |
| 930 | sampai teuku istri kamu perlu bukti kamu meman... | until my wife you need proof you do love think... |
| 931 | yang mksut dari lahir smpe skrng mba ga di sen... | what that means is that from birth until now, ... |
| 932 | bakti orang tua sebagai laki tapi lupa nafkah ... | devotee as a man but forget the main livellhoo... |
| 933 | siapa yg membenarkan si tante kan sini yg baha... | who invited the aunt here to discuss the ring |

Gambar 1: Hasil Translate

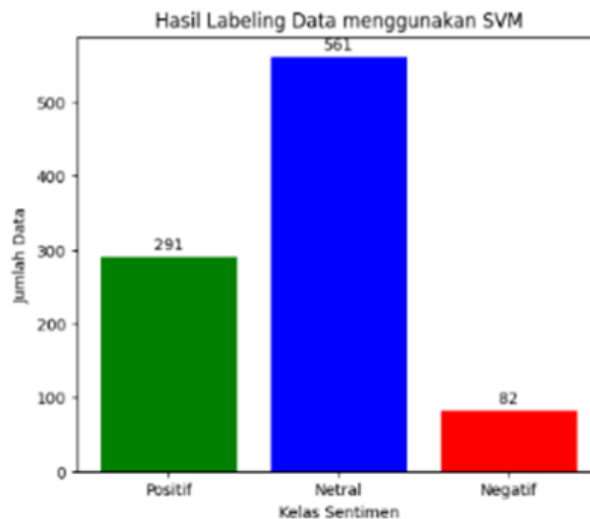
Setelah semua data diterjemahkan ke dalam bahasa Inggris, langkah selanjutnya adalah tahap pelabelan data. Pelabelan dilakukan dengan menggunakan TextBlob, yang memanfaatkan variabel polaritas untuk mengakumulasikan nilai polaritas data. Hasil dari labeling data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Labeling Data

| | Hasil |
|---------------------|--|
| <i>Cleaning</i> | Berbakti pada ortu sbg laki ² tentu harus tp tidak dg melupakan nafkah utama pada istri krn anda sudah menikah. Belajar lagi yang bener abang mantan teller gausah playing victim |
| <i>Case Folding</i> | berbakti pada ortu sbg laki ² tentu harus tp tidak dg melupakan nafkah utama pada istri krn anda sudah menikah. belajar lagi yang bener abang mantan teller gausah playing victim |

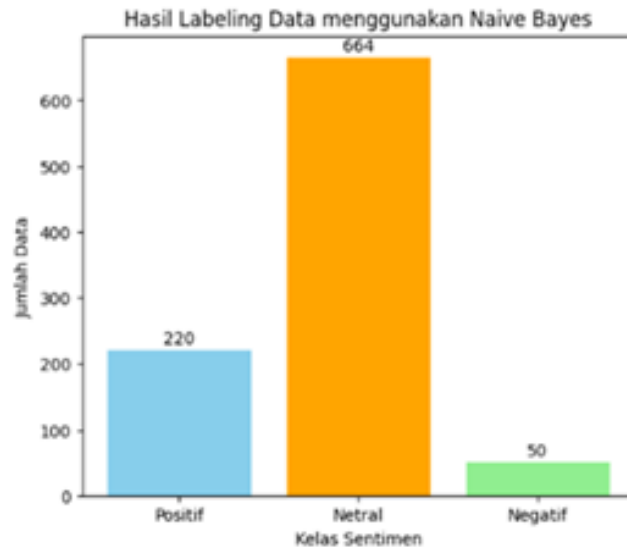
| | |
|----------------------|---|
| <i>Tokenizing</i> | [berbakti, pada, ortu, sbg, laki, tentu, harus, tp, tidak, dg, melupakan, nafkah, utama, pada, istri, krn, anda, sudah, menikah, belajar, lagi, yang, bener, abang, mantan, teller, gausah, plying, victim] |
| <i>Normalization</i> | berbakti pada orang tua sebagai laki tentu harus tapi tidak dengan melupakan nafkah utama pada istri karna anda sudah menikah belajar lagi yang bener abang mantan teller gausah plying victim |
| <i>Stemming</i> | bakti pada ortu sbg laki tentu harus tp tidak dg lupa nafkah utama pada istri krn anda sudah meni ajar lagi yang bener abang mantan teller gausah playing victim |

Dari total 934 komentar yang diberi label, 299 komentar memiliki sentimen positif, 126 komentar memiliki sentimen negatif, dan 509 komentar sisanya memiliki sentimen netral. Setelah dilakukan analisis data, langkah selanjutnya adalah melakukan pelabelan data dengan metode SVM untuk menentukan banyaknya data positif, negatif, dan netral. Hasil labeling data menggunakan metode SVM dapat dilihat pada Gambar 2.



Gambar 2: Hasil Labeling Data Menggunakan Metode SVM

Dari total 934 komentar, terdapat 291 komentar bersentimen positif, 82 komentar bersentimen negatif, dan 561 komentar bersentimen netral. Selanjutnya akan dilakukan pelabelan dengan metode Naive Bayes untuk mengetahui kuantitas data positif, negatif, dan negatif. Hasil Labeling data menggunakan metode Naive Bayes dapat dilihat pada Gambar 3.



Gambar 3: Hasil Labeling Data Menggunakan Metode Naive Bayes

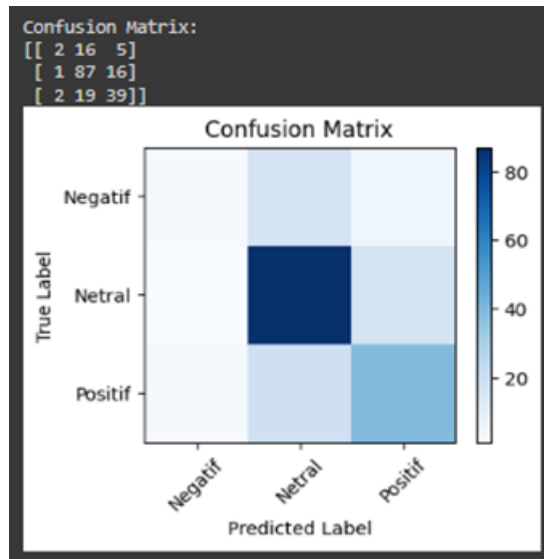
Dari total 934 komentar, terdapat 220 komentar bersentimen positif, 50 komentar bersentimen negatif, dan 664 komentar bersentimen netral.

Analisis sentimen menggunakan metode SVM dan Naive Bayes dilakukan dengan rasio pembagian data sebesar 80% berbanding 20%, dimana 20% digunakan untuk data latih dan 80% untuk data uji (Suryati et al., 2023). Hasil evaluasi metode SVM dapat dilihat pada Gambar 4.

```
Hasil Validasi Model:  
Accuracy: 0.6844919786096256  
Classification Report:  
              precision    recall  f1-score   support  
  
   Negatif      0.40      0.09      0.14         23  
   Netral       0.71      0.84      0.77        104  
   Positif      0.65      0.65      0.65         60  
  
 accuracy              0.68         187  
 macro avg              0.59         187  
 weighted avg           0.65         187
```

Gambar 4: Hasil Evaluasi Metode SVM

Hasil evaluasi model SVM menunjukkan kinerja yang baik secara keseluruhan dengan akurasi sebesar 68%. Untuk memastikan bahwa model klasifikasi telah bekerja dengan baik, maka akan dilakukan analisis dengan menggunakan confusion matrix. Hasil dari confusion matrix metode SVM dapat dilihat pada gambar 5.



Gambar5. Hasil Confusion Matrix Metode SVM

Dari Gambar 5, terlihat bahwa untuk data negatif aktual, ada 2 data yang diklasifikasikan dengan benar sebagai negatif, 16 data yang seharusnya negatif tetapi salah diprediksi sebagai netral, dan 5 data negatif yang diprediksi keliru sebagai positif. Untuk data netral aktual, terdapat 1 data netral yang salah diklasifikasikan sebagai negatif, 87 data netral yang diprediksi dengan benar sebagai netral, dan 16 data netral yang diprediksi salah sebagai positif. Sementara itu, untuk data positif aktual, ada 2 data positif yang salah diprediksi sebagai negatif, 19 data positif yang diprediksi keliru sebagai netral, dan 39 data positif yang diklasifikasikan dengan benar sebagai positif. Selanjutnya, dilakukan cross-validation untuk mengevaluasi kinerja metode SVM dengan membagi data secara acak dan mengelompokannya ke dalam beberapa grup sesuai dengan nilai K pada k-fold. K-fold cross-validation melibatkan proses validasi silang sebanyak k kali yang telah ditentukan. Hasil dari cross validation dapat dilihat pada Gambar 6.

```
# Lakukan validasi silang dan hitung skor cv_scores
cv_scores = cross_val_score(svm_model, X_test_tfidf, y_test, cv=kf)

# Tampilkan hasil validasi silang
print(f"\nCross-Validation Results (Accuracy) for {num_folds} Folds:")
print(cv_scores)
print(f"Mean Accuracy: {np.mean(cv_scores):.4f} +/- {np.std(cv_scores):.4f}")

# Secara opsional, tampilkan skor lipatan individual (fold scores)
for fold_idx, score in enumerate(cv_scores, start=1):
    print(f"Fold {fold_idx} Accuracy: {score:.4f}")

Cross-Validation Results (Accuracy) for 5 Folds:
[0.63157895 0.57894737 0.64864865 0.56756757 0.51351351]
Mean Accuracy: 0.5881 +/- 0.0482
Fold 1 Accuracy: 0.6316
Fold 2 Accuracy: 0.5789
Fold 3 Accuracy: 0.6486
Fold 4 Accuracy: 0.5676
Fold 5 Accuracy: 0.5135
```

Gambar 6. Hasil Cross Validation Metode SVM

Hasil cross-validation menunjukkan bahwa rata-rata akurasi adalah 58,81%, yang mengindikasikan bahwa model memiliki kinerja yang kurang memuaskan. Akurasi untuk masing-masing lipatan bervariasi antara 51,35% hingga 64,86%, menandakan bahwa model SVM kurang konsisten dalam performanya terhadap subset data yang berbeda. Sedangkan hasil evaluasi metode Naive Bayes dapat dilihat pada Gambar 7.

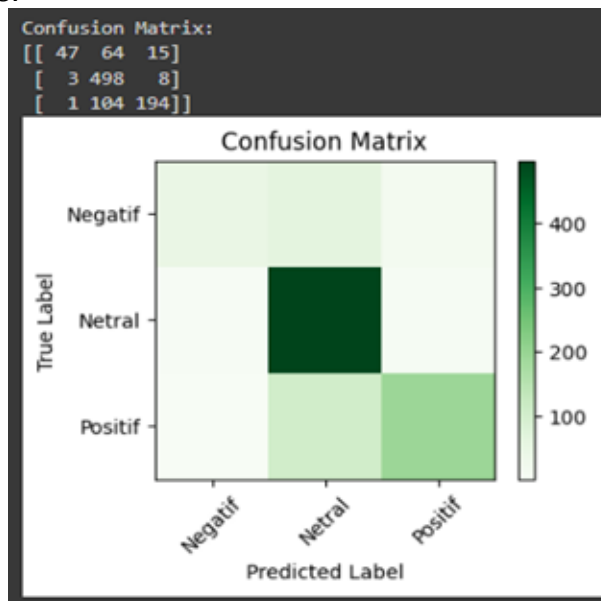
```
Akurasi Test: 0.791220556745182
Classification Report:

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Negatif | 0.92 | 0.37 | 0.53 | 126 |
| Netral | 0.75 | 0.98 | 0.85 | 509 |
| Positif | 0.89 | 0.65 | 0.75 | 299 |
| accuracy | | | 0.79 | 934 |
| macro avg | 0.85 | 0.67 | 0.71 | 934 |
| weighted avg | 0.82 | 0.79 | 0.77 | 934 |

Gambar 7. Hasil Evaluasi Metode Naive Bayes

Gambar 7 menunjukkan bahwa model Naive Bayes mencapai akurasi yang cukup tinggi, yaitu 79,12%. Secara umum, model Naive Bayes bekerja dengan sangat baik untuk kelas yang lebih sering muncul, seperti netral dan positif, namun masih memerlukan perbaikan dalam mendeteksi kelas yang jarang muncul, seperti negatif. Untuk memastikan bahwa model klasifikasi berfungsi dengan baik, analisis akan dilakukan menggunakan confusion matrix pada metode Naive Bayes. Hasil confusion matrix dapat dilihat pada Gambar 8.



Gambar 8. Hasil Confusion Matrix Metode Naive Bayes

Gambar 8 menunjukkan bahwa untuk data aktual negatif, terdapat 47 data yang benar-benar negatif dan diprediksi dengan tepat sebagai negatif, 64 data yang sebenarnya negatif namun salah diprediksi sebagai netral, serta 15 data negatif yang keliru diprediksi sebagai positif. Untuk data aktual netral, terdapat 3 data netral yang salah diprediksi sebagai negatif, 498 data netral yang diprediksi dengan benar sebagai netral, dan 8 data netral yang salah diprediksi sebagai positif. Sedangkan untuk data aktual positif, terdapat 1 data positif yang keliru diprediksi sebagai negatif, 104 data positif yang salah diprediksi sebagai netral, dan 194 data positif yang diprediksi dengan benar sebagai positif.

Selanjutnya, dilakukan validasi silang untuk menilai efisiensi metode Naive Bayes. Hasil dari cross-validation dengan metode Naive Bayes dapat dilihat pada Gambar 9.

```
# Lakukan cross-validation
cv_scores = cross_val_score(textblob_cl, X_train, y_train, cv=kf)

# Tampilkan hasil validasi silang
print(f"\nHasil Validasi Silang (Akurasi) untuk {num_folds} Lipatan:")
print(cv_scores)
print(f"Mean Accuracy: {np.mean(cv_scores):.4f} +/- {np.std(cv_scores):.4f}")

# Secara opsional, tampilkan skor lipatan individual (fold scores)
for fold_idx, score in enumerate(cv_scores, start=1):
    print(f"Fold {fold_idx} Accuracy: {score:.4f}")

Hasil Validasi Silang (Akurasi) untuk 5 Lipatan:
[0.67333333 0.70666667 0.60402685 0.63758389 0.69127517]
Mean Accuracy: 0.6626 +/- 0.0372
Fold 1 Accuracy: 0.6733
Fold 2 Accuracy: 0.7067
Fold 3 Accuracy: 0.6040
Fold 4 Accuracy: 0.6376
Fold 5 Accuracy: 0.6913
```

Gambar 9. Hasil Cross Validation Metode Naive Bayes

Akurasi dari hasil cross-validation metode Naive Bayes cukup tinggi, berkisar antara 60,40% hingga 70,67%, yang menunjukkan bahwa model Naive Bayes menunjukkan performa yang relatif konsisten pada berbagai subset data

3.2 Pembahasan

Pengumpulan data dilakukan secara daring melalui media sosial Instagram dengan metode *crawling* menggunakan web Export Comments. Sebanyak 1.088 komentar dari netizen yang diambil dari 11 akun Instagram *public figure*. Data tersebut akan dibersihkan melalui beberapa tahap text *preprocessing*, yaitu *Cleaning* data, *Case folding*, *Tokenizing*, *Normalization*, *Stopword*, dan *Stemming* seperti penelitian yang dilakukan oleh [1].

Setelah tahap *preprocessing*, jumlah data yang awalnya 1.088 berkurang menjadi 934, karena 154 data yang tidak relevan atau tidak diinginkan dihapus selama proses *preprocessing*, sesuai dengan penelitian yang dilakukan oleh [2]. Selanjutnya data akan diterjemahkan ke dalam bahasa Inggris sebelum lanjut ke tahap Labeling data seperti penelitian yang dilakukan oleh [3].

Setelah menguji data komentar *cyberbullying* di media sosial Instagram dengan menggunakan metode SVM dan Naive Bayes, hasil penelitian menunjukkan bahwa metode SVM memiliki akurasi sebesar 68%, sementara metode Naive Bayes mencapai akurasi sebesar 79%. Data dibagi dengan proporsi 80% untuk data uji dan 20% untuk data latih.

Analisis sentimen komentar *cyberbullying* di media sosial Instagram memiliki komentar sentimen netral sekitar 60,1% dari total data, komentar negatif sekitar 8,8% lebih rendah dibandingkan dengan komentar positif sekitar 31,2%.

Berdasarkan hasil analisis, proporsi label netral jauh lebih tinggi dibandingkan dengan label negatif dan positif. Meskipun komentar mungkin tampak netral, beberapa di antaranya dapat memiliki implikasi negatif atau mengandung penekanan tertentu yang sebenarnya bersifat merendahkan atau mengintimidasi. Dengan mengidentifikasi dominasi sentimen netral dalam komentar *cyberbullying* di Instagram, dapat mendorong pengguna Instagram untuk lebih berhati-hati dalam berkomentar, karena sentimen netral bisa menyebabkan ketidakpedulian terhadap dampak sebenarnya dari komentar tersebut.

4. KESIMPULAN

Setelah melakukan pengujian pada data komentar cyberbullying di media sosial Instagram dengan menggunakan metode SVM dan Naive Bayes, hasil penelitian menunjukkan bahwa akurasi metode SVM adalah 68%, sementara akurasi metode Naive Bayes mencapai 79%, dengan pembagian data testing sebesar 80% dan data training sebesar 20%. Oleh karena itu, dapat disimpulkan bahwa metode Naive Bayes memiliki akurasi yang lebih tinggi dibandingkan dengan metode SVM.

Dalam analisis sentimen komentar cyberbullying di media sosial Instagram, sekitar 60,1% dari total komentar memiliki sentimen netral, menunjukkan bahwa komentar-komentar ini bersifat informatif atau tidak memihak. Proporsi komentar negatif sekitar 8,8%, yang lebih rendah dibandingkan dengan komentar positif yang mencapai 31,2%. Meski demikian, komentar negatif tetap memerlukan perhatian khusus karena dapat memiliki dampak signifikan, terutama terkait dengan isu seperti cyberbullying.

5. DAFTAR PUSTAKA

- Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Media Informatika Budidarma*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Apriyani, H., & Kurniati. (2020). Perbandingan Metode Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), 133–143. <https://doi.org/10.51519/journalita.volume1.issuue3.year2020.page133-143>
- Asana, I. M. D. P., & Yanti, N. P. D. T. (2023). Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM). 06(02), 123–133.
- Dewi, C. (2023). Indonesia Nomor 1 Negara Dengan Kasus Cyberbullying Terbanyak Di Dunia, Etika Berjejaring: Jarimu Harimaumu! *Bernas.Id*. <https://www.bernas.id/2023/06/163975/indonesia-nomor-1-negara-dengan-kasus-cyberbullying-terbanyak-di-dunia-etika-berjejaring-jarimu-harimaumu/>
- Fadli, D. R. (2023). Mengenal Cyberbullying: Penyebab, Dampak, dan Cara Mengatasinya. *Halodoc*. Andrian, E., & Inain, A. R. (2023). Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Media Informatika Budidarma*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Apriyani, H., & Kurniati. (2020). Perbandingan Metode Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), 133–143. <https://doi.org/10.51519/journalita.volume1.issuue3.year2020.page133-143>
- Asana, I. M. D. P., & Yanti, N. P. D. T. (2023). Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM). 06(02), 123–133.
- Dewi, C. (2023). Indonesia Nomor 1 Negara Dengan Kasus Cyberbullying Terbanyak Di Dunia, Etika Berjejaring: Jarimu Harimaumu! *Bernas.Id*. <https://www.bernas.id/2023/06/163975/indonesia-nomor-1-negara-dengan-kasus-cyberbullying-terbanyak-di-dunia-etika-berjejaring-jarimu-harimaumu/>
- Fadli, D. R. (2023). Mengenal Cyberbullying: Penyebab, Dampak, dan Cara Mengatasinya. *Halodoc*. <https://www.halodoc.com/artikel/mengenal-cyberbullying-penyebab-dampak-dan-cara-mengatasinya>
- Imam, & Santoso, I. (2023). Analisis Sentimen Pada Twiter Terhadap Gagalnya Pelaksanaan Piala Dunia di Indonesia. 7(2).

- Kaburuan, E. R., & Setiawan, N. R. (2023). Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(1), 105–116. <https://doi.org/10.32736/sisfokom.v12i1.1614>
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), 302–307. <https://doi.org/10.30591/jpit.v8i3.5734>
- Sari, D. N., & Basit, A. (2020). Media Sosial Instagram Sebagai Media Informasi Edukasi. *Persepsi: Communication Journal*, 3(1), 23–36. <https://doi.org/10.30596/persepsi.v3i1.4428>
- Suryati, E., Styawati, & Aldino, A. A. (2023). Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Dan Sistem Informasi*, 4(1), 96–106. <https://doi.org/10.33365/jtsi.v4i1.2445>
- Widi, S. (2023). Pengguna Media Sosial di Indonesia Sebanyak 167 Juta pada 2023. *DataIndonesia*. <https://dataindonesia.id/internet/detail/pengguna-media-sosial-di-indonesia-sebanyak-167-juta-pada-2023>
- Yi, P., & Zubiaga, A. (2023). Session-based cyberbullying detection in social media: A survey. *Online Social Networks and Media*, 36(June), 100250. <https://doi.org/10.1016/j.osnem.2023.100250>