

## Analisis Sentimen Komentar Konten Edukatif Di Instagram Dengan Metode *Naïve Bayes* Dan *Support Vector Machine*

Ferry Andriansyah<sup>1</sup>, Puji Astuti<sup>2\*</sup>

<sup>1</sup> Universitas Nusa Mandiri  
e-mail: <sup>1</sup>ferji2007@gmail.com

<sup>2</sup>Universitas Bina Sarana Informatika  
e-mail: <sup>2</sup>puji.pat@bsi.ac.id

| Diterima   | Direvisi   | Disetujui  |
|------------|------------|------------|
| 25-10-2024 | 29-11-2024 | 21-01-2025 |

**Abstrak** - Pesatnya perkembangan teknologi informasi mempengaruhi penggunaan media sosial, termasuk *Instagram*. *Instagram* menjadi salah satu platform yang cukup populer untuk berbagi konten edukatif. Penelitian ini dimaksudkan untuk menganalisa sentimen menggunakan metode *Naïve Bayes* (NB) dan *Support Vector Machine* (SVM) yang ada pada komentar didalam unggahan akun *Instagram content creator* yang berfokus pada pembuatan konten edukatif. Adapula penggunaan 4600 data komentar *Instagram* yang diklasifikasikan dalam tiga kelas sentimen berbeda yaitu, positif, netral dan negatif. Proses analisis sentimen melibatkan beberapa proses dan tahapan, diantaranya yaitu *text preprocessing*, ekstraksi fitur, dan penerapan algoritma *Naïve Bayes* dan *SVM* untuk klasifikasi sentiment. Hasil penelitian menunjukkan bahwa metode algoritma *Naïve Bayes* lebih unggul dibandingkan algoritma *SVM*. Ini ditunjukkan dengan nilai akurasi pada metode *Naïve Bayes* menunjukkan tingkat akurasi yang lebih tinggi dalam mengklasifikasikan sentimen dibandingkan dengan metode *Support Vector Machine*. Acuan ini ditunjukkan terhadap nilai akurasi sebesar 84% pada *Naïve Bayes* dan 80% pada *Support Vector Machine*.

Kata Kunci: Analisis Sentimen, Sosial Media, Text Mining

**Abstract** - The rapid development of information technology is changing the use of social media, including *Instagram*. *Instagram* is becoming one of the most popular platforms for sharing educational content. This study was conducted to analyze how often using the Bayes method (NB) and Support Vector Machine (SVM) in the comments on the *Instagram content creator* account that helps in the creation of educational content. The 4600 comments on *Instagram* collected in three classes are often positive, positive, neutral and negative. Process analysis often involves several processes and stages, namely *text preprocessing*, feature analysis, and analysis of the overall algorithm and SVM for sentiment classification. The results showed that the Na Bay algorithm method is indicated by the value of accuracy in the *Naïve Bayes* method shows a higher level of accuracy in classifying the increase compared to the *Support Vector Machine* method. This reference is made to the accuracy value of 84% in Bayes and 80% in the supporting Vector Machine.

Keyword: Sentiment Analysis, Social Media, Text Mining

### PENDAHULUAN

Dalam era digital saat ini, media social telah menjadi platform yang penting untuk berbagi informasi dan pengetahuan. *Instagram*, sebagai salah satu platform social yang populer, sering digunakan untuk menyebarkan konten edukatif. Beberapa alasan dalam mencari informasi pada *Instagram* yaitu dapat memberikan kemudahan pada pengguna dalam melakukan penyebaran informasi secara visual dengan tampilan yang menarik (Anisah et al., 2021).

Komentar yang muncul dibawah konten ini dapat memberikan wawasan berharga tentang bagaimana netizen merespons informasi yang disajikan. Analisis sentiment terhadap komentar tersebut menjadi penting untuk memahami persepsi public serta efektifitas konten edukatif. Sentiment analysis juga merupakan text classification yang berfokus pada pengolahan Bahasa, dimana tujuannya untuk menganalisis pendapat sentiment baik negative maupun positif (Mailoa, 2021).



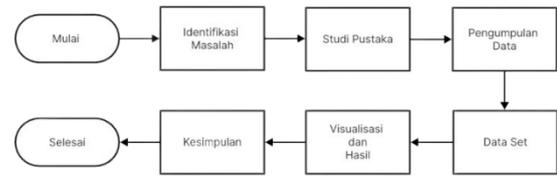
Metode analisis sentiment dapat dibagi menjadi beberapa pendekatan, diantaranya adalah *Naïve Bayes* dan *Support Vector Machine* (SVM). *naïve bayes*, yang berbasis pada teorema probabilitas, terkenal karena kesederhanaannya dan efisiensinya dalam klasifikasi teks. *Naïve bayes classifier* digunakan sebagai model dalam menentukan komentar dimana komentar tersebut termasuk kedalam komentar negative, positif ataupun netral (Chatrina Siregar et al., 2020). Metode *naïve bayes* mempunyai kelebihan yaitu lebih cepat dalam proses perhitungan, metode sederhana dan mampu memiliki nilai akurasi yang baik (Putro et al., 2020).

Sistem prediksi tingkat kelulusan mahasiswa dengan metode *naïve bayes* dengan data training 13 dan data testing 5 didapatkan nilai akurasi dengan *naïve bayes* sebesar 80% (Muqorobin & Bagoes Pakarti, 2022). Dalam proses sentiment analisis juga dapat menggunakan metode SVM, dimana algoritma SVM mampu dikategorikan sebagai algoritma yang baik, dimana mampu mengenali data latih dan data uji (Abdusyukur, 2023). Algoritma SVM mampu menganalisa sentiment pengguna twitter, hasil klasifikasi menggunakan SVM didapat tingkat akurasi sebesar 93% (Alhaq, dkk, 2021).

Perbandingan metode SVM dan *Naïve Bayes* dalam analisis sentimen twitter pada *marketplace* Tokopedia, Shoope, dan Lazada yaitu untuk *Naïve Bayes* sebesar 83.52% dan untuk SVM 88.93% dapat dikatakan bahwa metode SVM memiliki nilai akurasi yang baik dalam sentiment analysis (Kurniawan et al., 2023). Pada penelitian ini, peneliti menganalisa komentar yang ada pada Instagram menggunakan 2 metode yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB). *Text mining* merupakan teknik yang digunakan untuk menangani pola Bahasa yang tidak terstruktur kemudian di klasifikasikan, *clustering*, dan informasi (Agustina et al., 2021). Tujuannya adalah untuk mengetahui bagaimana setiap metode dapat memberikan hasil yang berbeda dalam pengklasifikasian sentiment, serta untuk mengevaluasi keakuratan dan efektivitas masing-masing metode dalam konteks analisis sentiment pada social media. Komparasi algoritma SVM dan NB untuk analisis sentiment analisis pada aplikasi BRImo, didapat nilai akurasi untuk SVM 97.69% dan NB 96.53% untuk itu maka SVM adalah algoritma yang lebih baik dibanding algoritma lain (Puji Astuti et al., 2022). Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan metode analisis sentiment serta memberikan wawasan yang lebih dalam mengenai interaksi pengguna pada platform media social.

## METODE PENELITIAN

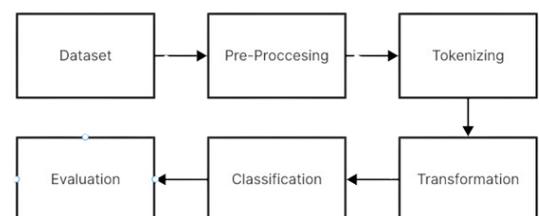
Metode penelitian ini dirancang secara sistematis untuk merancang, melaksanakan, dan menganalisis penelitian dengan tepat. Tahapan yang dilakukan dijelaskan melalui gambar berikut:



Sumber: Peneliti (2024)

Gambar 1. Metode Penelitian

- a. **Identifikasi Masalah**  
Penelitian ini berfokus pada analisis sentiment komentar di Instagram terhadap unggahan edukatif. Variasi komentar yang beraneka ragam membutuhkan metode untuk mengidentifikasi sentiment dan menyeleksi kata-kata yang relevan.
- b. **Studi Pustaka**  
Pengumpulan informasi dilakukan melalui media *social* Instagram serta referensi dari jurnal ilmiah terkait.
- c. **Pengumpulan Data**  
Data dikumpulkan dari komentar unggahan Instagram menggunakan ekstensi *IGCommentsExport* di peramban *Google Chrome* yang mengonversi data komentar menjadi *file csv* yang dapat diolah menggunakan *spreadsheet*.
- d. **Data Set**  
Setelah data diperoleh, proses pra-pemrosesan dilakukan termasuk *cleaning*, *tokenizing*, *filtering*, dan *stemming*. Proses ini menghasilkan data yang siap dianalisis untuk menemukan pola *sentiment*. Gambar dibawah ini menunjukkan alur dalam pengolahan data set.



Sumber: Peneliti (2024)

Gambar 2. Alur Pengolahan Data Set

- e. **Visualisasi Dan Hasil**  
Hasil penelitian divisualisasikan menggunakan *Word Cloud*, yang menampilkan kata-kata yang paling sering digunakan dalam komentar, kata-kata positif akan lebih besar ukurannya, sedangkan kata negative ditampilkan lebih kecil.
- f. **Proses**  
Penelitian ini dimulai dengan eksplorasi data dari Instagram @ferryirwandi. Data komentar diambil dari unggahan yang ditargetkan menggunakan *IGCommentsExport*. Data komentar kemudian melalui pra-pemrosesan

seperti *case folding*, *tokenizing*, *filtering* dan *stemming*. Tahapan ini diikuti oleh klasifikasi data menjadi sentiment positif, netral dan negative menggunakan metode *Naïve Bayes* dan SVM. Visualisasi pola kata dilakukan menggunakan *Word Cloud*, sementara evaluasi kinerja algoritma dilakukan menggunakan *confusion matrix*, menghitung nilai akurasi sentivitas, presisi, *recall* dan *F1 Score*.

## HASIL DAN PEMBAHASAN

Beberapa pengujian dilakukan untuk mendapatkan hasil dari pengujian pada penelitian kali ini. Tahapannya meliputi pengumpulan data, implementasi *pre-processing*, pelabelan kelas sentimen, pengujian klasifikasi serta evaluasi model algoritma.

### a. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil komentar dari akun Instagram yang diteliti dalam periode unggahan 22 Maret – 1 Juli 2024. Data yang digunakan berjumlah 4600 komentar dari unggahan Instagram Ferry Irwandi. Proses pengumpulan data ini sepenuhnya dibantu oleh *tools IGCommentExport* yang tersedia dalam ekstensi *browser Google Chrome*. Data yang berhasil dikumpulkan disimpan dalam format file *.csv*.

Data hasil scraping awalnya masih mengandung beberapa variabel yang tidak diperlukan, seperti profil pengguna, avatar, user ID, dan tanggal. Oleh karena itu, tahap *pre-processing* diperlukan untuk mendapatkan teks komentar yang bersih dari noise dan variabel yang tidak relevan.

### b. Implementasi Pre-Processing

Data komentar yang diperoleh dari proses scraping belum dapat digunakan secara langsung, mengingat bentuk dan struktur teksnya yang tidak teratur serta adanya banyak noise, seperti emoji. Oleh karena itu, langkah pertama yang dilakukan adalah pembersihan data (*cleaning*), yang bertujuan untuk menghapus teks yang tidak relevan.

Dalam tahap *cleaning*, variabel yang tidak diperlukan, seperti username (@), hashtag (#), tanda baca, angka, karakter tunggal, dan URL, akan dihapus untuk mengurangi beban dalam pemrosesan dan dianggap tidak relevan untuk analisis.

Untuk mendapatkan hasil yang lebih akurat, penghapusan data duplikat juga dilakukan guna meningkatkan presisi dan *recall* model. Dalam tahap ini, atribut yang tidak diperlukan dalam analisis telah dihapus, sehingga mempermudah proses klasifikasi dan pemrosesan data sebelum tahap analisis lebih lanjut. Data komentar dengan indeks "*comment*" mengandung kata-

kata yang perlu dinormalisasi, seperti kata tidak baku dan singkatan yang akan diubah menjadi kata baku. Proses selanjutnya adalah *Case Folding*, yang bertujuan untuk menyeragamkan huruf dari kapital ke huruf kecil. Penerapan proses ini menggunakan fungsi Python *lower* (). Selain itu, proses *tokenizing* dilakukan untuk memisahkan kalimat menjadi satuan kata atau token. Setelah proses *tokenizing*, langkah selanjutnya adalah *filtering*, yang bertujuan untuk menghilangkan kata yang tidak memiliki makna dan tidak diperlukan. Penggunaan *library Sastrawi* mempermudah deteksi kata yang akan difilter secara otomatis. Terakhir, dilakukan tahap *stemming* untuk mendapatkan kata dasar sesuai kaidah KBBI. Proses ini mengambil data dari tahap *tokenization*, yang telah dilakukan sebelumnya, untuk memastikan bahwa kata-kata yang dianalisis berada dalam bentuk dasarnya.

### c. Pelabelan Kelas Sentimen

Setelah tahap *preprocessing* selesai, langkah berikutnya adalah pelabelan kelas sentimen. Proses ini menggunakan *library TextBlob* untuk menerjemahkan data komentar ke dalam bahasa Inggris sebelum dilakukan *labeling*. *Labeling* dibagi menjadi tiga kelas, yaitu sentimen positif, negatif, dan netral.

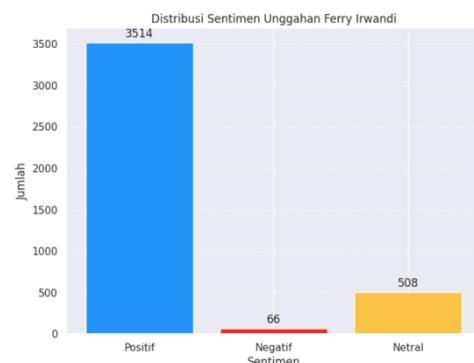
Hasil dari proses pelabelan dapat dilihat pada Tabel dibawah dan grafik perbandingan sentimen disajikan pada Gambar dibawah ini.

Tabel 1. Hasil Proses *Labeling* Sentimen

|             |      |
|-------------|------|
| Jumlah Data | 4088 |
| Positif     | 3514 |
| Netral      | 508  |
| Negatif     | 66   |

Sumber: Peneliti (2024)

Berikut Grafik Perbandingannya untuk kategori positif dan negatif



Sumber: Peneliti (2024)

Gambar 3. Grafik Perbandingan Kelas Sentimen

### d. Visualisasi Data

Visualisasi data dilakukan berdasarkan

keseluruhan data yang telah melalui tahap pelabelan. Penggunaan *word cloud* bertujuan untuk menampilkan kata-kata yang sering muncul pada komentar. Hasil visualisasi menunjukkan frekuensi kata yang muncul, di mana kata yang lebih sering muncul ditampilkan dengan ukuran *font* yang lebih besar.

**e. Pengujian Klasifikasi**

Pengujian dilakukan menggunakan dua metode, yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*. Sebelum proses klasifikasi, dilakukan *splitting* data dengan perbandingan 80% untuk data latih dan 20% untuk data uji.

**1) Support Vector Machine**

Pengujian klasifikasi menggunakan SVM menunjukkan hasil yang baik dalam mengklasifikasikan kelas positif dengan skor *precision*, *recall*, dan *f1-score* yang tinggi.

**2) Naïve Bayes**

Hasil pengujian dengan metode *Naïve Bayes* juga menunjukkan kinerja yang baik dalam mengklasifikasikan data dengan kelas positif, meskipun performa pada kelas netral dan negatif perlu diperbaiki.

**f. Evaluasi**

Evaluasi model dilakukan menggunakan *confusion matrix* untuk kedua metode klasifikasi. Tabel dibawah menyajikan perbandingan hasil klasifikasi antara SVM dan *Naïve Bayes*, dengan nilai akurasi, *precision*, *recall*, dan *f1-score* yang dihasilkan oleh masing-masing metode.

Tabel 2. Perbandingan Hasil Klasifikasi

| Model                        | Accuracy Score | Precision | Recall   | F1-Score |
|------------------------------|----------------|-----------|----------|----------|
| Support Vector Machine (SVM) | 0.798289       | 0.728165  | 0.798289 | 0.759480 |
| Naïve Bayes (NB)             | 0.844743       | 0.713591  | 0.844743 | 0.773648 |

Sumber: Peneliti (2024)

**KESIMPULAN**

Berdasarkan hasil analisis sentiment terhadap komentar pada media social Instagram, dengan menggunakan algoritma *Naïve Bayes* dan *Support Vectro Machine (SVM)*, diperoleh data sebanyak 4600 komentar. Hasil pelabelan kelas sentiment menunjukan distribusi 3514 komentar positif, 508 komentar netral dan 66 komentar negative. Tahapan pelabelan otomatis menawarkan kepraktisan dalam pengolahan data, meskipun dapat menimbulkan bias, terutama pada klasifikasi sentiment netral dan negative. Penerapan metode klasifikasi dilakukan dengan membagi data menjadi rasio pembanding 80:20, dimana *naïve bayes* menunjukkan tingkat akurasi yang lebih tinggi 84% disbanding dengan SVM sebesar 80%.

## REFERENSI

- Abdusyukur, F. (2023). Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter. *Komputa: Jurnal Ilmiah Komputer Dan Informatika*, 12(1), 73–82. <https://doi.org/10.34010/komputa.v12i1.9418>
- Agustina, D. A., Subanti, S., & Zukhronah, E. (2021). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine. *Indonesian Journal of Applied Statistics*, 3(2), 109. <https://doi.org/10.13057/ijas.v3i2.44337>
- Alhaq, Zidna, Mustopa, Ali, Mulyatun, Sri, Santoso, J. D. (2021). Optimasi Algoritma Support Vector Machine untuk Analisis Sentimen pada Ulasan Produk Tokopedia Menggunakan PSO. *Media Informatika*, 20(2), 97–108. <https://doi.org/10.37595/mediainfo.v20i2.59>
- Anisah, N., Sartika, M., & Kurniawan, H. (2021). Penggunaan Media Sosial Instagram dalam Meningkatkan Literasi Kesehatan Pada Mahasiswa. *Jurnal Peurawi: Media Kajian Komunikasi Islam*, 4(2), 94. <https://doi.org/10.22373/jp.v4i2.11080>
- Chatrina Siregar, N., Ruli, R., Siregar, A., Yoga, ; M., & Sudirman, D. (2020). Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ). *Jurnal Teknologi Aliansi Perguruan Tinggi (APERTI) BUMN*, 3(1), 102–110.
- Kurniawan, I., Lia Hananto, A., Shofia Hilabi, S., Hananto, A., Priyatna, B., & Yuniar Rahman, A. (2023). Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(1), 731–740. <http://jurnal.mdp.ac.id>
- Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), 44. <https://doi.org/10.22146/jisph.44455>
- Muqorobin, & Bagoes Pakarti, M. (2022). Sistem Prediksi Lama Studi Kuliah Menggunakan Metode Naive Bayes. *Jurnal Informatika, Komputer Dan Bisnis*, 2(1), 117–129. <https://jurnal.itbaas.ac.id/index.php/jikombis>
- Puji Astuti, A., Alam, S., & Jaelani, I. (2022). Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo. *Jurnal Bangkit Indonesia*, 11(2), 1–6. <https://doi.org/10.52771/bangkitindonesia.v11i2.196>
- Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>