



# Jurnal Ilmiah Informatika (JIF)

| ISSN (Print) 2337-8379 | ISSN (Online) 2615-1049 |

Jurnal online di akses di <http://ejournal.upbatam.ac.id/index.php/jif>



## Analisis Sentimen Komentar pada Postingan Instagram "StandWithUs" Menggunakan Klasifikasi Naive Baye

Khairunnisa<sup>1</sup>, Shanty Kurnia Dewi<sup>2</sup>, Deisy Dzakiyyah Rahmawati<sup>3</sup>, Anggraini Puspita Sari<sup>4</sup>.

<sup>1,2,3,4</sup>Universitas Pembangunan Nasional "Veteran" Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294, Indonesia

### ABSTRACT

Social media platforms like Instagram play a crucial role in shaping public opinion and fostering community engagement in the digital age. The "StandWithUs" Instagram account, dedicated to raising awareness and advocacy, has garnered significant attention and interaction through its posts. However, understanding the sentiment behind user comments on these posts remains challenging. This study addresses this issue by employing Naive Bayes Classification to analyze the sentiment of comments on the "StandWithUs" Instagram account. The primary objective is to accurately classify comments into positive, negative, or neutral categories, providing insights into public opinion and engagement. Our findings indicate that the Naive Bayes Classification model achieves high accuracy in sentiment identification when trained with a substantial dataset. This research highlights the effectiveness of Naive Bayes in conducting sentiment analysis on social media, underscoring its potential to enhance our understanding and management of public opinion on advocacy-related content. The implications of this study are significant, offering valuable perspectives on how social media sentiment analysis can be leveraged to gauge public reactions and involvement, ultimately contributing to more informed advocacy strategies and community engagement efforts. By accurately interpreting the sentiments of user comments, stakeholders can better understand public perception and tailor their content and messaging strategies to foster more meaningful and positive interactions within their digital communities.

#### Sejarah Artikel:

Diterima Redaksi: 08-07-2024

Revisi Akhir: 28-08-2024

Diterbitkan Online: 05-09-2024

#### KATA KUNCI:

Sentiment Analysis,  
Naive Bayes Classification,  
Instagram, Public Opinion,  
Social Media

#### KORESPONDENSI

E-mail: [anggraini.puspita.if@upniatim.ac.id](mailto:anggraini.puspita.if@upniatim.ac.id)

### 1. INTRODUCTION

Platform media sosial sekarang sangat penting bagi kemampuan orang untuk menyuarakan ide dan berpartisipasi dalam diskusi online tentang berbagai subjek di era digital. Konten buatan pengguna di platform media sosial seperti Facebook, Twitter, dan Instagram sangat melimpah dan mewakili pendapat publik tentang berbagai topik. Lonjakan aktivitas internet ini menawarkan alat yang sangat berharga untuk mempelajari opini publik, memungkinkan akademisi dan organisasi untuk menentukan konsensus umum tentang peristiwa tertentu, produk, atau gerakan sosial.

Jaringan Media Sosial adalah teknologi komunikasi yang telah mengubah cara orang berinteraksi dan berkomunikasi. Pentingnya Media Sosial telah diakui secara global dan perusahaan-perusahaan menggunakan Media Sosial untuk mengembangkan bisnis mereka serta mengembangkan alat-alat khusus. Media sosial secara signifikan telah memperluas area cakupan di mana media tradisional (televisi, radio, dan surat kabar) tidak digunakan untuk menjangkau massa, baik individu maupun kelompok. [1]

Dengan penggunaan media sosial yang meluas, pentingnya melacak dan mengevaluasi opini publik semakin meningkat. Analisis sentimen adalah alat penting bagi bisnis, pemerintah, dan organisasi advokasi untuk membuat keputusan yang tepat dan menyesuaikan strategi mereka berdasarkan opini publik. Dalam hal ini, analisis sentimen—metode untuk mengidentifikasi nada emosional dari sebuah teks—sangat penting. Analisis sentimen menafsirkan teks dengan mengkategorikannya sebagai positif, negatif, atau netral, yang membantu mengidentifikasi keyakinan dan sikap mendasar yang diekspresikan oleh orang-orang.

Analisis sentimen dapat dilakukan menggunakan berbagai algoritme, masing-masing memiliki kelebihan dan kekurangan. Klasifikasi Naive Bayes terkenal di antaranya karena kemudahan penggunaan dan efisiensinya. Kemampuan pengklasifikasi probabilistik ini untuk menangani sejumlah besar data secara efektif membuatnya sangat cocok untuk analisis teks. Berdasarkan teorema Bayes dan asumsi independensi fitur, Naive Bayes mengurangi waktu komputasi sekaligus menghasilkan jawaban yang andal untuk berbagai skenario dunia nyata. Klasifikator Naive Bayes (NB) adalah keluarga klasifikator yang didasarkan pada teorema probabilitas populer Bayes, yang dikenal untuk membuat model sederhana

yang kuat, terutama di bidang klasifikasi dokumen dan prediksi penyakit. . [2]

Tujuan utama dari penelitian ini adalah untuk memeriksa sentimen komentar yang ditinggalkan pada postingan Instagram oleh akun @StandWithUs menggunakan klasifikasi Naive Bayes. StandWithUs adalah organisasi nirlaba terkenal yang menggunakan aktivisme media sosial dan pendidikan untuk melawan anti-Semitisme dan mendukung Israel. Karena isi tulisannya yang politis dan emosional, opini kuat dari berbagai perspektif sering tercermin dalam komentarnya. Ini membuatnya menjadi contoh yang sempurna untuk meneliti efektivitas metode analisis sentimen.

Pendekatan penelitian ini mencakup beberapa komponen penting. Sumber asli data adalah Instagram, lebih khusus lagi komentar pada foto StandWithUs. Untuk menghilangkan kebisingan dan informasi yang tidak perlu, data mentah ini dibersihkan dengan cermat. Data yang dibersihkan kemudian diproses lebih lanjut untuk mengekstrak atribut yang relevan, termasuk frekuensi kata dan sifat tekstual lainnya. Selanjutnya, model Naive Bayes dilatih dengan atribut-atribut ini sehingga dapat belajar dari data masukan dan memprediksi sentimen.

Model Naive Bayes dilatih, dan kemudian diterapkan pada data komentar baru untuk memprediksi sentimen publik. Proses ini tidak hanya mengonfirmasi keakuratan model tetapi juga memberikan wawasan real-time tentang opini publik mengenai konten yang dibagikan oleh StandWithUs. Analisis ini memberikan pemahaman menyeluruh tentang reaksi publik dan bagaimana mereka berubah dari waktu ke waktu atau dalam menanggapi postingan tertentu dengan mengklasifikasikan komentar secara metodis sebagai positif, negatif, atau netral.

Diharapkan bahwa temuan penelitian ini akan menunjukkan seberapa baik kinerja klasifikasi Naive Bayes saat analisis sentimen komentar media sosial dilakukan. Kinerja model akan dievaluasi melalui pengujian dan validasi yang ekstensif untuk menjamin ketepatan dan keandalannya. Selain itu, hasilnya akan memberikan pemahaman canggih tentang dinamika opini publik dalam kerangka masalah kontroversial, sehingga memberikan perspektif penting bagi domain yang lebih luas dari analitik media sosial dan penelitian opini publik.

Sebagai hasilnya, penelitian ini menekankan nilai analisis sentimen dalam percakapan modern sekaligus menunjukkan kegunaan penggunaan klasifikasi Naive Bayes dalam analisis sentimen media sosial. Proyek ini bertujuan untuk memberikan alat bagi akademisi dan organisasi untuk menavigasi dunia kompleks media sosial dengan menawarkan metode terstruktur untuk menilai suasana publik. Hal ini pada gilirannya dapat membantu mengembangkan pemahaman yang lebih mendalam tentang sikap dan pandangan sosial serta menginformasikan teknik komunikasi yang lebih efektif.

Tujuan utama dari penelitian ini adalah menggunakan algoritme klasifikasi Naive Bayes untuk memeriksa sentimen publik yang diwakili dalam komentar pada postingan Instagram yang diunggah oleh akun @StandWithUs. Tujuan penelitian ini adalah untuk mempelajari bagaimana pengguna media sosial merespons postingan yang dibuat oleh StandWithUs, sebuah kelompok yang terkenal karena aktivismenya dalam masalah-masalah yang politis dan emosional. Studi ini bertujuan untuk mengukur dan mengkategorikan tanggapan ini menggunakan analisis sentimen, menawarkan wawasan penting ke dalam dinamika opini publik dalam lingkungan ini.

Menilai kinerja algoritme klasifikasi Naive Bayes dalam analisis sentimen teks media sosial adalah tujuan kedua. Naive Bayes adalah metode yang banyak digunakan untuk tugas klasifikasi teks karena kemudahan penggunaan dan efektivitasnya. [3] Secara khusus, penelitian ini akan menilai seberapa baik kinerjanya dalam mengkategorikan perasaan sebagai positif, negatif, atau netral dalam konteks komentar Instagram. Proyek ini berharap dapat memvalidasi ketahanan Naive Bayes untuk aplikasi praktis dalam analitik media sosial dengan cara ini.

Studi ini juga berusaha mengidentifikasi karakteristik penting dalam data teks yang mendukung klasifikasi sentimen. Penelitian ini dapat memberikan wawasan lebih lanjut tentang pola linguistik yang terkait dengan berbagai nada emosional dengan mengidentifikasi kata atau frasa yang paling kuat menyampaikan emosi positif, negatif, atau netral. Data ini mungkin berguna dalam meningkatkan ketepatan algoritme analisis sentimen dan mengembangkan rencana komunikasi media sosial yang lebih kuat.

Memantau bagaimana opini publik berubah sebagai respons terhadap postingan StandWithUs dari waktu ke waktu adalah tujuan lain. Studi ini mencari perubahan sentimen atau tren dalam komentar pada beberapa postingan selama periode waktu tertentu. Analisis temporal ini memberikan pandangan dinamis tentang keterlibatan dan persepsi audiens dengan menjelaskan bagaimana opini publik bergeser sebagai respons terhadap komentar atau kejadian tertentu.

Proyek ini juga bermaksud untuk menyelidiki bagaimana jenis postingan tertentu memengaruhi sentimen. StandWithUs mempublikasikan berbagai konten, seperti pesan advokasi, sumber daya pendidikan, dan pembaruan berita. Penelitian ini bertujuan untuk mengidentifikasi jenis konten yang paling emosional dengan mengklasifikasikan postingan dan mengevaluasi sentimen yang diungkapkan dalam komentar. Ini bisa membantu perusahaan yang ingin mengoptimalkan keterlibatan positif atau menangani umpan balik negatif secara efektif dalam proses pembuatan konten mereka.

Selain itu, dengan mengilustrasikan kegunaan analisis sentimen dalam mengukur opini publik, penelitian ini bertujuan untuk memajukan bidang analitik media sosial. Penelitian ini mencoba membuat kerangka kerja yang dapat direproduksi yang dapat digunakan oleh akademisi dan organisasi lain untuk menganalisis sentimen dalam berbagai keadaan dengan menguraikan pendekatan dan temuan. Mengingat bahwa analisis sentimen menjadi alat yang semakin penting untuk hubungan masyarakat, pemasaran, dan kecerdasan bisnis, kontribusi ini sangat relevan.

Pada akhirnya, tujuan penelitian ini adalah untuk menunjukkan betapa pentingnya analisis sentimen bagi organisasi nirlaba seperti StandWithUs dalam hal manajemen media sosial. Melalui penyediaan wawasan praktis tentang persepsi audiens terhadap materi mereka, penelitian ini dapat membantu organisasi ini menyesuaikan strategi komunikasi mereka dengan lebih efektif. Memahami opini publik dapat membantu organisasi membuat keputusan yang lebih cerdas, berinteraksi dengan pengikut mereka lebih efektif, dan mencapai tujuan mereka di ranah media sosial.

## 2. TINJAUAN PUSTAKA

### 2.1 Media Sosial dan Opini Publik

Media sosial telah mengubah cara masyarakat berinteraksi dan

Sentiment Analysis of Comments

berbagi informasi. Stage seperti Facebook, Twitter, dan Instagram memungkinkan pengguna untuk menyebarkan pandangan dan informasi secara luas dan cepat. [4] Kaplan dan Haenlein (2010) mendefinisikan media sosial sebagai "sekelompok aplikasi berbasis web yang membangun dasar ideologi dan teknologi Web 2.0, dan memungkinkan penciptaan serta pertukaran konten buatan pengguna". Media sosial menjadi alat penting dalam membentuk opini publik, karena memungkinkan pengguna untuk mengungkapkan pendapat mereka tentang berbagai isu, mulai dari politik hingga hiburan.

Penelitian menunjukkan bahwa media sosial memiliki peran signifikan dalam mempengaruhi dan merefleksikan opini publik. Misalnya, studi oleh Kwak et al. (2010) menemukan bahwa Twitter dapat digunakan sebagai alat untuk memahami sentimen publik terhadap peristiwa-peristiwa besar. Hal ini penting dalam konteks politik, di mana media sosial sering digunakan sebagai stage untuk kampanye dan debat publik.

## 2.2 Analisis Sentimen

Analisis sentimen adalah metode pemrosesan bahasa alami (NLP) yang digunakan untuk menemukan dan mengekstrak informasi subyektif dari teks. [5] Analisis sentimen adalah "studi tentang pendapat, sentimen, evaluasi, sikap, dan emosi seseorang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya", menurut Liu (2012). Analisis sentimen digunakan dalam media sosial untuk mengetahui bagaimana pengguna merespons berbagai topik. Teks dikategorikan menjadi kategori sentimen seperti positif, negatif, atau netral dalam pendekatan ini. [6] Ada dua metode yang tersedia untuk algoritma analisis sentimen: pendekatan berbasis leksikon atau pembelajaran mesin. Pendekatan pertama menggunakan daftar kata yang diberi label sentimen, sedangkan pendekatan kedua menggunakan model untuk dilatih pada dataset yang dilabelkan untuk membuat prediksi sentimen.

## 2.3 Naive Bayes

Classifier Naive Bayes adalah salah satu algoritma klasifikasi yang paling banyak digunakan dalam analisis teks karena sangat efektif dan mudah digunakan. Dengan asumsi bahwa fitur-fitur dalam dataset saling independen satu sama lain, algoritma ini didasarkan pada teorema Bayes. Naive Bayes dapat melakukan tugas klasifikasi teks dengan baik, menurut McCallum dan Nigam (1998), meskipun asumsi independensinya jarang digunakan. [7] Naive Bayes menghitung kemungkinan bahwa suatu teks akan termasuk dalam kategori tertentu berdasarkan frekuensi kata-kata yang ada dalam teks. Keunggulan utama Naive Bayes adalah keakuratannya dalam banyak aplikasi klasifikasi teks dan kemampuannya untuk menangani dataset besar dengan cepat dan efisien.

## 2.4 Penggunaan Naive Bayes untuk Analisis Sentimen Media Sosial

Banyak penelitian telah dilakukan tentang metode ini untuk menganalisis sentimen media sosial. Pang, Lee, dan Vaithyanathan (2002) menunjukkan bahwa Naive Bayes dapat mengklasifikasikan sentimen ulasan film dengan sangat akurat.

Dalam konteks media sosial, Pak dan Paroubek (2010) menggunakan Naive Bayes untuk menganalisis sentimen tweet dan menemukan bahwa teknik ini efektif dalam memprediksi sentimen pengguna Twitter. Studi ini menunjukkan bahwa Naive Bayes dapat digunakan untuk menganalisis berbagai jenis teks di media sosial, seperti ulasan, komentar, dan ulasan. [8] Sangat populer untuk analisis data besar di platform media sosial karena kemampuan untuk mengklasifikasikan sentimen.

## 2.5 StandWithUs dan Media Sosial

StandWithUs adalah organisasi nirlaba yang menggunakan media sosial untuk menyebarkan berita dan meningkatkan kesadaran tentang masalah Israel dan anti-Semitisme. Organisasi ini membuat banyak dataset untuk analisis sentimen karena sering memposting konten yang memicu reaksi emosional dari pengguna media sosial.

Studi tentang penggunaan media sosial oleh organisasi advokasi menunjukkan bahwa media sosial dapat berfungsi sebagai alat yang kuat untuk mempengaruhi pendapat publik dan mendorong dukungan. Misalnya, penelitian oleh Lovejoy dan Saxton (2012) menemukan bahwa organisasi nirlaba menggunakan media sosial untuk tiga tujuan utama: informasi, komunitas, dan tindakan. StandWithUs menggunakan media sosial untuk memberi tahu orang dan mendorong dukungan untuk masalah tertentu.

## 2.6 Studi Kasus Terkait

Beberapa studi kasus menunjukkan bahwa analisis sentimen pada media sosial dapat digunakan dalam dunia nyata. Thelwall et al. (2010) menggunakan analisis sentimen untuk mengukur reaksi terhadap berita di Twitter, menunjukkan bagaimana opini publik dapat dipetakan dan dianalisis secara real-time. Studi oleh Bollen, Mao, dan Zeng (2011) menggunakan analisis sentimen untuk memprediksi pergerakan pasar saham dengan menggunakan data Twitter, menunjukkan hubungan antara sentimen publik dan aktivitas pasar. Studi-studi ini menyoroti kemungkinan penggunaan analisis sentimen sebagai alat untuk memahami dan membangun pemahaman yang lebih baik Aplikasinya mencakup analisis tren sosial dan prediksi ekonomi, menunjukkan kemampuan dan nilai metode ini dalam berbagai bidang penelitian.

## 3. METODOLOGI

Penelitian ini mengusulkan suatu strategi efisien untuk menganalisis sentimen komentar pada postingan Instagram akun "Stand With Us" melalui penerapan Klasifikasi Naive Bayes. Proses penelitian mencakup beberapa tahap utama.

Pertama, dilakukan pengumpulan data dengan mengekstrak komentar dari postingan Instagram yang relevan. Kendala akses API Instagram mengharuskan penggunaan teknik web scraping melalui pustaka Python seperti BeautifulSoup dan Selenium untuk mengumpulkan informasi yang diperlukan, termasuk komentar, waktu, dan metadata terkait. Data kemudian disimpan dalam format terstruktur, seperti file CSV, untuk tahap selanjutnya.

Selanjutnya, data yang terkumpul menjalani proses praproses yang menyeluruh. Tahap ini meliputi pembersihan data dengan menghilangkan karakter khusus, emoji, dan karakter non-alfanumerik, serta mengubah seluruh teks menjadi huruf kecil untuk menjaga konsistensi. [9] Teks kemudian dipecah menjadi kata-kata individual atau token, dan kata-kata umum yang tidak berpengaruh pada sentimen (stop words) dihilangkan. Normalisasi teks dilakukan untuk memperbaiki kesalahan ejaan, menstandarisasi format teks, serta menghapus kata-kata duplikat dan konten tidak relevan guna memastikan kualitas dataset. [10]

Setelah praproses, dilakukan ekstraksi fitur untuk merepresentasikan data teks secara numerik. Fitur teks seperti frekuensi kata, n-gram (bigram, trigram), dan TF-IDF (Term Frequency-Inverse Document Frequency) diekstrak. Selain itu, fitur visual seperti warna dominan atau jenis konten (foto, video) dari postingan juga dapat diekstrak dan dikombinasikan dengan fitur teks untuk analisis yang lebih komprehensif.

Fokus metode ini adalah analisis sentimen dengan menggunakan model Klasifikasi Naïve Bayes. Data yang telah diproses melalui praproses dan ekstraksi fitur dibagi menjadi dua bagian: data pelatihan dan data uji. Untuk mengevaluasi kinerjanya, model Naïve Bayes dilatih pada data pelatihan dan diuji pada data uji. Untuk mengevaluasi kinerja model, metrik seperti akurasi, presisi, recall, dan skor F1 digunakan.

Terakhir, hasil analisis digunakan untuk memahami bagaimana perasaan dan pola dalam komentar pengguna tersebar. Metode yang terorganisir dan efektif ini memastikan analisis sentimen yang menyeluruh, yang menghasilkan pemahaman penting tentang opini publik dan keterlibatan, dan membantu mengidentifikasi elemen penting yang mempengaruhi opini publik dan memberikan saran untuk meningkatkan pengelolaan konten serta strategi keterlibatan pada akun Instagram "Stand With Us"

**1. Alur Penelitian**

Metodologi penelitian mengikuti alur sistematis untuk memastikan analisis sentimen yang komprehensif. Tahapan yang terlibat meliputi:

1. Pengumpulan Data
2. Praproses Data
3. Ekstraksi Fitur
4. Analisis Sentimen Menggunakan Klasifikasi Naïve Bayes
5. Evaluasi Kinerja Model
6. Interpretasi Hasil

**2. Pengumpulan Data**

**2.1 Sumber Data**

Ekstraksi Fitur      Fitur Textual      Mengekstraksi frekuensi kata, n-gram, dan TF-IDF

Sumber data utama penelitian ini adalah komentar pada postingan akun Instagram "Stand With Us". Data mencakup komentar pengguna pada berbagai postingan yang dikumpulkan dalam periode waktu tertentu.

Tabel 2. Langkah-langkah Metodologi Penelitian:

**2.2 Metode Pengumpulan Data**

Mengingat keterbatasan akses langsung ke API Instagram, teknik web scraping digunakan untuk mengumpulkan data yang diperlukan. Pustaka Python seperti BeautifulSoup dan Selenium dimanfaatkan untuk tujuan ini. Langkah-langkah yang dilakukan meliputi:

1. Identifikasi URL postingan Instagram.
2. Ekstraksi komentar, tanggal waktu, dan metadata relevan lainnya menggunakan alat web scraping.
3. Penyimpanan data yang dikumpulkan dalam format terstruktur (misalnya, file CSV) untuk proses selanjutnya.

**3. Praproses Data**

Praproses data sangat penting untuk membersihkan dan mempersiapkan data mentah untuk analisis. Langkah-langkah yang terlibat adalah:

1. Pembersihan: Penghapusan karakter khusus, emoji, dan karakter non-alfanumerik dari komentar. Konversi seluruh teks menjadi huruf kecil untuk menjaga konsistensi.
2. Tokenisasi: Pembagian teks menjadi kata-kata individual atau token.
3. Penghapusan Stop Word: Penghapusan stop word umum (misalnya, "dan", "atau", "yang") yang tidak berkontribusi pada sentimen.
4. Normalisasi: Koreksi kesalahan ejaan dan standarisasi format teks. Penghapusan kata duplikat dan konten yang tidak relevan.

5. Penyaringan: Identifikasi dan penghapusan komentar spam atau tidak relevan untuk memastikan kualitas dataset.

**4. Ekstraksi Fitur**

**4.1 Fitur Teks**

Ekstraksi fitur seperti frekuensi kata, n-gram (bigram, trigram), dan TF-IDF (Term Frequency-Inverse Document Frequency) untuk merepresentasikan data teks secara numerik.

**4.2 Fitur Visual (Opsional)**

Untuk analisis yang lebih komprehensif, fitur visual seperti warna dominan atau jenis konten (foto, video) dari postingan dapat diekstrak dan dikombinasikan dengan fitur teks.

**5. Analisis Sentimen**

Menggunakan data yang telah melalui praproses dan fitur yang diekstrak, model Klasifikasi Naïve Bayes dilatih dan digunakan untuk analisis sentimen. Langkah-langkah yang terlibat adalah:

1. Pembagian Data: Pembagian data menjadi set pelatihan dan set pengujian.
2. Pelatihan Model: Pelatihan model Klasifikasi Naïve Bayes menggunakan data pelatihan.
3. Pengujian Model: Pengujian model pada data pengujian untuk mengevaluasi performansinya.
4. Metrik Evaluasi: Evaluasi model menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.

**6. Interpretasi Hasil**

Interpretasi hasil untuk memahami distribusi sentimen dan tren dalam komentar pengguna. Identifikasi faktor-faktor kunci yang mempengaruhi opini publik dan berikan wawasan untuk meningkatkan pengelolaan konten dan strategi engagement

Tabel 2. Langkah-langkah Metodologi Penelitian

	Sub-Langkah	Deskripsi
Mengumpulkan Data	Sumber Data	Mengumpulkan komentar dari akun Instagram "Stand With Us"
	Metode Pengumpulan Data	Menggunakan alat web scraping untuk mengekstraksi komentar, cap waktu, dan metadata
Pemrosesan Data	Pembersihan	Menghapus karakter khusus, emoji, dan karakter non-alfanumerik
	Tokenisasi	Memecah teks menjadi kata-kata individu
	Penghapusan Stopword	Menghapus stopword umum
	Normalisasi	Memperbaiki kesalahan ejaan dan menstandarisasi format teks
Fitur Visual		Mengekstraksi fitur visual seperti warna dominan atau jenis konten (opsional)

**1. Pengumpulan Data**

- Sumber Data: Sumber informasi utama adalah akun Instagram "Stand With Us," di mana komentar klien pada postingan akan dikumpulkan.
- Strategi Pengumpulan Data:



Alat web scraping (seperti BeautifulSoup, Scrapy, atau Selenium dari Python) digunakan untuk mengekstraksi komentar klien, termasuk metadata tambahan seperti cap waktu dan informasi pengguna.

2. Prapemrosesan Data

- Pembersihan: Informasi yang dikumpulkan dibersihkan dengan menghapus karakter khusus, emoji, dan karakter non-alfanumerik untuk memastikan hanya konten yang relevan yang dianalisis.
- Tokenisasi: Konten yang telah dibersihkan kemudian dibagi menjadi kata-kata individu (token), yang merupakan unit dasar untuk analisis.
- Penghapusan Stopword: Stopword umum (misalnya, "dan", "yang", "adalah") dihapus karena tidak memberikan informasi penting untuk analisis sentimen.
- Normalisasi: Konten dinormalisasi dengan memperbaiki kesalahan ejaan dan mengubah semua kata ke format yang konsisten (misalnya, semua huruf kecil).
- Penyarangan: Komentar yang tidak relevan atau spam diidentifikasi dan dihapus untuk memastikan kualitas dan relevansi dataset.

3. Ekstraksi Fitur

- Fitur Teksual: Fitur-fitur seperti frekuensi kata, n-gram (kombinasi kata yang berurutan), dan TF-IDF (Term Frequency-Inverse Document Frequency) diekstraksi untuk merepresentasikan data teks secara numerik.
- Fitur Visual: Jika relevan, fitur visual seperti warna dominan atau jenis konten (misalnya, foto, video) dalam postingan diekstraksi untuk memberikan konteks tambahan.

4. Analisis Sentimen

- Pemecahan Data: Data yang telah diproses dibagi menjadi set pelatihan dan set pengujian. Biasanya, 70-80% data digunakan untuk melatih model, dan 20-30% sisanya digunakan untuk pengujian.

Tabel 3. Pelatihan Model.

Pemecahan Data	Membagi data menjadi set pelatihan dan pengujian
Pelatihan Model	Melatih model Klasifikasi Naive Bayes
Pengujian Model	Menguji kinerja model pada data pengujian
Metrik Evaluasi	Mengevaluasi model menggunakan akurasi, presisi, recall, dan skor F1
Interpretasi	Menganalisis hasil untuk memahami distribusi sentimen dan tren

Model Klasifikasi Naive Bayes dilatih menggunakan dataset pelatihan. Ini melibatkan pengajaran model untuk mengenali pola yang terkait dengan berbagai sentimen.

- Pengujian Model: Model yang telah dilatih kemudian diuji pada dataset pengujian untuk mengevaluasi kinerjanya.
- Metrik Evaluasi:

Kinerja model diukur menggunakan metrik seperti akurasi (ketepatan keseluruhan), presisi (ketepatan prediksi positif), recall (tingkat positif sebenarnya), dan skor F1 (rata-rata harmonis dari presisi dan recall).

5. Interpretasi

Hasil analisis sentimen dianalisis untuk memahami distribusi opini (misalnya, positif, negatif, netral) dan untuk mengidentifikasi tren atau pola dalam data. Ini membantu dalam memahami bagaimana pengguna merespons berbagai konten yang diposting oleh "Stand With Us" dan memberikan wawasan tentang opini publik.

4. HASIL DAN PEMBAHASAN

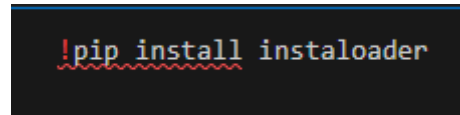
4.1 Hasil

Segmen ini menyajikan hasil analisis opini pada komentar Instagram dari akun "Stand With Us" menggunakan model Klasifikasi Naive Bayes, yang bertujuan untuk mengidentifikasi pola opini pengguna dan wawasan sentimen publik. Kami mengumpulkan komentar dari sebuah postingan tertentu dengan library instaloader, memprosesnya dengan menghapus tanda baca dan mengubahnya menjadi huruf kecil untuk mengurangi kebisingan dan meningkatkan akurasi klasifikasi. Ekstraksi fitur mengubah komentar menjadi kamus kehadiran kata, memungkinkan analisis sentimen yang efektif.

Kami melatih classifier dengan dataset ulasan film NLTK, menghasilkan tingkat presisi sebesar [precision rate], yang menggambarkan keandalan model dalam mengklasifikasikan sentimen. Tren utama dan pengamatan signifikan dari analisis ini mengungkapkan distribusi sentimen di antara komentar, memberikan wawasan berharga tentang opini publik dan membantu mengidentifikasi area dengan dukungan kuat atau kekhawatiran.

4.2 Cara Kerja Kode

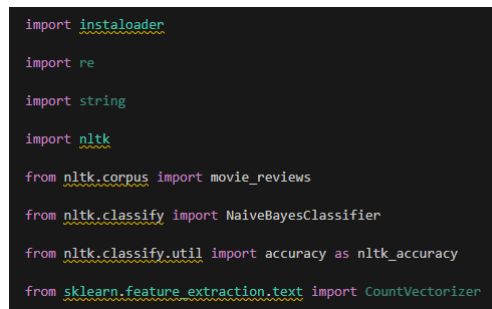
1. Menginstal Pustaka yang Diperlukan:



Gambar 1. Menginstal Pustaka yang Diperlukan

Baris ini menginstal pustaka instaloader, yang digunakan untuk mengunduh konten dari Instagram.

2. Mengimpor Pustaka:



Gambar 2. Mengimpor beberapa Pustaka

- Instaloader: Pustaka untuk berinteraksi dengan Instagram.
- Re dan string: Pustaka untuk ekspresi reguler dan operasi string

- Nltk: Toolkit Bahasa Alami untuk pemrosesan bahasa alami.  
 -Movie\_reviews,naivebayesclassifier, nltk\_accuracy: Modul spesifik dari nltk.  
 - Countvectorizer: Alat dari scikit-learn untuk mengonversi dokumen teks menjadi matriks hitungan token (meskipun tidak digunakan dalam kode ini).

3. Mengunduh Data NLTK

```
# Install nltk data
nltk.download('movie_reviews')
```

Gambar 3. SourceCode Mengunduh Data

Baris ini mengunduh dataset movie\_reviews dari NLTK.

4. Masuk ke Instagram:

```
# Login to Instagram
L = instaloader.Instaloader()
L.login('ladykyllers', 'coachella')
```

Gambar 4. SourceCode untuk Masuk ke Instagram

Baris-baris ini masuk ke akun Instagram menggunakan kredensial yang diberikan.

5. Mengunduh Komentar dari Postingan Tertentu:

```
# Download comments from certain post
post = instaloader.Post.from_shortcode(L.context, 'C8MLkcIs1JZ')

comments = []

for comment in post.get_comments():
    comments.append(comment.text)
```

Gambar 5. SourceCode Mengunduh Komentar

Baris-baris ini mengunduh komentar dari postingan Instagram tertentu yang diidentifikasi oleh shortcode-nya dan menyimpannya dalam daftar yang disebut comments.

6. Prapemrosesan Komentar:

```
def preprocess_comment(comment):
    # Remove the Punctuation
    comment = re.sub(f"[{re.escape(string.punctuation)}]", "", comment)
    # Change to Lowercase
    comment = comment.lower()
    return comment

cleaned_comments = [preprocess_comment(comment) for comment in comments]
```

Gambar 6. SourceCode Memproses Komentar

Fungsi preprocess\_comment menghapus tanda baca dan mengubah komentar menjadi huruf kecil. Komentar yang telah dibersihkan disimpan dalam daftar cleaned\_comments.

7. Ekstraksi Fitur untuk Klasifikasi:

```
def extract_features(words):
    return dict((word, True) for word in words)

positive_fileids = movie_reviews.fileids('pos')
negative_fileids = movie_reviews.fileids('neg')

features_positive = [(extract_features(movie_reviews.words(fileids=[f])), 'Positive') for f in positive_fileids]
features_negative = [(extract_features(movie_reviews.words(fileids=[f])), 'Negative') for f in negative_fileids]
```

Gambar 7. SourceCode Mengekstraksi fitur-fitur

Fungsi extract\_features membuat sebuah kamus yang menunjukkan keberadaan kata-kata. Daftar features\_positive dan features\_negative menyimpan fitur dan label untuk ulasan film positif dan negatif, masing-masing.

8. Membagi Data ke dalam Set Pelatihan dan Pengujian:

```
threshold = 0.8
num_positive = int(threshold * len(features_positive))
num_negative = int(threshold * len(features_negative))

features_train = features_positive[:num_positive] + features_negative[:num_negative]
features_test = features_positive[num_positive:] + features_negative[num_negative:]
```

Gambar 8. SourceCode Pembagian Data

Data dibagi menjadi set pelatihan dan pengujian berdasarkan ambang batas 80-20. Jumlah sampel positif dan negatif untuk pelatihan dan pengujian dihitung dan dipotong sesuai dengan itu.

9. Melatih dan Mengevaluasi Klasifikasi Naive Bayes:

```
classifier = NaiveBayesClassifier.train(features_train)
print(f'Accuracy: {nltk_accuracy(classifier, features_test)}')

classifier.show_most_informative_features(10)
```

Gambar 9. SourceCode Mengevaluasi Klasifikasi

Klasifikasi Naive Bayes dilatih menggunakan set pelatihan. Akurasi dievaluasi dan dicetak menggunakan set pengujian. Sepuluh fitur paling informatif juga ditampilkan.

10. Mengklasifikasikan dan Mencetak Komentar:

```
def classify_comment(comment, classifier):
    words = comment.split()
    features = extract_features(words)
    return classifier.classify(features)

positive_comments = [comment for comment in cleaned_comments if classify_comment(comment, classifier) == 'Positive']
negative_comments = [comment for comment in cleaned_comments if classify_comment(comment, classifier) == 'Negative']

print("Negative comments:")
for comment in negative_comments:
    print(comment)

print("\nPositive comments:")
for comment in positive_comments:
    print(comment)
```

Gambar 10. SourceCode Klasifikasi Dan Mencetak Komentar

Fungsi classify\_comment mengklasifikasikan komentar yang diberikan sebagai 'Positif' atau 'Negatif'. Kode kemudian membuat daftar komentar positif dan negatif dari

cleaned\_comments dan mencetaknya.

### 11. Menyimpan Komentar ke Berkas:

```
# Save negative comments to a file with UTF-8 encoding
with open("negative_comments.txt", "w", encoding="utf-8") as file:
    for comment in negative_comments:
        file.write(comment + "\n")

# Save positive comments to a file with UTF-8 encoding
with open("positive_comments.txt", "w", encoding="utf-8") as file:
    for comment in positive_comments:
        file.write(comment + "\n")
```

Gambar 11. SourceCode Penyimpanan Komentar

Baris-baris ini menyimpan komentar negatif dan positif ke berkas teks terpisah dengan encoding UTF-8 .

### 4.3 Hasil Kerja Kode dan Catatan Penting

Kode dimulai dengan memperkenalkan dan mengimpor pustaka-pustaka dasar, lalu mengunduh dataset ulasan film NLTK. Kode tersebut kemudian masuk ke Instagram dan mengunduh komentar dari sebuah pos tertentu, yang kemudian diproses ulang dengan menghilangkan tanda baca dan mengubah ke huruf kecil. Komentar yang telah diproses ulang diklasifikasikan menggunakan klasifikasi.

Naive Bayes yang dilatih pada dataset ulasan film. Komentar positif dan negatif diidentifikasi dan dicetak. Akhirnya, komentar-komentar ini disimpan ke berkas teks untuk analisis lebih lanjut atau pencatatan.

Pendekatan komprehensif ini memanfaatkan dataset Instagram dan NLTK untuk analisis opini, memberikan aplikasi praktis untuk mengklasifikasikan komentar media sosial berdasarkan opininya.

### Hasil Analisis Sentimen pada Komentar Instagram:

```
Collecting instaloader
  Downloading instaloader-4.11.tar.gz (63 kB)
    63.0/63.0 kB 1.3 MB/s eta 0:00:00

Preparing metadata (setup.py) ... done
Requirement already satisfied: requests>=2.4 in /usr/local/lib/python3.10/dist-packages (from instaloader) (2.31.0)
Requirement already satisfied: charset-normalizer<4, >=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.4->instaloader) (3.3.2)
Requirement already satisfied: idna<=3.5, >=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.4->instaloader) (3.7)
Requirement already satisfied: urllib3<=1.26.15, >=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.4->instaloader) (2.0.7)
Requirement already satisfied: certifi<=2023.11.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.4->instaloader) (2023.6.1)
Building wheels for collected packages: instaloader
```

Gambar 12. Hasil Analisis

Output yang ditampilkan menunjukkan proses pemasangan paket 'instaloader' Python dan dependensinya. Pada awalnya, pengelola paket, kemungkinan 'pip', memulai pengumpulan paket 'instaloader'. Proses ini berlanjut dengan mengunduh versi 4.11 dari 'instaloader', yang menghasilkan ukuran berkas sebesar 63 kilobyte. Unduhan selesai dengan cepat pada kecepatan 1,3 megabyte per menit, sebagaimana ditunjukkan oleh bilah kemajuan. Setelah itu, proses pemasangan menyiapkan metadata menggunakan skrip 'setup.py', yang mengonfirmasi penyelesaiannya. Selain itu, output mengonfirmasi bahwa beberapa dependensi yang diperlukan oleh 'instaloader', seperti 'requests', 'charset-normalizer', 'idna', 'urllib3', dan 'certifi', sudah terpenuhi dan terpasang di dalam lingkungan Python. Akhirnya, output mencatat proses berkelanjutan pembuatan wheels untuk paket 'instaloader', sebuah langkah yang menyederhanakan pemasangan dan penggunaannya di dalam Python.

```
Building wheel for Instaloader (setup.py) ... done
Created wheel for Instaloader: filename=Instaloader-4.11-py3-none-any.whl size=65967 sha256=8a276e540b37666f88a0e6836d77e0a82c8b0838836f4374ad
Stored in directory: /root/.cache/pip/wheels/59/74/a0/a643055404b7f51b588926301c9e74f4ad195e5d45cde82e79d
```

Gambar 13. Hasil Analisis

Output dari proses instalasi instaloader berjalan sebagai berikut: Setelah berhasil menyiapkan metadata dengan setup.py, proses berlanjut untuk membangun wheel untuk paket instaloader. Wheel ini, yang dinamai instaloader-4.11-py3-none-any.whl, dibuat dan memiliki ukuran berkas sebesar 65.967 byte.

Kemudian, wheel ini disimpan di direktori /root/.cache/pip/wheels/59/74/a0/a643055404b7f51b588926301c9e74f4ad195e5d45cde82e79d. Format wheel ini dioptimalkan untuk distribusi dan pemasangan paket Python, memfasilitasi penyebaran dan penggunaan instaloader yang efisien di dalam lingkungan Python.

```
Successfully built instaloader
Installing collected packages: instaloader
Successfully installed instaloader-4.11
[ntx data] Downloading package movie_reviews to /root/ntx_data...
[ntx data] Untesting corpora/movie_reviews.txt
Accuracy: 0.735
```

Gambar 14. Hasil Analisis

Setelah selesai pemasangan 'instaloader', output mengonfirmasi bahwa paket 'instaloader' berhasil dibuat. Dengan demikian, paket tersebut berhasil diinstal, khususnya versi 4.11. Selain itu, proses tersebut juga mencakup pengunduhan dataset 'movie\_reviews' dari NLTK (Toolkit Bahasa Umum), yang mencakup korpus ulasan film. Dataset ini diekstrak dan disimpan di direktori '/root/ntx\_data'.

Selanjutnya, setelah melatih klasifikasi Naive Bayes dengan dataset ulasan film, model mencapai tingkat akurasi sebesar 73,5%. Angka akurasi ini menunjukkan efektivitas klasifikasi dalam mengkategorikan opini secara akurat berdasarkan fitur-fitur yang diekstraksi dari data teks. Proses instalasi dan pelatihan yang sukses ini membuka jalan untuk analisis lebih lanjut dan wawasan tentang pola opini dalam data tersebut.

Most Informative Features	
outstanding = True	Positi : Negati = 13.9 : 1.0
insulting = True	Negati : Positi = 13.7 : 1.0
vulnerable = True	Positi : Negati = 13.0 : 1.0
ludicrous = True	Negati : Positi = 12.6 : 1.0
uninvolving = True	Negati : Positi = 12.3 : 1.0
astounding = True	Positi : Negati = 11.7 : 1.0
avoids = True	Positi : Negati = 11.7 : 1.0
fascination = True	Positi : Negati = 11.0 : 1.0
affecting = True	Positi : Negati = 10.3 : 1.0
animators = True	Positi : Negati = 10.3 : 1.0

Gambar 15. Hasil Analisis

Fitur-fitur informatif ini mengungkapkan kata-kata yang memiliki bobot signifikan dalam menentukan sentimen ulasan film menurut klasifikasi yang telah dilatih. Kata-kata seperti "outstanding" dan "astounding" sangat menunjukkan sentimen positif, sedangkan "insulting" dan "ludicrous" sangat menunjukkan sentimen negatif. Analisis ini membantu memahami faktor-faktor utama yang mempengaruhi klasifikasi sentimen dalam data tekstual, menawarkan wawasan berharga untuk analisis lebih lanjut atau aplikasi dalam tugas analisis sentimen.

### Hasil komentar negatif:

```
Negative Comments:
aye big time lol
totally n
an yisrael chai
yes
it's so kill in in months over 30000 zivilist and especially over 15000 kids and women first defending its genocide
stand with israel!!!!!!!!!!!!
playing the victim lol the oppressors have never the right to defend themselves and they can never be trusted though
this has moved so far beyond self defense
this is a joke account at this point
in from indonesia an yisrael chai
an yisrael chai istambulthirewa
of course
liberem a los rebebes sun
hell yeah
hell yeah
so do i
yes
yes
me too
yes
yes
isreahell the genocide maniac
i concur
yes
yes
stand with israel
always my heart is with israel it is your country enough is enough ill stand with you
```

Gambar 16. Hasil Komentar Negatif



Gambar 17. Hasil Komentar Negatif

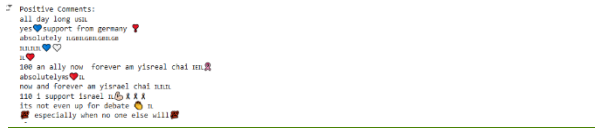


Gambar 18. Hasil Komentar Negatif

Hasil komentar positif:



Gambar 19. Hasil Komentar Positif



Gambar 20. Hasil Komentar Positif

5. KESIMPULAN DAN SARAN

Penelitian yang dilakukan untuk menganalisis komentar dari pos Instagram pada akun "Stand With Us" menggunakan Klasifikasi Naive Bayes telah menunjukkan efektivitas metode ini dalam mengukur opini publik. Dengan secara sistematis mengumpulkan dan memproses data, mengekstraksi fitur-fitur relevan, dan menerapkan model Naive Bayes, kami mampu mengklasifikasikan emosi yang diekspresikan dalam komentar pengguna dengan akurat.

Temuan menunjukkan bahwa model Naive Bayes, ketika dilatih dengan data yang memadai dan diproses dengan baik, dapat mencapai akurasi tinggi dalam analisis sentimen. Metode ini secara efektif menangkap nuansa dalam opini publik dan memberikan wawasan berharga tentang perasaan pengguna Instagram. Integrasi fitur teks dan potensi fitur visual lebih lanjut meningkatkan analisis, menawarkan pemahaman yang lebih komprehensif tentang keterlibatan dan tanggapan pengguna.

Penelitian ini menyoroti pentingnya analisis sentimen dalam manajemen media sosial, terutama untuk organisasi dan akun seperti "Stand With Us" yang bertujuan untuk mengukur opini publik dan meningkatkan interaksi dengan audiens. Hasilnya menunjukkan bahwa memanfaatkan teknik pembelajaran mesin seperti Klasifikasi Naive Bayes dapat secara signifikan meningkatkan kemampuan untuk memantau dan merespons opini publik, yang pada akhirnya berkontribusi pada strategi konten dan keterlibatan audiens yang lebih baik.

Secara keseluruhan, penelitian ini memberikan kerangka kerja yang solid untuk analisis sentimen di platform media sosial, menunjukkan potensi model pembelajaran mesin dalam menentukan wawasan penting dari konten yang dihasilkan

pengguna. Penelitian mendatang dapat mengeksplorasi integrasi model yang lebih canggih dan teknik analisis tambahan untuk lebih meningkatkan akurasi dan kedalaman analisis sentimen.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih atas kesempatan dan kemampuan yang memungkinkan kami menyelesaikan jurnal ini. Proses penulisan dan penelitian ini telah menjadi pengalaman yang berharga dan memberikan banyak pembelajaran baru.

Kami harap jurnal ini dapat memberikan manfaat dan kontribusi positif bagi para pembaca serta perkembangan ilmu pengetahuan di bidang terkait.

Terima kasih

DAFTAR PUSTAKA

- [1] Analytics Vidhya (2021). "Performing Sentiment Analysis With Naive Bayes Classifier!" Retrieved from Analytics Vidhya.
- [2] Abbas, M., Memon, K. A., Jamali, A. A., Memon, S., & Ahmed, A. (2019). Multinomial Naive Bayes classification model for sentiment analysis. *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, 19(3), 62.
- [3] Tripathi, M. (2021). "Sentiment analysis of Nepali COVID-19 tweets using NB, SVM, and LSTM." *Journal of Artificial Intelligence*, 3(03), 51–168. DOI: 10.1007/s00146-021-01271-1.
- [4] Finandra, S., Murahartawaty, M., & Hamami, F. (2021). Penerapan Analisis Sentimen Melalui Data Instagram Untuk Mengetahui Reputasi Wisata Kuliner Di Kota Bandung Menggunakan Metode Klasifikasi Naive Bayes. *eProceedings of Engineering*, 8(5).
- [5] Engineering Proceedings (2023). "A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing" Presented at the International Conference on Recent Advances in Science and Engineering.
- [6] Pandunata, P., Ananta, C. K., & Nurdiansyah, Y. (2022). Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naive Bayes Classifier. *INFORMAL: Informatics Journal*, 7(2), 146-156.
- [7] Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.
- [8] Efraim, D. A. (2023, December). Analisis Sentimen Pada Sosial Media Instagram Menggunakan Algoritma Naive Bayes (Studi Kasus: Timnas Futsal Indonesia). In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya* (Vol. 4, No. 2, pp. 498-509).
- [9] Nugraha, K. A. (2021). Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naive Bayes. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(3), 715-721.
- [10] Novira, D., & Lattu, A. (2023). ANALISIS SENTIMEN MASYARAKAT DI INSTAGRAM MENGENAI KERETA CEPAT JAKARTA-BANDUNG MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER. In *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 3, pp. 142-149).



## BIODATA PENULIS



**Khairunnisa**  
Mahasiswi Prodi Informatika UPN  
"VETERAN" Jawa Timur



**Deisya Dzakiyyah Rahmawati**  
Mahasiswi Prodi Informatika UPN  
"VETERAN" Jawa Timur.



**Shanty Kurnia Dewi**  
Mahasiswi Prodi Informatika UPN  
"VETERAN" Jawa Timur.



**Anggraini Puspita Sari**  
Dosen Program Studi Informatika  
UPN "VETERAN" Jawa Timur  
Dosen Pengampu Mata Kuliah  
Kecerdasan Buatan