

## Analisis Eksperimental Kinerja Transformers, VADER, dan Naive Bayes dalam Analisis Sentimen Teks Bahasa Indonesia (Studi Kasus Komentar Terkait Judi Online)

Sugiyono<sup>1</sup>, Agung Budi Susanto<sup>2</sup>, Sajarwo Anggai<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Kota, Tangerang Selatan

Email: <sup>1</sup>sugiyoku@gmail.com, <sup>2</sup>dosen02680@unpam.ac.id, <sup>3</sup>sajarwo@mail.com

### ABSTRACT

*Sentiment analysis is a subfield of Natural Language Processing (NLP) that focuses on detecting and classifying opinions expressed in textual data. In the digital social context, the increasing volume of public comments related to online gambling in Indonesia highlights the need to map public perception. This study aims to conduct an experimental analysis of the performance of three popular sentiment analysis approaches: VADER (Valence Aware Dictionary and sEntiment Reasoner), Naive Bayes, and Transformers-based models, specifically on Indonesian-language text. The dataset consists of public comments from social media and digital platforms containing keywords related to online gambling. The research process involves text preprocessing, data labeling, model training (for Naive Bayes and Transformers), and performance testing. Evaluation metrics include accuracy, precision, recall, and F1-score. The experimental results show that the Transformers model (using IndoBERT) achieves the highest performance in terms of accuracy and generalization ability, while VADER performs less optimally due to its limitations in understanding Indonesian linguistic context. Naive Bayes demonstrates moderate and consistent performance but lacks the capability to capture complex contextual meanings. These findings contribute to selecting appropriate sentiment analysis methods for non-English languages and support the development of more accurate public opinion detection systems in the future.*

**Keywords:** *sentiment analysis, VADER, Naive Bayes, Transformers, online gambling, Indonesian language*

### ABSTRAK

*Analisis sentimen merupakan salah satu cabang dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang berfokus pada pengenalan dan klasifikasi opini dalam teks. Dalam konteks sosial digital, meningkatnya komentar publik terkait fenomena judi online di Indonesia mendorong perlunya pemetaan persepsi masyarakat. Penelitian ini bertujuan untuk melakukan analisis eksperimental terhadap kinerja tiga pendekatan populer dalam analisis sentimen, yaitu VADER (Valence Aware Dictionary and sEntiment Reasoner), Naive Bayes, dan model berbasis Transformers terhadap teks berbahasa Indonesia. Data yang digunakan berupa komentar-komentar publik dari media sosial dan platform digital yang mengandung kata kunci terkait judi online. Proses penelitian mencakup tahap pra-pemrosesan teks, pelabelan data, pelatihan model (untuk Naive Bayes dan Transformers), serta pengujian akurasi. Evaluasi kinerja dilakukan dengan membandingkan metrik akurasi, presisi, recall, dan F1-score dari ketiga metode. Hasil penelitian menunjukkan bahwa model Transformers (menggunakan IndoBERT) memiliki performa paling tinggi dalam hal akurasi dan kemampuan generalisasi, sementara VADER cenderung kurang optimal karena keterbatasan dalam menangkap konteks bahasa Indonesia. Naive Bayes menempati posisi tengah dengan performa stabil namun kurang mampu mengenali makna kontekstual yang kompleks. Temuan ini memberikan kontribusi penting dalam pemilihan metode analisis sentimen yang tepat untuk bahasa non-Inggris serta menjadi dasar bagi pengembangan sistem deteksi opini publik yang lebih akurat di masa depan.*

**Kata kunci:** *analisis sentimen, VADER, Naive Bayes, Transformers, judi online, bahasa Indonesia*

## 1. PENDAHULUAN

Kemajuan teknologi digital telah menjadikan media sosial sebagai ruang utama bagi masyarakat dalam menyampaikan opini dan respons terhadap berbagai isu sosial. Salah satu isu yang menonjol dalam ruang digital Indonesia adalah praktik judi online, yaitu aktivitas mempertaruhkan uang secara daring melalui permainan seperti *poker*, *slot*, atau taruhan olahraga. Fenomena ini tidak hanya meluas dengan cepat, tetapi juga menimbulkan berbagai dampak serius dari sisi sosial, ekonomi, hukum, hingga pendidikan, khususnya pada kelompok remaja dan pelajar [3], [9]. Di Indonesia, judi online merupakan aktivitas ilegal yang dilarang secara hukum dan bertentangan dengan nilai-nilai moral serta tatanan sosial masyarakat. Namun demikian, kemudahan akses dan anonimitas dalam dunia digital mendorong tingginya partisipasi masyarakat dalam praktik ini. Diskursus publik di media sosial mengenai isu ini mencerminkan spektrum sentimen yang luas—mulai dari yang positif, netral, hingga negatif. Analisis terhadap opini-opini ini penting untuk memahami persepsi masyarakat secara real-time.

Salah satu metode komputasional yang relevan untuk menangani persoalan ini adalah analisis sentimen, sebuah cabang dari *Natural Language Processing* (NLP) yang berfokus pada identifikasi dan klasifikasi opini berdasarkan isi teks [1], [6]. Dengan analisis sentimen, kita dapat memetakan pola emosi masyarakat terhadap isu tertentu secara sistematis dan efisien. Dalam konteks ini, beberapa pendekatan analisis sentimen digunakan, yakni Naive Bayes, Transformers, dan VADER. Naive Bayes merupakan algoritma statistik klasik yang dikenal karena kesederhanaannya dan efisiensinya dalam klasifikasi teks [2], [8], [10]. Sementara itu, pendekatan berbasis Transformers—terutama model seperti *Bidirectional Encoder Representations from Transformers* (BERT)—memiliki keunggulan dalam memahami konteks dan relasi antar kata melalui mekanisme *attention* yang kompleks [11]. Adapun VADER (*Valence Aware Dictionary and sEntiment Reasoner*) adalah metode berbasis leksikon yang dirancang untuk teks informal seperti media sosial, dan mampu menangkap intensitas emosional dari kata atau frasa [1], [5].

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa dari ketiga pendekatan tersebut dalam klasifikasi sentimen terhadap komentar publik berbahasa Indonesia yang berkaitan dengan isu judi online. Dengan menggunakan pendekatan eksperimental dan mengukur metrik kinerja seperti *accuracy*, *precision*,

*recall*, dan *F1-score*, hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pemilihan metode analisis sentimen yang paling efektif dan akurat, khususnya untuk digunakan dalam konteks sosial yang serupa di masa mendatang.

## 2. METODE

Metode merupakan pendukung penelitian baik dari segi pengumpulan data, tempat penelitian, jenis dan sumber penelitian, tahapan penelitian, metode pengembangan, dan lainnya yang diperlukan dalam metode.

### 2.1. Objek Data Penelitian

Objek dalam penelitian ini adalah komentar publik berbahasa Indonesia yang membahas isu **judi online**, dikumpulkan dari media sosial x. Teknik pengumpulan data dilakukan melalui *web scraping* selama periode Januari hingga November 2024. Pada awalnya data yang diperoleh sebanyak 7.655 komentar. Setelah dilakukan proses seleksi dan pembersihan, diperoleh sebanyak **2.500 komentar** yang telah dilabeli secara manual ke dalam tiga kategori sentimen, yaitu **positif**, **negatif**, dan **netral**, berdasarkan interpretasi kontekstual terhadap isi komentar.

Dalam menanggapi fenomena judi online, penting bagi kita untuk memberikan komentar yang objektif, baik dalam perspektif positif, negatif, maupun netral. Penilaian ini dapat dilakukan dengan mengacu pada sejumlah faktor, seperti dampak ekonomi, aspek hukum, serta implikasi sosial dan budaya yang ditimbulkan. Berikut adalah acuan yang digunakan dalam menentukan apakah suatu pernyataan atau pandangan terhadap judi online bersifat positif, negatif, atau netral:

1. Positif
  - 1) Komentar yang secara jelas menentang atau mengkritik judi online.
  - 2) Komentar yang mengedukasi tentang bahaya judi online.
  - 3) Komentar yang mendorong masyarakat untuk menjauhi judi online.
2. Negatif
  - 1) Komentar yang mempromosikan judi online, termasuk yang menyertakan link pendaftaran atau kata-kata ajakan seperti "GACOR", "JACKPOT", "BONUS NEW MEMBER".
  - 2) Komentar yang berusaha meyakinkan orang untuk bermain judi online.
3. Netral

- 1) Komentar yang hanya berupa berita atau laporan fakta tanpa ada opini (misalnya data statistik tanpa menyatakan dukungan atau penolakan).
- 2) Komentar yang membahas judi online tetapi tidak menunjukkan sikap yang jelas mendukung atau menolak.

Berdasarkan acuan diatas, table yang telah disaring kemudian diberi komentar positif, negatif, ataupun netral, serta diberi alasan atas komentar yang diberikan. File selengkapnya dari data yang telah diberi komentar tersimpan dalam file Bernama Master.xlsx. Berikut ini adalah beberapa baris contoh dari tabel yang telah diberi komentar positif, negatif, atau netral.

Tabel 2.1 Table yang berisi beberapa baris contoh yang telah diberi label positif, negative, ataupun netral.

Full_text	Manual_sentiment	Alasan
INFO SLOT ONLINE GACOR HARIINI LINK GACOR DI BIO <a href="https://t.co/BWe5FxM3D0">https://t.co/BWe5FxM3D0</a>	Negatif	Mengandung ajakan untuk bermain judi online dan membagikan tautan pendaftaran.
🎰 INFO SLOT GACOR TERBARU Selasa Pagi Ini 🎰 🎰 AGEN SLOT ONLINE TERPERCAYA 🎰 BONUS 50% KHUSUS NEW MEMBER LINK GACOR : <a href="https://t.co/J8tvwIQV73">#slotgacor</a> <a href="https://t.co/q17K90MfhO">https://t.co/q17K90MfhO</a>	Negatif	Mempromosikan judi online dengan iming-iming bonus dan tautan pendaftaran.
Serbubet slot online terpercaya daftar slot idn daftar idn slot slot online 24jam #slotonlineterpercaya #agenjudislot2020 #judionlineterpercaya #depositpulsa #depositDANA Sore #AsalBukan02 IISTIMEWA JIPLAK Bawaslu All Out Fahri UU Ciptaker Reva Kalimantan Timur <a href="https://t.co/lb6BxKazcU">https://t.co/lb6BxKazcU</a>	Negatif	Menawarkan layanan judi online dengan berbagai metode deposit.

Rekomendasi situs online terpercaya terlengkap angka togel casino online slot game penghasil uang tercepat dan mudah jackpot dengan pola RTP tertinggi situs luar negeri dijamin <a href="https://t.co/Nw1Rl2j1vp">https://t.co/Nw1Rl2j1vp</a> <a href="https://t.co/3uOSCEFOuf">https://t.co/3uOSCEFOuf</a>	Negatif	Mengajak orang bermain judi online dengan klaim keuntungan tinggi.
INFO SLOT GACOR HARI INI! ID BARU PASTI CUAN - BONUS NEW MEMBER 50% - JOIN DAN MAINKAN SEKARANG LINK GACOR : <a href="https://t.co/Nw1Rl2j1vp">https://t.co/Nw1Rl2j1vp</a> <a href="https://t.co/v5uS5te31U">https://t.co/v5uS5te31U</a>	Negatif	Menggunakan kata-kata ajakan dan promosi bonus untuk menarik pemain baru.
INFO SLOT GACOR HARI INI ID BARU AUTO MAXWIN LINK GACOR DI BAWAH LINK GACOR : <a href="https://t.co/Nw1Rl2j1vp">https://t.co/Nw1Rl2j1vp</a> <a href="https://t.co/8c7bwXrfjz">https://t.co/8c7bwXrfjz</a>	Negatif	Mengandung promosi judi online dengan klaim kemenangan tinggi.
Ini salah satu korban judi online. Beliau dari golongan ekonomi lemah yang semakin melarat setelah kenal dengan judol. Ada berjuta-juta orang bernasib sama dengan pemuda ini. Dengan uang pas-pasan yang ia miliki, ia ikut judol berharap keberuntungan dan akhirnya... <a href="https://t.co/vcDpatnhcM">https://t.co/vcDpatnhcM</a>	Positif	Mengedukasi tentang dampak negatif judi online terhadap masyarakat ekonomi lemah.
Film *No More Bets* di Netflix layak ditonton, terutama mereka yang kecanduan judi online agar paham bagaimana para bandar menjebak dan mengibuli para pemainnya. Judol melahirkan kejahatan lanjutan seperti halnya	Positif	Mengedukasi masyarakat tentang bahaya judi online dan mendukung upaya pemberantasannya.

narkoba. Kita sangat mengapresiasi tekad Pak Prabowo, Polri, dan BNN. <a href="https://t.co/Z7H9TEC5kB">https://t.co/Z7H9TEC5kB</a>		
100 Orang Dirawat di RSCM akibat Judi Online. Remaja dan Dewasa Muda Lebih Berisiko Kecanduan Judol. <a href="https://t.co/kjA6qPpH6a">https://t.co/kjA6qPpH6a</a>	Netral	Menyajikan berita tanpa opini, hanya melaporkan data dan fakta tentang dampak judi online.
100 Orang Dirawat di RSCM akibat Judi Online. Remaja dan Dewasa Muda Lebih Berisiko Kecanduan Judol. <a href="https://t.co/N5rtdq7MR1">https://t.co/N5rtdq7MR1</a>	Netral	Memberikan informasi faktual tanpa menunjukkan dukungan atau penolakan secara eksplisit.

## 2.2. Pra-pemrosesan Data

Data yang telah dikumpulkan kemudian melalui serangkaian tahap pra-pemrosesan guna meningkatkan akurasi model analisis sentimen. Tahapan tersebut meliputi:

1. Pembersihan teks (*cleaning*): Menghapus karakter yang tidak relevan seperti tanda baca, angka, simbol, dan karakter khusus.
2. *Case folding*: Mengubah seluruh huruf menjadi huruf kecil untuk menyamakan representasi kata.
3. Tokenisasi: Memecah kalimat menjadi unit-unit kata (token) sebagai dasar pemrosesan lanjutan.
4. Penghapusan *stopword* dan *stemming*: Menghapus kata-kata yang tidak bermakna penting (seperti "dan", "atau") dan mengekstraksi akar kata. Tahapan ini secara khusus diterapkan pada model *Naive Bayes* yang sangat bergantung pada representasi kata secara eksplisit.

Selanjutnya, teks yang telah diproses diubah menjadi bentuk representasi numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*), yang memungkinkan pemberian bobot terhadap kata-kata berdasarkan frekuensi kemunculan relatif dalam dokumen.

### 2.3. Model Algoritma Yang Digunakan

Penelitian ini menggunakan dan membandingkan tiga pendekatan analisis sentimen yang berbeda:

1. *Naive Bayes*: Algoritma klasifikasi berbasis probabilistik dan teorema Bayes, yang dikenal efisien untuk teks pendek, namun kurang dalam memahami konteks semantik yang kompleks.
2. *Transformers* (BERT): Menggunakan model pralatih *nlptown/bert-base-multilingual-uncased-sentiment*, yang mendukung berbagai bahasa termasuk Bahasa Indonesia, dan mengandalkan arsitektur *self-attention* untuk memahami konteks dua arah dalam teks secara mendalam.

*VADER (Valence Aware Dictionary and sEntiment Reasoner)*: Metode berbasis leksikon dan aturan (*rule-based*), dirancang untuk menganalisis sentimen dalam teks informal seperti komentar media sosial, serta mampu menangkap intensitas sentimen.

### 2.4. Evaluasi Model

Evaluasi kinerja ketiga model dilakukan dengan menggunakan empat metrik utama:

1. Akurasi: Proporsi prediksi yang benar terhadap total data.
2. Presisi: Kemampuan model dalam mengidentifikasi komentar dengan sentimen tertentu secara tepat.
3. *Recall*: Kemampuan model dalam menangkap semua komentar yang relevan untuk setiap kategori sentimen.
4. *F1-score*: Rata-rata harmonis antara presisi dan *recall*, digunakan untuk memberikan penilaian yang seimbang terutama dalam kondisi distribusi data yang tidak merata.

Dataset dibagi menjadi dua bagian dengan rasio 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Seluruh proses eksperimental dilakukan menggunakan bahasa pemrograman *Python*, dengan bantuan pustaka NLP seperti *scikit-learn*, NLTK, dan *Transformers* dari *Hugging Face*.

### 3. PEMBAHASAN

#### 3.1 Analisis Sentimen dengan *Naive Bayes*

Dalam analisis sentimen Bahasa Indonesia terkait komentar tentang judi online, penulis menggunakan dua pendekatan Naive Bayes. *Script* pertama tidak menggunakan pustaka NLTK, sedangkan *script* kedua menyertakan proses pra-pemrosesan teks menggunakan NLTK, seperti *stopword removal* dan *stemming*.

##### 3.1.1 *Naive Bayes* Tanpa NLTK

Hasil dari script pertama menunjukkan:

1. Akurasi data uji: 73,70%
2. Akurasi seluruh dataset: 93,36%

Laporan klasifikasi:

Tabel 3.1 laporan klasifikasi proses *Naïve Bayes* tanpa NLTK

Sentimen	Precision	Recall	F1-score	Support
Negatif	0.91	0.91	0.91	106
Netral	0.74	0.74	0.74	236
Positif	0.59	0.61	0.60	137
<b>Accuracy</b>			<b>0.74</b>	<b>479</b>
<b>Macro avg</b>	0.75	0.75	0.75	
<b>Weighted avg</b>	0.74	0.74	0.74	

##### 3.1.2 *Naive Bayes* dengan NLTK

Script kedua menggunakan teknik tambahan seperti *stopword removal* dan *stemming*.

1. Akurasi data uji: 75,37%
2. Akurasi seluruh dataset: 94,03%

Laporan klasifikasi:

Tabel 3.2 laporan klasifikasi *Naïve Bayes* dengan NLTK

Sentimen	Precision	Recall	F1-score	Support
Negatif	0.93	0.91	0.92	106
Netral	0.75	0.76	0.76	236
Positif	0.62	0.63	0.63	137

Sentimen	Precision	Recall	F1-score	Support
Accuracy			<b>0.75</b>	<b>479</b>
<b>Macro avg</b>	0.77	0.76	0.77	
<b>Weighted avg</b>	0.76	0.75	0.75	

### 3.1.3 Uraian Perbandingan *Script* Pertama vs *Script* Kedua

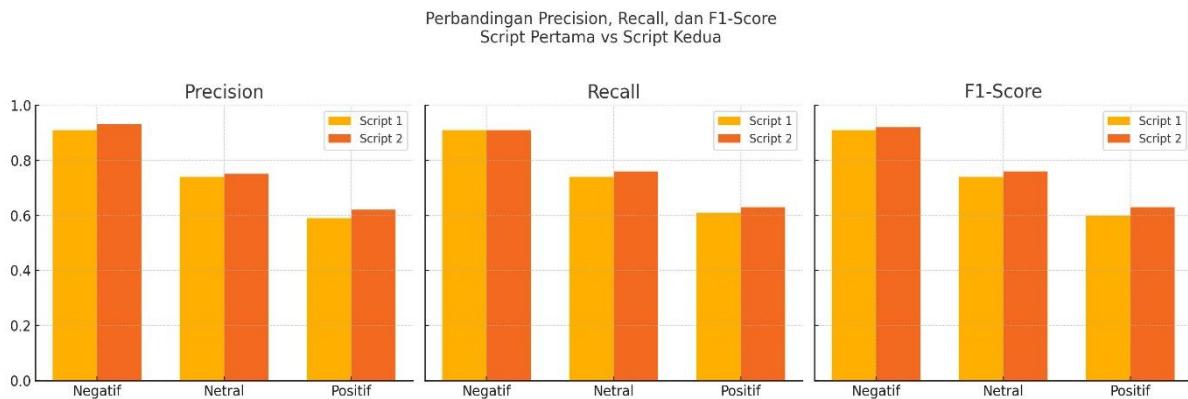
1. Akurasi pada Data Uji Eksternal: Script pertama diuji dengan data eksternal dan menghasilkan akurasi 76.0%, sementara script kedua diuji menggunakan train-test split internal dengan akurasi 75.37%. Ada penurunan sebesar 0.63%. Meski selisihnya kecil, penggunaan data eksternal menunjukkan bahwa model dari script pertama memiliki kemampuan generalisasi yang baik.
2. Akurasi pada Data Uji Internal: Script pertama menghasilkan akurasi sebesar 73.70%, sedang script kedua menghasilkan akurasi sebesar 75.37%. *Script* kedua menunjukkan peningkatan sebesar 1,67% dibanding script pertama. Ini berkat adanya *preprocessing* teks seperti *stopword removal* dan *stemming*, yang membuat fitur teks lebih bersih dan representatif.
3. Akurasi pada Seluruh *Dataset*: *Script* pertama mencatat akurasi 93.36%, sedangkan *script* kedua meningkat menjadi 94.03%. Ada peningkatan sebesar 0.67%. Peningkatan ini, meskipun kecil (0.67%), tetap menunjukkan dampak positif dari pra-pemrosesan.
4. Kelas Negatif: *Precision* meningkat dari 0.91 menjadi 0.93, *F1-score* naik dari 0.91 ke 0.92, sementara *recall* stabil. Ini menunjukkan peningkatan ketepatan dalam mengenali komentar negatif.
5. Kelas Netral: Peningkatan *precision*, *recall*, dan *F1-score* (masing-masing sekitar 0.01–0.02 poin) menunjukkan model kedua lebih mampu mengklasifikasikan komentar ambigu atau netral dengan lebih akurat.
6. Kelas Positif: Terdapat peningkatan *precision* dari 0.59 ke 0.62, *recall* dari 0.61 ke 0.63, dan *F1-score* dari 0.60 ke 0.63. Ini mencerminkan perbaikan dalam mengidentifikasi komentar positif, meskipun tidak terlalu signifikan.
7. Akurasi Umum: *Accuracy* naik dari 0.74 ke 0.75, mencerminkan perbaikan keseluruhan dalam klasifikasi.

8. *Macro Average*: Peningkatan dari 0.75 ke 0.77 menunjukkan bahwa model kedua memberikan kinerja yang lebih seimbang antar kategori sentimen.
9. *Weighted Average*: Terdapat peningkatan dari 0.74 ke 0.76 (*precision*) dan dari 0.74 ke 0.75 (*recall* dan *F1-score*). Artinya, script kedua memberikan kinerja yang lebih baik pada kelas-kelas yang memiliki jumlah data lebih besar, sehingga berdampak langsung pada skor keseluruhan model.

Berikut adalah ringkasan perbandingan antara *Script Naive Bayes* Pertama (tanpa NLTK) dan *Script Kedua* (dengan NLTK) dalam bentuk tabel yang rapi:

Tabel 3.3 Ringkasan Perbandingan Script Pertama vs Script Kedua

No.	Aspek Evaluasi	Script Pertama (Tanpa NLTK)	Script Kedua (Dengan NLTK)	Perbedaan / Catatan
1	Akurasi Data Uji	73.70%	75.37%	Naik 1.67%
2	Akurasi Seluruh Dataset	93.36%	94.03%	Naik 0.67%
3	Precision - Negatif	0.91	0.93	Naik 0.02
4	Recall - Negatif	0.91	0.91	Stabil
5	F1-score - Negatif	0.91	0.92	Naik 0.01
6	Precision - Netral	0.74	0.75	Naik 0.01
7	Recall - Netral	0.74	0.76	Naik 0.02
8	F1-score - Netral	0.74	0.76	Naik 0.02
9	Precision - Positif	0.59	0.62	Naik 0.03
10	Recall - Positif	0.61	0.63	Naik 0.02
11	F1-score - Positif	0.60	0.63	Naik 0.03
12	Akurasi Total	0.74	0.75	Naik 0.01
13	Macro Avg (Precision / Recall / F1-score)	0.75 / 0.75 / 0.75	0.77 / 0.76 / 0.77	Meningkat semua
14	Weighted Avg (Precision / Recall / F1-score)	0.74 / 0.74 / 0.74	0.76 / 0.75 / 0.75	Meningkat semua



Gambar 3.1 Grafik perbandingan *script 1* versus *script 2*

### 3.2 Analisis Sentimen dengan *Transformers*

Peneliti menggunakan dua pendekatan dalam menganalisis komentar terkait judi online menggunakan model *transformer*:

1. Pendekatan pertama: tanpa *preprocessing* seperti NLTK.
2. Pendekatan kedua: dengan *preprocessing* berupa *stopword removal* menggunakan NLTK.

#### 3.2.1 Hasil Analisis Tanpa NLTK

Model yang digunakan dalam penelitian ini menunjukkan tingkat akurasi sebesar 21,55% dari total 2.394 data komentar yang dianalisis. Berdasarkan pelabelan manual, distribusi komentar terdiri dari 1.111 komentar netral, 694 komentar positif, dan 586 komentar negatif. Sementara itu, hasil prediksi model memperlihatkan distribusi yang berbeda secara signifikan, yaitu 1.772 komentar diprediksi sebagai negatif, 477 komentar sebagai positif, dan hanya 145 komentar sebagai netral. Dari keseluruhan prediksi, hanya 516 yang sesuai dengan label sebenarnya, sedangkan 1.878 sisanya merupakan prediksi yang tidak tepat. Rincian kecocokan antara label dan hasil prediksi menunjukkan bahwa sebanyak 82 komentar netral berhasil diprediksi dengan benar sebagai netral, 96 komentar positif diprediksi dengan benar sebagai positif, dan 338 komentar negatif diprediksi dengan tepat sebagai negatif. Temuan ini mengindikasikan bahwa meskipun model memiliki kemampuan dalam mengenali sentimen tertentu, performa keseluruhannya masih rendah dan perlu ditingkatkan, terutama dalam menangani variasi konteks dalam bahasa Indonesia

### 3.2.2 Hasil Analisis dengan NLTK

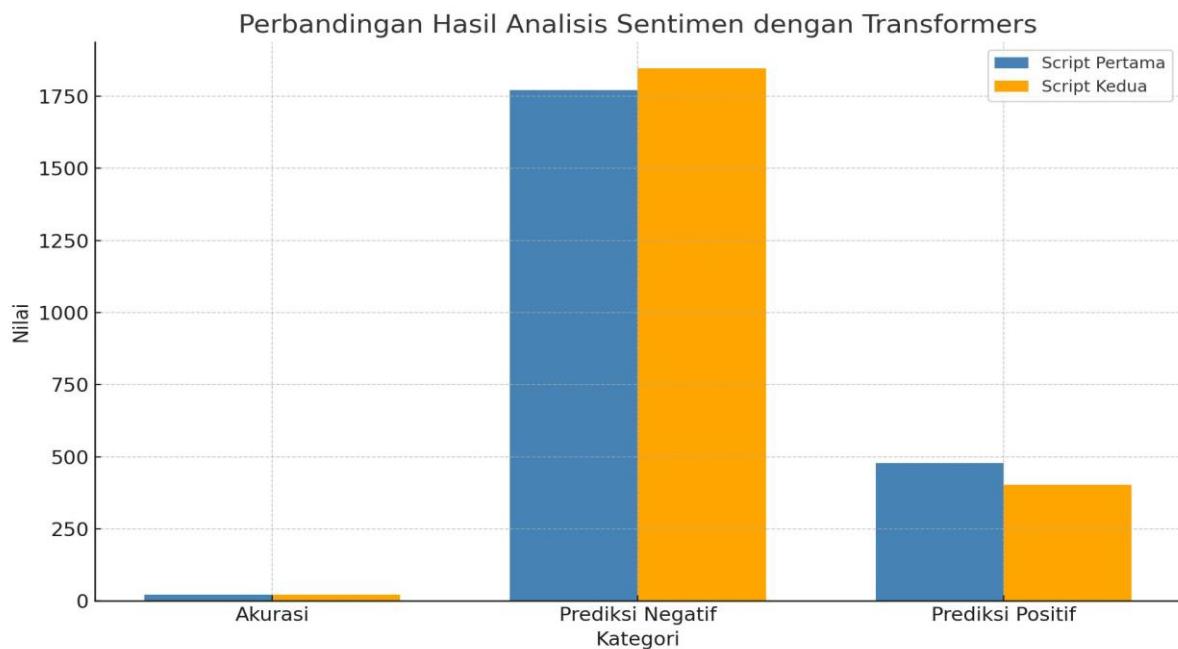
Model analisis sentimen yang diuji dalam penelitian ini menunjukkan akurasi sebesar 21,89% dari total 2.394 komentar yang dianalisis. Distribusi label manual tetap terdiri atas 1.111 komentar netral, 694 komentar positif, dan 586 komentar negatif. Sementara itu, hasil prediksi model menunjukkan 1.846 komentar diprediksi sebagai negatif, 402 sebagai positif, dan 146 sebagai netral. Dari keseluruhan prediksi, sebanyak 524 komentar berhasil diprediksi dengan benar sesuai label aslinya, sedangkan 1.870 komentar lainnya diprediksi secara tidak tepat. Adapun rincian kecocokan antara prediksi dan label sebenarnya menunjukkan bahwa 74 komentar netral berhasil dikenali dengan benar sebagai netral, 85 komentar positif diprediksi secara akurat sebagai positif, dan 365 komentar negatif juga berhasil dikenali secara tepat oleh model. Hasil ini menunjukkan adanya kecenderungan model dalam memprediksi sentimen negatif secara dominan, meskipun belum memberikan performa klasifikasi yang optimal secara keseluruhan. Akurasi yang masih rendah mengindikasikan perlunya peningkatan dalam aspek representasi data, teknik prapemrosesan, atau pemilihan algoritma yang lebih sesuai untuk konteks bahasa Indonesia.

### 3.2.3 Perbandingan Dua Pendekatan

Setelah dilakukan penghapusan *stopword*, akurasi model mengalami peningkatan dari 21,55% menjadi 21,89%. Perubahan ini menunjukkan adanya pengaruh positif dari proses *preprocessing* terhadap kinerja model, meskipun peningkatannya masih relatif kecil. Salah satu dampak utama yang diamati adalah peningkatan jumlah prediksi terhadap sentimen negatif, yang kemungkinan disebabkan oleh pembersihan kata-kata umum yang tidak memiliki bobot sentimen kuat. Sebaliknya, jumlah prediksi terhadap sentimen positif justru menurun, mengindikasikan bahwa beberapa kata bermuatan positif kemungkinan ikut terhapus selama proses penghapusan *stopword*. Sementara itu, prediksi terhadap sentimen netral tidak mengalami perubahan signifikan. Dari sisi kecocokan prediksi, model menunjukkan peningkatan dalam mengenali komentar negatif secara benar, dari sebelumnya 338 menjadi 365. Namun, kecocokan terhadap label positif dan netral mengalami sedikit penurunan. Temuan ini menunjukkan bahwa penghapusan *stopword* dapat membantu meningkatkan deteksi untuk sentimen tertentu, namun juga berisiko mengurangi informasi penting jika tidak dilakukan secara selektif terhadap konteks bahasa Indonesia..

Tabel 3.4 Ringkasan analisis perbandingan *Script* Pertama *versus* *Script* Kedua

No	Aspek Analisis	Script Pertama	Script Kedua	Perbedaan / Catatan
1	Akurasi Model	21.55%	21.89%	Peningkatan kecil berkat proses stopword removal
2	manual_sentiment: Netral	1111	1111	Tidak berubah, label manual tidak dipengaruhi preprocessing
3	manual_sentiment: Positif	694	694	Tidak berubah
4	manual_sentiment: Negatif	586	586	Tidak berubah
5	Prediksi Sentimen Model: Negatif	1772	1846	Meningkat, kemungkinan akibat hilangnya nuansa positif setelah stopword removal
6	Prediksi Sentimen Model: Positif	477	402	Menurun, kata-kata positif terhapus saat stopword removal
7	Prediksi Sentimen Model: Netral	145	146	Hampir tidak berubah
8	is_accurate = False (Prediksi Salah)	1878	1870	Menurun sedikit, indikasi peningkatan performa model
9	is_accurate = True (Prediksi Benar)	516	524	Meningkat, dukungan akurasi oleh preprocessing
10	Kecocokan Sentimen Negatif (manual = prediksi)	338	365	Meningkat, model lebih baik mengenali komentar negative
11	Kecocokan Sentimen Positif (manual = prediksi)	96	85	Menurun, kata positif mungkin terhapus dalam preprocessing
12	Kecocokan Sentimen Netral (manual = prediksi)	82	74	Menurun sedikit, beberapa komentar netral bisa tampak emosional setelah preprocessing



Gambar 3.2 Grafik perbandingan script pertama versus kedua.

### 3.3 Analisis Sentimen dengan VADER

Pada bagian ini, penulis melakukan analisis sentimen menggunakan model VADER (*Valence Aware Dictionary and Sentiment Reasoner*). Uji coba dilakukan dengan dua pendekatan, yaitu: (1) tanpa menyisipkan pendekripsi kata kunci terkait judi dan indikasi sarkasme, serta (2) dengan menyisipkan pendekripsi kedua aspek tersebut. Tujuan dari pendekatan ini adalah untuk mengetahui apakah penambahan komponen analisis semantik tersebut mampu meningkatkan akurasi klasifikasi sentimen.

#### 3.3.1 Tanpa Pendekripsi Kata Kunci dan Sarkasme

Analisis awal dilakukan tanpa melibatkan pendekripsi khusus terhadap kata kunci yang berhubungan dengan praktik judi online maupun indikasi sarkasme dalam teks. Pada tahap ini, model *Valence Aware Dictionary and sEntiment Reasoner* (VADER) digunakan untuk mengklasifikasikan sentimen komentar. Hasil analisis menunjukkan bahwa VADER mengklasifikasikan 1.200 komentar (50,13%) sebagai positif, 760 komentar (31,75%) sebagai negatif, dan 434 komentar (18,13%) sebagai netral. Jika dibandingkan dengan pelabelan manual (ground truth), diketahui bahwa komentar sebenarnya terdiri atas 694 komentar positif, 586 komentar negatif, dan 1.111 komentar netral. Perbandingan ini menunjukkan adanya perbedaan signifikan antara hasil prediksi

VADER dan label yang sebenarnya. Dari evaluasi performa model, diketahui bahwa akurasi model hanya mencapai 20,51%, dengan jumlah prediksi yang sesuai (benar) sebanyak 491 komentar, sedangkan sisanya, yaitu 1.903 komentar, diprediksi secara tidak tepat. Selain itu, jumlah *false positives* tercatat sebanyak 500 kasus, dan *false negatives* sebanyak 349 kasus. Performa ini juga tercermin dalam nilai *Mean Absolute Error* (MAE) yang cukup tinggi, yaitu 0,8974. Hasil ini menunjukkan bahwa VADER belum mampu menangani analisis sentimen dalam bahasa Indonesia secara optimal, terutama ketika tidak didukung oleh deteksi kata kunci kontekstual maupun elemen linguistik seperti sarkasme yang khas dalam komunikasi informal di media sosial.

### 3.3.2 Dengan Pendekstrian Kata Kunci dan Sarkasme

Analisis lanjutan dilakukan dengan menambahkan mekanisme pendekstrian terhadap kata kunci yang berhubungan dengan praktik judi serta indikasi sarkasme dalam teks. Dengan pendekatan ini, model VADER kembali digunakan untuk mengklasifikasikan sentimen komentar. Hasilnya menunjukkan bahwa VADER memprediksi 1.184 komentar (49,46%) sebagai positif, 821 komentar (34,29%) sebagai negatif, dan 389 komentar (16,25%) sebagai netral. Sementara itu, berdasarkan pelabelan manual (*ground truth*), jumlah sebenarnya adalah 694 komentar positif, 586 komentar negatif, dan 1.111 komentar netral. Dibandingkan dengan analisis sebelumnya, model mengalami sedikit pergeseran distribusi prediksi, khususnya pada kategori negatif yang meningkat.

Evaluasi performa model menunjukkan adanya peningkatan akurasi menjadi 22,18%, dengan jumlah prediksi yang sesuai sebanyak 531 komentar, meningkat dari 491 pada analisis pertama. Jumlah prediksi yang tidak sesuai menurun menjadi 1.863 komentar. Selain itu, jumlah *false positives* tercatat sebanyak 493, dan *false negatives* sebanyak 353. Nilai *Mean Absolute Error* (MAE) juga mengalami sedikit penurunan menjadi 0,8673. Peningkatan akurasi ini menunjukkan bahwa integrasi deteksi kata kunci kontekstual dan sarkasme dapat membantu VADER dalam memahami makna sentimen yang lebih kompleks dalam teks berbahasa Indonesia, meskipun performanya masih tergolong rendah dan memerlukan pengembangan lebih lanjut.

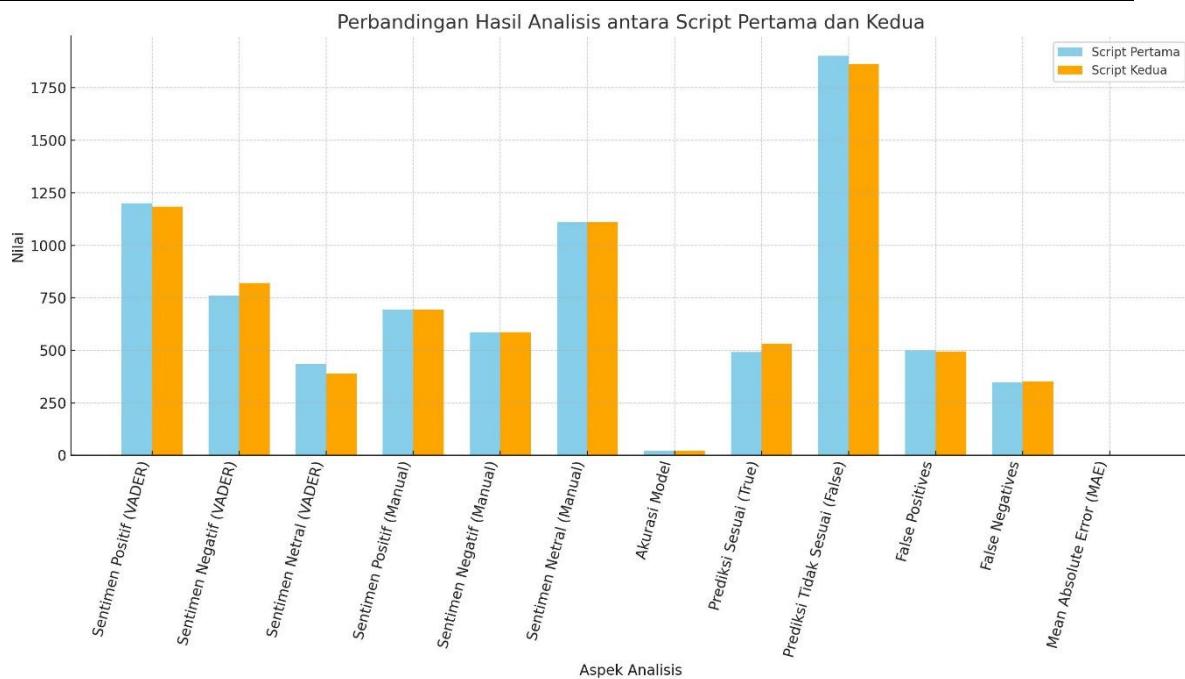
### 3.3.3 Perbandingan Hasil Script Pertama dan Kedua

Berdasarkan dua pendekatan di atas, dilakukan perbandingan kinerja model sebagaimana dirangkum dalam Tabel 4.6 berikut:

Tabel 3.5 Perbandingan kinerja model *script* pertama *versus script* kedua.

No.	Aspek Analisis	Script Pertama	Script Kedua	Catatan
1	 Sentimen Positif (VADER)	1200 (50,13%)	1184 (49,46%)	Menurun sedikit, akibat penyesuaian proses analisis
2	 Sentimen Negatif (VADER)	760 (31,75%)	821 (34,29%)	Naik, menunjukkan sensitivitas terhadap nuansa negatif
3	 Sentimen Netral (VADER)	434 (18,13%)	389 (16,25%)	Turun, sebagian diklasifikasi ulang
4	 Sentimen Positif (Manual)	694	694	Tetap, tidak berubah
5	 Sentimen Negatif (Manual)	586	586	Tetap, tidak berubah
6	 Sentimen Netral (Manual)	1111	1111	Tetap, tidak berubah
7	 Akurasi Model	20,51%	22,18%	Naik 1,67%, menunjukkan sedikit perbaikan
8	 Prediksi Sesuai (True)	491	531	Naik 40 prediksi benar
9	 Prediksi Tidak Sesuai (False)	1903	1863	Turun 40 kesalahan
10	 False Positives	500	493	Turun, klasifikasi positif lebih akurat
11	 False Negatives	349	353	Naik tipis, kesalahan negatif ke netral sedikit meningkat

No.	Aspek Analisis	Script Pertama	Script Kedua	Catatan
12	Mean Absolute Error (MAE)	0,8974	0,8673	Menurun, menunjukkan sedikit peningkatan presisi model



Gambar 3.3 Grafik perbandingan *script* pertama *versus* *script* kedua.

#### 4. KESIMPULAN

Penelitian ini menyimpulkan bahwa dalam analisis sentimen terhadap komentar masyarakat terkait fenomena judi online, model *Naïve Bayes* dengan preprocessing menggunakan NLTK menunjukkan performa paling unggul dibandingkan Transformer dan VADER. *Naïve Bayes* berhasil mencapai akurasi tertinggi, baik pada data keseluruhan maupun data uji, serta menunjukkan peningkatan konsisten pada metrik evaluasi lainnya setelah dilakukan *stopword removal*. Sebaliknya, model Transformer yang secara teori lebih kompleks justru menunjukkan performa rendah dan tidak mengalami peningkatan signifikan, menunjukkan keterbatasannya tanpa fine-tuning pada konteks lokal. Adapun model VADER, meskipun ringan secara komputasi, tetap menunjukkan akurasi yang rendah, namun peningkatan kecil terjadi setelah ditambahkan deteksi kata kunci dan sarkasme. Secara umum, hasil penelitian ini menegaskan

pentingnya proses *preprocessing* dalam meningkatkan akurasi model analisis sentimen, dan menunjukkan bahwa model sederhana seperti Naïve Bayes dapat memberikan hasil yang lebih baik dibanding model yang lebih kompleks, asalkan dilatih dan disesuaikan dengan baik. Selain itu, pendekatan leksikal dan semantik tambahan juga terbukti membantu meningkatkan performa model berbasis aturan seperti VADER.

## 5. DAFTAR PUSTAKA

- [1] Abimanyu, D., Budianita, E., Pandu Cynthia, E., Yanto, F., Studi Teknik Informatika, P., & Sains Dan Teknologi, F. (2022). Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 5(3). <https://techno.kompas.com>
- [2] Akbar, Y., & Sugiharto, T. (2023). Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes (Yuma Akbar 1\*, Tri Sugiharto 2 ) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes. *Jurnal Sains Dan Teknologi*, 5(1), 115–122. <https://doi.org/10.55338/saintek.v4i3.1368>
- [3] Annisa Laras, Najwa Salvabillah, Cindy Caroline, Jusini Delas H, Farra Dinda, & Mic Finanto. (2024). Analisis Dampak Judi Online di Indonesia. *Concept: Journal of Social Humanities and Education*, 3(2), 320–331. <https://doi.org/10.55606/concept.v3i2.1304>
- [4] Aprian Putra dan, J., & Budi, A. (n.d.). *PENERAPAN NATURAL LANGUAGE PROCESSING DALAM APLIKASI CHATBOT SEBAGAI MEDIA PENCARIAN INFORMASI DENGAN MENGGUNAKAN REACT (STUDI KASUS: INSTITUT BISNIS DAN INFORMATIKA KWIK KIAN GIE)*.
- [5] Gultom, M., Marikros, J., & Rusli, W. (n.d.). *SEMINAR NASIONAL CORISINDO Penerapan Vader Sentiment untuk Mendeteksi Sentimen Bahasa Inggris berbasis Website*.
- [6] Hendrawan, A., & Sela, E. I. (2024). Analisis Sentimen Komentar Youtube Tentang Resesi Global 2023 Menggunakan LSTM. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(1), 587–593. <https://doi.org/10.35870/jimik.v5i1.526>

- [7] Maghfiroh Universitas Negeri Surabaya Jl Ketintang No, N., Gayungan, K., Surabaya, K., & Timur, J. (n.d.). *BAHASA INDONESIA SEBAGAI ALAT KOMUNIKASI MASYARAKAT DALAM KEHIDUPAN SEHARI-HARI*.
- [8] Martantoh, E., & Yanih, N. (2022). Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL Implementation of Naive Bayes Method for Classification of Student's Personality Characteristics at MTS Darussa'adah School Using PHP Mysql. In *JTSI* (Vol. 3, Issue 2).
- [9] Naufal Zaidan Nayottama. (n.d.). *Dampak Judi Online terhadap Kondisi Finansial, Hubungan Sosial, dan*.
- [10] Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- [11] Putra, D. H., Wahyu, G., & Noto, I. (n.d.). *IMPLEMENTASI GENERATIVE PRE-TRAINED TRANSFORMERS 2 UNTUK MENGOREKSI KESALAHAN PENULISAN BAHASA INDONESIA PADA DOKUMEN JURNAL*.
- [12] Putri, I. (n.d.). *MEDIA SOSIAL SEBAGAI MEDIA PERGESERAN INTERAKSI SOSIAL REMAJA* (Vol. 2, Issue 2).