

## Analisis Sentimen Supporter Pss Sleman Terhadap Peringkat Klub Di Twitter/X Menggunakan Naive Bayes

M Djauhar Fardi<sup>1</sup>, Mutaqin Akbar<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta

e-mail: [1fardisanss@gmail.com](mailto:fardisanss@gmail.com), [2mutaqin@mercubuana-yogya.ac.id](mailto:mutaqin@mercubuana-yogya.ac.id)

**Abstract** - *Social media has become a primary channel for the public to express opinions on various issues, including those related to sports. Twitter/X, as one of the most popular platforms, enables users to openly share their views on football club performance. This research focuses on examining how supporters feel about the PSS Sleman football club using the Multinomial Naive Bayes algorithm. Data were collected using the Twitter API and SNScrape with relevant keywords and processed through several preprocessing stages such as cleaning, case folding, filtering, tokenizing, stemming, and transformation into numerical form using the TF-IDF method. Sentiments were classified into three categories: positive, negative, and neutral. The evaluation results using a confusion matrix indicate that the model accurately classified all 100 test tweets, consisting of 52 negative, 5 neutral, and 43 positive tweets. The model achieved precision, recall, and F1-score values of 1.00 across all classes, with an overall accuracy of 100%, reflecting excellent classification performance. These findings demonstrate that the Naive Bayes algorithm is highly effective for analyzing public opinion on social media, particularly in the context of football. The system is expected to assist club management in understanding supporter perceptions and in formulating more targeted communication strategies.*

**Keywords** - *Sentiment Analysis, Naive Bayes, Twitter, PSS Sleman.*

**Abstrak** - Media sosial telah menjadi saluran utama bagi masyarakat dalam mengekspresikan opini terhadap berbagai isu, termasuk dalam bidang olahraga. Twitter/X, sebagai salah satu platform yang populer, memungkinkan pengguna untuk menyampaikan pendapat secara terbuka mengenai performa klub sepak bola. Penelitian ini berfokus pada menelaah bagaimana perasaan para supoter terhadap klub PSS Sleman dengan menggunakan algoritma Multinomial Naive Bayes. Data dikumpulkan melalui API Twitter dan SNScrape dengan kata kunci yang relevan, kemudian diproses melalui tahapan praproses data seperti cleaning, case folding, filtering, tokenizing, stemming, hingga transformasi ke bentuk numerik menggunakan metode TF-IDF. Sentimen diklasifikasikan ke dalam tiga kategori: positif, negatif, dan netral. Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa model mampu mengklasifikasikan seluruh data uji sebanyak 100 tweet dengan tepat, masing-masing terdiri dari 52 tweet negatif, 5 tweet netral, dan 43 tweet positif. Model menghasilkan nilai precision, recall, dan F1-score sebesar 1.00 untuk ketiga kelas, serta akurasi 100%, yang mencerminkan performa klasifikasi yang sangat optimal. Temuan ini menunjukkan bahwa algoritma Naive Bayes sangat efektif dalam menganalisis opini publik di media sosial, khususnya dalam konteks persepakbolaan. Sistem ini diharapkan dapat menjadi alat bantu manajemen klub dalam memahami persepsi supoter serta merumuskan strategi komunikasi yang lebih tepat sasaran.

**Kata Kunci** - *Analisis Sentimen, Naive Bayes, Twitter, PSS Sleman.*

### I. PENDAHULUAN

Saat ini, di tengah perkembangan era digital, media sosial kini menjadi bagian penting dalam kehidupan modern, terutama dalam mendukung komunikasi dan penyebaran informasi yang cepat serta luas. Perubahan ini turut memengaruhi cara individu dalam mengungkapkan opini serta perasaan mereka terhadap berbagai situasi. Salah satu platform yang cukup populer adalah Twitter, karena kemampuannya dalam memfasilitasi penyampaian pesan singkat secara langsung [1]. Twitter kerap digunakan sebagai media untuk menyalurkan pandangan, kritik, maupun respons terhadap isu-isu yang sedang berkembang, termasuk dalam ranah olahraga seperti sepak bola.

PSS Sleman merupakan tim sepak bola Indonesia yang berasal dari Kabupaten Sleman, Daerah Istimewa Yogyakarta. Klub ini dikenal memiliki basis pendukung yang sangat loyal dan militan, yakni Slemania dan Super Elja, yang aktif mengikuti perkembangan klub baik di dalam stadion maupun di ranah digital. Dalam beberapa musim terakhir, PSS Sleman dihadapkan pada berbagai kendala, mulai dari penurunan performa, pergantian pelatih, hingga kebijakan manajemen yang kerap menuai kritik. Situasi ini memicu gelombang reaksi dari para pendukung, yang banyak disuarakan melalui media sosial, terutama Twitter. Platform tersebut menjadi medium utama bagi para supoter untuk mengekspresikan rasa kecewa, kritik tajam, maupun bentuk dukungan terhadap klub secara terbuka. Banyak cuitan yang mencerminkan sentimen kuat terhadap performa tim, kebijakan internal, hingga isu-isu yang berhubungan langsung dengan loyalitas supoter. Oleh karena itu, penting untuk menganalisis dinamika opini ini secara terstruktur guna memahami bagaimana persepsi publik terbentuk dan berubah seiring waktu.

Dalam bidang ini, analisis sentimen merupakan salah satu metode yang secara luas digunakan untuk mengidentifikasi sikap atau emosi pengguna terhadap suatu permasalahan melalui data teks. Beberapa algoritma machine learning yang sering diterapkan dalam proses ini antara lain Convolutional Neural Network (CNN) [2], Support Vector Machine (SVM)

[3], [4], Decision Tree [5], serta k-Nearest Neighbors (k-NN) [6], karena kemampuannya dalam menghasilkan klasifikasi teks yang cukup presisi [7]. Selain itu, algoritma Naive Bayes Classifier juga menjadi pilihan populer karena berbasis pendekatan probabilistik yang sederhana namun efektif dalam menganalisis dan mengelompokkan data teks [8].

Sejumlah studi sebelumnya telah membuktikan bahwa algoritma Naive Bayes memiliki kemampuan yang tinggi dalam mengklasifikasikan opini pengguna di media sosial dengan tingkat akurasi yang memadai. Algoritma ini telah diaplikasikan dalam berbagai penelitian, termasuk dalam menganalisis tanggapan masyarakat terhadap klub sepak bola melalui media seperti Twitter [9]. Di Indonesia, media sosial telah menjadi saluran penting bagi para penggemar klub sepak bola seperti suporter PSS Sleman untuk mengekspresikan pendapat mereka terhadap performa tim kesayangan.

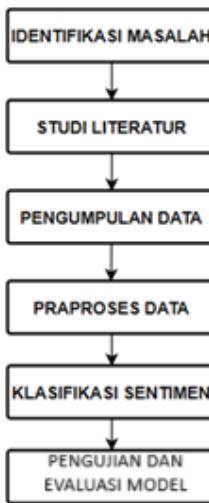
Penerapan metode Naive Bayes Classifier telah banyak digunakan dalam berbagai studi terdahulu. Sebagai satu contoh ialah penelitian yang diuji oleh [10]. Secara umum, algoritma Naive Bayes Classifier dikenal sebagai metode klasifikasi berbasis probabilistik yang sederhana namun cukup andal. Keunggulan utamanya terletak pada kemampuannya untuk membangun model klasifikasi dengan jumlah data pelatihan yang tidak terlalu besar. Karena asumsi adanya independensi antar fitur, algoritma ini hanya membutuhkan informasi varians per fitur di tiap kelas, tanpa perlu menghitung seluruh matriks kovarian. Proses klasifikasi dengan metode ini merujuk pada prinsip dasar dari Teorema Bayes. Akan tetapi, melakukan analisis terhadap opini suporter secara menyeluruh merupakan tantangan tersendiri. Hal ini dikarenakan oleh tingginya jumlah tweet serta keberagaman dalam pilihan kata, gaya bahasa, dan bentuk ekspresi yang digunakan [11]. Kompleksitas tersebut menuntut digunakannya metode analisis yang terstruktur dan berbasis teknologi agar dapat memperoleh hasil yang benar-benar mewakili opini publik. Dalam hal ini, kombinasi pemanfaatan data Twitter dengan penerapan algoritma Naive Bayes Classifier dinilai sebagai pendekatan yang tepat dan efisien [12]. Oleh karena itu, dibutuhkan sistem klasifikasi sentimen otomatis yang mampu merangkum persepsi suporter terhadap performa klub secara komprehensif.

## II. PENELITIAN YANG TERKAIT

Beberapa penelitian terdahulu menunjukkan bahwa algoritma Naive Bayes cukup andal dalam penerapan analisis sentimen pada media sosial, pada penelitian ini membangun sistem klasifikasi sentimen menjadi tujuan dengan menggunakan metode Naive Bayes yang dapat mengelompokkan opini suporter PSS Sleman ke dalam tiga kelas utama: positif, negatif, dan netral, berdasarkan konten cuitan mereka [13]. Menunjukkan bahwa algoritma ini efektif dalam mengklasifikasikan opini masyarakat terhadap klub sepak bola. Temuan serupa juga dilaporkan oleh [14] yang menyebutkan bahwa *Naive Bayes* mampu menangani berbagai ragam ekspresi bahasa dengan baik. Sementara itu [15], menekankan pentingnya peran media sosial sebagai saluran utama bagi suporter di Indonesia, termasuk pendukung PSS Sleman, dalam menyampaikan opini mereka. Dalam studi lainnya CNN diterapkan untuk mengklasifikasikan tweet terkait isu Jogja Darurat Sampah, dengan dataset berjumlah 505 tweet. Sebanyak 381 tweet digunakan sebagai data latih, dan 124 sisanya sebagai data uji. Hasil pelatihan menunjukkan bahwa model CNN mampu mencapai akurasi sebesar 90% pada epoch ke-19. Namun demikian, saat diuji terhadap data uji, akurasi menurun signifikan menjadi hanya 53%, yang mengindikasikan adanya kemungkinan overfitting. Temuan ini menunjukkan bahwa meskipun CNN memiliki potensi besar dalam mempelajari pola kompleks pada data teks pendek seperti tweet, tantangan dalam hal generalisasi terhadap data baru tetap perlu diperhatikan. Oleh karena itu, penerapan CNN dalam analisis sentimen perlu didukung oleh strategi tambahan seperti regularisasi atau penyesuaian arsitektur agar dapat menghasilkan performa yang lebih stabil di berbagai jenis data [2]. Penelitian lainnya menggunakan Support Vector Machine (SVM) untuk analisis sentimen masyarakat terhadap revisi Undang-Undang TNI yang dipublikasikan di Twitter selama periode 1–31 Maret 2025. Data diambil melalui scraping, kemudian dipraproses dan diubah menjadi representasi numerik dengan metode TF-IDF sebelum diberi label sentimen secara manual. Evaluasi model dilakukan menggunakan teknik 5-fold cross-validation dan confusion matrix, menghasilkan akurasi rata-rata sebesar 78,99% dan nilai F1-score 0,83 untuk kelas sentimen negatif [3]. Penelitian yang dilakukan oleh [4] menunjukkan penerapan algoritma Support Vector Machine (SVM) dalam menganalisis sentimen bagi opini di masyarakat mengenai RUU Perampasan Aset di media sosial Twitter dan YouTube. Data dikumpulkan melalui scraping dengan Google Colab, lalu diproses dengan menggunakan tahapan preprocessing yaitu cleansing, case folding, tokenizing, filtering, stemming, dan normalisasi. Untuk penilaian pada kata, digunakan metode TF-IDF, dan ketidakseimbangan kelas diatasi menggunakan SMOTE. Model SVM dilatih dengan skema ShuffleSplit dan dievaluasi dengan memakai confusion matrix. Evaluasi memberikan hasil yang menunjukkan performa yang baik, dengan rata-rata akurasi sebesar 79,8%, precision 80,58%, recall 79,8%, dan F1-score 79,96%. Selain itu, visualisasi hasil klasifikasi menunjukkan bahwa mayoritas opini publik terhadap RUU tersebut bernada negatif. Penelitian-penelitian tersebut menjadi dasar pertimbangan dalam pengembangan sistem klasifikasi pada studi ini.

## III. METODE PENELITIAN

Penelitian ini terdiri dari tujuh tahapan pokok, yang mencakup proses pengumpulan data, perancangan sistem, pelaksanaan implementasi, serta tahap pengujian. Keseluruhan alur dari setiap tahapan tersebut disajikan secara visual pada gambar berikut.



Gbr. 1 Alur penelitian.

#### A. Identifikasi Masalah

Di era digital, media sosial menjadi platform yang aktif digunakan untuk menyampaikan opini dan aspirasi publik. Platform seperti Twitter memungkinkan pengguna untuk menyampaikan pendapat secara langsung terhadap isu-isu yang sedang terjadi, termasuk dalam ranah olahraga. Suporter klub sepak bola seperti PSS Sleman sangat aktif dalam menggunakan media sosial untuk menyuarakan dukungan, kritik, atau harapan terhadap performa tim kesayangan mereka, khususnya ketika menyangkut posisi klub di klasemen liga. Opini-opini tersebut disampaikan dalam berbagai bentuk gaya bahasa, pilihan kata, serta nada emosional yang beragam, sehingga menimbulkan tantangan tersendiri dalam hal pengelompokan atau klasifikasinya secara manual.

Masalah pada penelitian kali ini adalah bagaimana mengklasifikasikan opini masyarakat dalam bentuk tweet berbahasa Indonesia dalam bentuk kategori sentimen berupa positif, netral, dan negatif secara otomatis dan akurat menggunakan algoritma *Naive Bayes*. Tantangan utama terletak pada pengolahan teks tidak terstruktur dalam jumlah besar, yang memerlukan proses prapemrosesan data yang sistematis, serta pemilihan algoritma klasifikasi yang sesuai. Dengan membangun sistem klasifikasi berbasis *Naive Bayes*, diharapkan dapat diperoleh gambaran umum yang objektif mengenai persepsi publik terhadap performa PSS Sleman, serta menyediakan dasar analitis yang dapat dimanfaatkan dalam pengambilan keputusan strategis oleh manajemen klub.

#### B. Studi Literatur

Studi literatur dilakukan sebagai landasan teoretis dan metodologis dalam menyusun sistem klasifikasi sentimen. Tahap ini mencakup pengumpulan dan kajian berbagai referensi ilmiah yang memiliki hubungan dengan analisis sentimen dengan algoritma *Naive Bayes*, metode *text mining*, serta penelitian-penelitian terdahulu yang relevan. Analisis sentimen merupakan teknik dalam *natural language processing* (NLP) memiliki tujuan untuk mengenali serta membuat pengelompokan opini kedalam sebuah teks menjadi sentimen positif, negatif, netral. Proses ini sangat berguna dalam memahami respons pengguna terhadap suatu topik di media sosial secara otomatis. Algoritma *Naive Bayes* dipilih dalam penelitian ini karena memiliki keunggulan dalam hal kesederhanaan implementasi, efisiensi komputasi, dan performa yang cukup baik dalam pengklasifikasian data teks. Algoritma ini bekerja berdasarkan prinsip Teorema Bayes dengan asumsi independensi antar fitur. Selain itu, metode *text mining* juga dibahas dalam studi literatur sebagai pendekatan dalam mengekstraksi informasi dari data teks tidak terstruktur. Penelitian terdahulu yang menjadi acuan menunjukkan bahwa algoritma *Naive Bayes* telah berhasil digunakan untuk mengklasifikasikan opini publik terhadap isu-isu politik, pelayanan publik, hingga performa klub sepak bola. Hasil dari studi-studi tersebut menunjukkan bahwa pendekatan ini mampu menghasilkan akurasi yang cukup tinggi dalam berbagai konteks data, sehingga menjadi dasar yang kuat untuk diterapkan dalam penelitian ini.

#### C. Pengumpulan Data

Data dalam penelitian ini dikumpulkan melalui platform media sosial Twitter, yang dipilih karena karakteristiknya sebagai media komunikasi terbuka dan cepat dalam menyebarkan opini publik. Data diambil menggunakan dua metode, yaitu melalui pemanfaatan Twitter API v2 dan alat bantu scraping bernama SNScrape. Untuk memperoleh data yang relevan, digunakan sejumlah kata kunci seperti "PSS Sleman", "peringkat", dan "klasemen", yang berfokus pada opini terkait performa klub.

Tweet yang dikumpulkan merupakan ungkapan berbahasa Indonesia dan dihimpun dalam periode waktu tertentu, yaitu dari tanggal 11 Agustus 2024 hingga 24 Mei 2025. Rentang waktu ini dipilih agar dapat mencerminkan dinamika sentimen suporter selama berlangsungnya kompetisi sepak bola di musim tersebut.

Setelah proses pengumpulan selesai, data mentah diseleksi untuk menghilangkan entri yang tidak relevan, seperti tweet yang tidak mengandung opini, spam, atau retweet. Hasil seleksi tersebut kemudian disimpan dalam format CSV sebagai

korpus data utama yang akan digunakan dalam tahap analisis selanjutnya. Tahapan ini penting untuk memastikan bahwa data yang digunakan memiliki kualitas serta relevansi tinggi terhadap tujuan penelitian, yaitu mengklasifikasikan sentimen publik terhadap kinerja PSS Sleman menggunakan algoritma *Naive Bayes*.

#### D. Praproses Data

Tahapan ini menjadi langkah awal setelah data dibagi ke dalam set pelatihan dan pengujian. Pada tahap praprosesan teks (*text preprocessing*), dilakukan serangkaian proses untuk membersihkan dan mempersiapkan data teks mentah agar dapat digunakan secara optimal dalam analisis selanjutnya. Proses ini bertujuan untuk mengurangi kompleksitas bahasa alami dan mengubah teks menjadi format yang lebih terstruktur sehingga dapat diproses oleh algoritma pembelajaran mesin.

##### a. Pembersihan Teks (*Cleaning*)

Langkah pertama dalam pengolahan teks adalah menghapus elemen-elemen yang tidak relevan, seperti tanda baca, angka, tagar, tautan, dan *mention* pengguna. Tujuannya adalah untuk menyederhanakan teks dari unsur yang tidak berkontribusi terhadap analisis.

Contoh:

Teks asli: @pssleman Luar biasa! Menang 3-0! #EljaPride

<https://pss.co.id>

Hasil setelah cleaning: luar biasa menang

##### b. Normalisasi Huruf (*Case Folding*)

Pada tahap ini, seluruh karakter dalam teks diubah menjadi huruf kecil. Proses ini penting untuk menyamakan format kata dan menghindari pembacaan ganda akibat perbedaan huruf kapital.

Contoh:

PSS Sleman Menang → pss sleman menang

##### c. Penyaringan Kata (*Filtering*)

*Filtering* dilakukan dengan menghapus kata-kata umum atau *stopwords* yang tidak memiliki nilai analitis, seperti "dan", "yang", "di", dan sebagainya. Tujuannya adalah menyaring kata-kata yang tidak relevan agar hanya kata penting yang tersisa.

Contoh:

pss adalah tim yang sangat hebat → pss tim hebat

##### d. Tokenisasi (*Tokenizing*)

Tokenisasi bertujuan memecah teks menjadi unit-unit kata atau token. Setiap kata yang terpisah oleh spasi dianggap sebagai satu token yang akan dianalisis secara individual.

Contoh:

pss sleman main bagus → ['pss', 'sleman', 'main', 'bagus']

##### e. Stemming

Stemming adalah proses mengembalikan kata ke bentuk dasar atau akarnya, sesuai dengan aturan dalam Kamus Besar Bahasa Indonesia (KBBI). Proses ini dilakukan menggunakan pustaka Python seperti Sastraw.

Contoh:

diberikan, berlari, penyerangan → beri, lari, serang

##### f. Lematisasi (*Lemmatization*)

Lematisasi berfungsi menyederhanakan kata menjadi bentuk lema yang benar secara gramatikal dan semantik (Tri Julianto, n.d.). Meskipun mirip dengan *stemming*, proses ini mempertimbangkan konteks kalimat agar hasilnya lebih presisi.

Contoh:

lebih cepat → cepat, bermainlah → main

#### E. Klasifikasi Sentimen

Setelah data teks melalui tahapan pra-pemrosesan, langkah selanjutnya adalah melakukan transformasi data menjadi bentuk numerik melalui ekstraksi fitur menggunakan metode seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) atau *Count Vectorizer*. Representasi numerik ini penting agar data dapat diproses oleh algoritma pembelajaran mesin.

Dalam penelitian ini, digunakan algoritma *Naive Bayes Classifier* untuk mengelompokkan data menjadi tiga jenis sentimen, yaitu *positif*, *negatif*, dan *netral* berdasarkan isi tweet

##### a. Perhitungan Bobot TF-IDF

Metode TF-IDF digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan seluruh dokumen lain dalam korpus. Metode ini terdiri dari dua komponen utama yaitu *Term Frequency* (TF), yaitu frekuensi kemunculan kata tertentu dalam satu dokumen

$$TF(d, t) \quad (1)$$

dan *Inverse Document Frequency* (IDF), yaitu ukuran seberapa jarang kata tersebut muncul di seluruh dokumen

$$IDF(t) = \frac{N}{df(t)} \quad (2)$$

Gabungan dari kedua nilai tersebut membentuk skor akhir TF-IDF

$$TF - IDF(t, d) = TF(d, t) \times IDF(t) \quad (3)$$

Di mana :

N adalah total jumlah dokumen

$df(t)$  adalah jumlah dokumen yang mengandung kata

#### F. Pengujian dan Evaluasi Model

Setelah proses pelatihan model dilakukan menggunakan algoritma *Multinomial Naive Bayes*, langkah selanjutnya adalah melakukan pengujian untuk mengevaluasi kinerja model dalam mengklasifikasikan data uji. Pengujian ini bertujuan untuk mengetahui seberapa baik model dapat memprediksi kategori sentimen terhadap data baru yang belum pernah dilatih sebelumnya. Proses evaluasi dilakukan dengan membagi dataset menjadi dua bagian, yaitu data latih dan data uji menggunakan rasio 80:20. Data latih digunakan untuk membentuk model klasifikasi, sedangkan data uji digunakan untuk mengukur performa model. Evaluasi model dilakukan menggunakan confusion matrix, yang digunakan untuk menghitung jumlah prediksi benar dan salah dari masing-masing kelas sentimen (positif, negatif, dan netral).

Dari confusion matrix tersebut, akan dihitung beberapa metrik evaluasi seperti:

1. Akurasi, yaitu persentase data yang diklasifikasikan dengan benar oleh model.
2. Precision, yaitu tingkat ketepatan klasifikasi terhadap satu kelas tertentu.
3. Recall, yaitu kemampuan model dalam mengenali data dari kelas tertentu.
4. F1-score, yaitu rata-rata harmonis antara precision dan recall.

Pengujian ini bertujuan untuk memberikan gambaran objektif terhadap kinerja algoritma yang digunakan, serta memastikan bahwa model tidak hanya bekerja baik pada data pelatihan, tetapi juga dapat menggeneralisasi dengan baik terhadap data uji

## IV. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan pendekatan *cross-section* atau pengumpulan data insidental, di mana data dikumpulkan hanya dalam satu periode waktu tertentu tanpa berulang. Pada penelitian ini, proses pengumpulan dilakukan selama kurun waktu 11 Agustus 2024 hingga 24 Mei 2025, dengan total 1.000 komentar yang diambil dari unggahan akun resmi PSS Sleman di *Twitter*. Untuk mendapatkan data tersebut, digunakan alat bantu yang terhubung dengan *Twitter API*, yang dijalankan melalui browser Google Chrome, dan hasil *scraping* disimpan dalam format file CSV. Analisis dilakukan dengan menerapkan algoritma *Naive Bayes Classifier*, menggunakan sejumlah *library Python*, seperti Sastrawi untuk proses *stemming* dan *Wordcloud* untuk visualisasi data. Sebelum pemrosesan dilakukan, semua *library* diinstal terlebih dahulu melalui terminal dengan perintah pip. Pengambilan data tweet dilakukan menggunakan SNScrape dengan menerapkan kata kunci "PSS Sleman peringkat OR klasemen". Perintah ini bertujuan untuk mengumpulkan hingga 1.000 tweet yang memiliki keterkaitan dengan topik tersebut, kemudian hasilnya disimpan dalam format .json sebagai bahan untuk analisis sentimen. Proses inisialisasi dan autentikasi terhadap API *Twitter* dilakukan dengan memanfaatkan *library* Tweepy. Dalam tahap ini, digunakan *consumer\_key*, *consumer\_secret*, *access\_token*, dan *access\_token\_secret* untuk memastikan aplikasi memiliki izin akses terhadap data *Twitter* melalui API. Pengambilan data tweet menggunakan *Twitter API v2* melalui modul *requests* di Python. Kode ini menggunakan metode pencarian recent search dengan query "PSS Sleman lang:id -is:retweet" untuk mengambil tweet berbahasa Indonesia yang menyebut "PSS Sleman" dan menghindari *retweet*. Autentikasi dilakukan menggunakan Bearer Token yang dimasukkan ke dalam *header* permintaan.

TABEL I  
CONTOH HASIL PENGAMBILAN DATA TWITTER API v2

No	Isi Tweet
1	PSS Sleman mainnya bagus malam ini, semangat terus!
2	Posisi klasemen PSS Sleman turun lagi, manajemen harus evaluasi.
3	Akhirnya menang juga, PSS Sleman kembali ke jalur positif.
4	Pertandingan PSS Sleman besok krusial banget.
5	Klasemen makin padat, semoga PSS Sleman tetap konsisten.
6	Manajemen PSS Sleman perlu gerak cepat kalau mau naik peringkat.

No	Isi Tweet
7	Performa PSS Sleman mulai stabil, pelatih baru cocok nih.
8	Masih ada peluang untuk PSS Sleman naik 5 besar klasemen.
9	Kritik keras buat lini belakang PSS Sleman malam ini.
10	Gak nyangka bisa comeback, mantap PSS Sleman!
11	PSS Sleman harus lebih fokus di babak kedua, terlalu banyak kesalahan.
12	Klasemen sementara belum menguntungkan, tapi masih bisa dikejar.
13	Terima kasih sudah berjuang, semangat terus PSS Sleman.
14	Beberapa keputusan wasit merugikan PSS Sleman, sayang banget.
15	Semoga laga selanjutnya jadi titik balik buat PSS Sleman.

Proses cleaning data tweet dilakukan menggunakan bahasa pemrograman Python. Dalam tahap ini, modul `re` dan `nltk` dimanfaatkan untuk menghilangkan elemen-elemen yang tidak relevan seperti tautan (URL), *mention*, *hashtag*, simbol, serta karakter non-alfabet. Setelah itu, teks diubah menjadi huruf kecil (lowercase), kemudian melalui proses tokenisasi dan penghapusan *stopwords* dalam bahasa Indonesia. Output dari proses ini adalah kumpulan tweet yang telah dibersihkan dan siap untuk digunakan dalam analisis sentimen. Hasil dari proses cleaning dapat dilihat pada Tabel II.

TABEL II  
CONTOH HASIL PEMBERSIHAN DATA TWEET

No	Isi Tweet	Isi Tweet Setelah Dibersihkan
1	PSS Sleman mainnya bagus malam ini, semangat terus!	pss sleman main bagus malam semangat
2	Posisi klasemen PSS Sleman turun lagi, manajemen harus evaluasi.	posisi klasemen pss sleman turun manajemen evaluasi
3	Akhirnya menang juga, PSS Sleman kembali ke jalur positif.	akhirnya menang pss sleman kembali jalur positif
4	Pertandingan PSS Sleman besok krusial banget.	pertandingan pss sleman besok krusial
5	Klasemen makin padat, semoga PSS Sleman tetap konsisten.	klasemen makin padat semoga pss sleman tetap konsisten
6	Manajemen PSS Sleman perlu gerak cepat kalau mau naik peringkat.	manajemen pss sleman perlu gerak cepat naik peringkat
7	Performa PSS Sleman mulai stabil, pelatih baru cocok nih.	performa pss sleman mulai stabil pelatih baru cocok
8	Masih ada peluang untuk PSS Sleman naik 5 besar klasemen.	peluang pss sleman naik besar klasemen
9	Kritik keras buat lini belakang PSS Sleman malam ini.	kritik keras lini belakang pss sleman malam
10	Gak nyangka bisa comeback, mantap PSS Sleman!	gak nyangka comeback mantap pss sleman
11	PSS Sleman harus lebih fokus di babak kedua, terlalu banyak kesalahan.	pss sleman fokus babak kedua banyak kesalahan
12	Klasemen sementara belum menguntungkan, tapi masih bisa dikejar.	klasemen sementara belum menguntungkan bisa dikejar
13	Terima kasih sudah berjuang, semangat terus PSS Sleman.	terima kasih berjuang semangat pss sleman
14	Beberapa keputusan wasit merugikan PSS Sleman, sayang banget.	keputusan wasit merugikan pss sleman sayang
15	Semoga laga selanjutnya jadi titik balik buat PSS Sleman.	semoga laga selanjutnya titik balik pss sleman

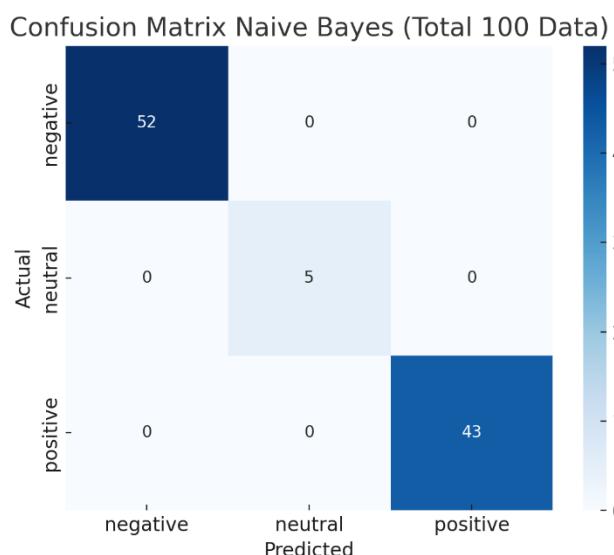
Dari Tabel II dapat dilihat bahwa hasil preprocessing berhasil menyaring kata-kata yang tidak relevan dan menyisakan inti makna dari setiap tweet. Tweet-tweet yang semula mengandung berbagai unsur seperti tanda baca, kata hubung, dan kata-kata tidak penting (*stopwords*), telah disederhanakan menjadi kumpulan kata kunci yang lebih representatif terhadap isi pesan. Misalnya, pada tweet nomor 1 yang semula berbunyi “*PSS Sleman mainnya bagus malam ini, semangat terus!*”, setelah dibersihkan menjadi “*pss sleman main bagus malam semangat*” menunjukkan bahwa kata-kata penting yang merepresentasikan sentimen tetap dipertahankan, sementara kata pelengkap yang tidak terlalu berpengaruh dihilangkan. Proses ini sangat penting untuk memastikan bahwa data yang akan dianalisis benar-benar mencerminkan opini inti dari pengguna, tanpa terganggu oleh elemen-elemen bahasa yang bersifat teknis atau umum. Melalui penyajian tabel ini, terlihat

jelas bagaimana setiap tweet mengalami transformasi secara sistematis, dari bentuk mentah menjadi bentuk siap olah. Setelah melalui tahap preprocessing, setiap tweet dianalisis menggunakan algoritma Naive Bayes untuk menentukan polaritas sentimennya, yaitu positif, negatif, atau netral. Tabel berikut menyajikan hasil klasifikasi terhadap 15 tweet yang telah dibersihkan, di mana setiap baris mencerminkan satu tweet, teks hasil preprocessing, serta hasil prediksi sentimennya.

TABEL III  
PELABELAN SENTIMEN

No	Isi Tweet	Sentimen
1	pss sleman main bagus malam semangat	positive
2	posisi klasemen pss sleman turun manajemen evaluasi	negative
3	akhirnya menang pss sleman kembali jalur positif	positive
4	pertandingan pss sleman besok krusial	neutral
5	klasemen makin padat semoga pss sleman tetap konsisten	positive
6	manajemen pss sleman perlu gerak cepat naik peringkat	negative
7	performa pss sleman mulai stabil pelatih baru cocok	positive
8	peluang pss sleman naik besar klasemen	positive
9	kritik keras lini belakang pss sleman malam	negative
10	gak nyangka comeback mantap pss sleman	positive
11	pss sleman fokus babak kedua banyak kesalahan	negative
12	klasemen sementara belum menguntungkan bisa dikejar	neutral
13	terima kasih berjuang semangat pss sleman	positive
14	keputusan wasit merugikan pss sleman sayang	negative
15	semoga laga selanjutnya titik balik pss sleman	positive

Dalam proses analisis sentimen, pelabelan data merupakan tahapan krusial yang menentukan kualitas hasil klasifikasi. Tabel III menyajikan hasil pelabelan secara manual terhadap lima belas tweet yang berkaitan dengan klub sepak bola PSS Sleman. Proses pelabelan dilakukan dengan membaca dan memahami konteks serta nada emosional dari masing-masing tweet, kemudian menetapkan label sentimen yang sesuai, yaitu *positive*, *neutral*, atau *negative*. Misalnya, tweet seperti “pss sleman main bagus malam semangat” dan “akhirnya menang pss sleman kembali jalur positif” jelas menunjukkan nada optimis dan dukungan, sehingga diberi label *positive*. Sebaliknya, tweet seperti “manajemen pss sleman perlu gerak cepat naik peringkat” dan “keputusan wasit merugikan pss sleman sayang” mengandung kritik dan kekecewaan, sehingga dilabeli *negative*. Sementara itu, beberapa tweet seperti “pertandingan pss sleman besok krusial” atau “klasemen sementara belum menguntungkan bisa dikejar” tidak menunjukkan sentimen yang dominan, sehingga dilabeli *neutral*.



Gbr. 2 Hasil Confusion Matrix.

Gambar 2 menampilkan visualisasi confusion matrix hasil klasifikasi sentimen terhadap data tweet yang telah dianalisis menggunakan algoritma Naive Bayes. Visualisasi ini disajikan dalam bentuk heatmap menggunakan pustaka Seaborn di

Python, yang berfungsi untuk menunjukkan perbandingan antara label aktual dan prediksi model secara visual. Warna intensitas pada matriks membantu menekankan distribusi klasifikasi, dengan warna biru gelap merepresentasikan jumlah prediksi yang tinggi. Berdasarkan hasil evaluasi melalui confusion matrix tersebut, terlihat bahwa model klasifikasi bekerja secara sempurna tanpa kesalahan prediksi. Seluruh data uji yang terdiri dari tiga kategori sentimen negatif, netral, dan positif berhasil diklasifikasikan dengan tepat sesuai label aslinya. Tercatat sebanyak 52 tweet diklasifikasikan dengan benar sebagai sentimen negatif (True Negative), 5 tweet sebagai sentimen netral (True Neutral), dan 43 tweet sebagai sentimen positif (True Positive). Tidak ditemukan adanya kesalahan klasifikasi dalam bentuk False Positive, False Negative, maupun kesalahan antar kelas lainnya. Keberhasilan model dalam menghasilkan confusion matrix yang bersih secara diagonal dengan nilai prediksi hanya muncul pada elemen diagonal utama dan nol di posisi lainnya menunjukkan bahwa akurasi model mencapai 100% pada data uji yang digunakan. Visualisasi ini sekaligus memperkuat temuan kuantitatif sebelumnya, yang menunjukkan bahwa pipeline preprocessing, ekstraksi fitur, serta metode klasifikasi yang digunakan dalam penelitian ini mampu menghasilkan performa yang sangat optimal.

TABEL IV  
HASIL KLASIFIKASI SENTIMEN

Metric	Negative	Neutral	Positive	Accuracy	Macro Avg	Weighted Avg
Precision	1	1	1	—	1	1
Recall	1	1	1	—	1	1
F1-Score	1	1	1	1	1	1
Support	52	5	43	100	100	100

Evaluasi performa model dilakukan menggunakan beberapa metrik utama, yaitu precision, recall, dan F1-score, yang masing-masing dihitung untuk tiga kategori sentimen: negatif, netral, dan positif. Berdasarkan hasil yang ditampilkan pada Tabel IV, model menunjukkan performa yang sangat tinggi dengan nilai 1.00 pada seluruh metrik, baik untuk masing-masing kelas maupun pada rata-rata keseluruhan, yaitu macro average dan weighted average. Selain itu, nilai accuracy juga mencapai 100%, yang mengindikasikan bahwa seluruh data uji berhasil diklasifikasikan dengan tepat oleh model tanpa satu pun kesalahan prediksi. Nilai precision sebesar 1.00 menunjukkan bahwa setiap tweet yang diprediksi sebagai negatif, netral, maupun positif oleh model, seluruhnya benar sesuai dengan label aslinya. Demikian pula dengan nilai recall sebesar 1.00, yang mencerminkan kemampuan model dalam mendeteksi seluruh tweet yang benar-benar termasuk dalam masing-masing kategori, tanpa ada yang terlewat. F1-score, sebagai rata-rata harmonis dari precision dan recall, juga menunjukkan nilai sempurna di ketiga kelas, menandakan bahwa model memiliki performa yang konsisten dan stabil. Sementara itu, pada metrik support, diketahui bahwa jumlah data uji terdiri dari 52 tweet dengan sentimen negatif, 5 tweet dengan sentimen netral, dan 43 tweet dengan sentimen positif, dengan total keseluruhan sebanyak 100 tweet. Meskipun distribusi data tidak sepenuhnya seimbang, model tetap mampu menunjukkan performa yang sempurna di semua kategori tanpa adanya kecenderungan bias terhadap salah satu kelas. Hal ini menunjukkan bahwa pipeline yang diterapkan, mulai dari preprocessing, ekstraksi fitur TF-IDF, hingga klasifikasi dengan algoritma Naive Bayes, telah bekerja dengan sangat efektif dalam membangun sistem analisis sentimen yang akurat.

## V. KESIMPULAN

Artikel ini menyajikan analisis sentimen para suporter PSS Sleman terhadap peringkat klub di platform Twitter/X menggunakan algoritma Naive Bayes. Metodologi yang digunakan melibatkan beberapa tahapan, dimulai dari pengumpulan data menggunakan SNScrape dan Twitter API v2, dengan pemilihan kata kunci yang relevan seperti “PSS Sleman”, “peringkat”, dan “klasemen”. Setelah data berhasil dikumpulkan, dilakukan proses praproses data secara menyeluruh, yang meliputi tahapan cleaning, case folding, filtering, tokenizing, stemming, serta transformasi ke dalam bentuk numerik menggunakan metode TF-IDF. Data yang telah direpresentasikan secara numerik kemudian diklasifikasikan menggunakan algoritma Multinomial Naive Bayes. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan seluruh 100 data uji secara akurat, dengan distribusi hasil klasifikasi yaitu 52 tweet sebagai sentimen negatif, 5 tweet sebagai sentimen netral, dan 43 tweet sebagai sentimen positif. Visualisasi melalui confusion matrix menunjukkan bahwa seluruh prediksi tepat sasaran tanpa adanya kesalahan klasifikasi (False Positive = 0, False Negative = 0). Evaluasi lanjutan menggunakan metrik precision, recall, dan F1-score juga menghasilkan nilai 1.00 untuk ketiga kategori sentimen, serta nilai accuracy sebesar 100%, yang mengindikasikan kinerja model yang sangat optimal dan konsisten. Keberhasilan model dalam mengklasifikasikan opini publik secara tepat menunjukkan bahwa pendekatan Naive Bayes sangat cocok diterapkan dalam analisis sentimen berbasis media sosial, terutama dalam konteks persepakbolaan. Sistem ini memberikan peluang bagi manajemen klub untuk memahami persepsi publik secara lebih komprehensif, yang dapat digunakan dalam menyusun strategi komunikasi, meningkatkan interaksi dengan suporter, hingga mendukung pengambilan keputusan internal yang lebih responsif terhadap opini masyarakat. Namun, penelitian ini juga memiliki sejumlah keterbatasan. Salah satunya adalah ukuran data yang masih relatif kecil dan ketidakseimbangan distribusi antar kelas, di mana kelas netral jumlahnya jauh lebih sedikit dibandingkan dua kelas lainnya. Selain itu, model belum diuji

dengan metode cross-validation atau dataset dari periode waktu yang berbeda, sehingga generalisasi model belum dapat dipastikan secara menyeluruh. Untuk penelitian selanjutnya, disarankan untuk melakukan pengumpulan data dalam jumlah lebih besar dengan distribusi kelas yang lebih seimbang, menerapkan metode validasi tambahan, serta membandingkan performa algoritma Naive Bayes dengan algoritma lain seperti SVM, KNN, atau Random Forest, agar diperoleh pemahaman yang lebih luas terhadap pendekatan terbaik dalam analisis sentimen suporter di media sosial.

## DAFTAR PUSTAKA

- [1] S. Styawati, N. Hendrastuty dan A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan It*, 2021.
- [2] Y. Yusanto dan M. Akbar, "Analisis Sentimen Jogja Darurat Sampah di Twitter menggunakan Ekstraksi Fitur Model Word2VecdanConvolutional Neural Network," *TIN: Terapan Informatika Nusantara*, 2024.
- [3] A. A. Nurhasananda dan M. Akbar, "Analisis Sentimen Masyarakat Terhadap Kebijakan Undang-Undang Tentara Nasional Indonesia (UU TNI) Menggunakan Support Vector Machine," *KOMITEK: Jurnal Komputer, Informasi dan Teknologi*, 2025.
- [4] L. Rofiqi dan M. Akbar, "Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine," *JEKIN*, 2024.
- [5] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer*, 2021.
- [6] S. R. Cholil, T. Handayani, R. Prathivi dan T. Ardianita, "Implementasi Algoritma K-Nearest Neighbor(KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT(Indonesian Journal on Computer and Information Technology)*, 2021.
- [7] F. F. Irfani, M. Triyanto dan A. D. Hartanto, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, Dan Informatika)*, 2020.
- [8] D. Normawati dan S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-Sakti (Jurnal Sains Komputer Dan Informatika)*, 2021.
- [9] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *Jurnal Informatika Dan Teknik Elektro Terapan*, 2023.
- [10] F. V. Sari dan A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 2019.
- [11] Y. A. Singgalen, "Pemilihan Metode Dan Algoritma Dalam Analisis Sentimen Di Media Sosial: Sistematic Literature Review," *Journal Of Information Systems And Informatics*, 2021.
- [12] N. M. S. Hadna, P. I. Santosa dan W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. Dan Komunikasi*, 2016.
- [13] F. A. Larasati, D. E. Ratnawati dan B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana Dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2022.
- [14] A. Safira dan F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zonasi: Jurnal Sistem Informasi*, 2023.
- [15] A. D. A. Putra dan S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma Knn," *Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi)*, 2021.