

# OPINI PUBLIK DI MEDIA X TERHADAP PATRICK KLUIVERT SEBAGAI PELATIH TIMNAS INDONESIA YANG BARU DENGAN METODE NAÏVE BAYES

Bagus Febrianto<sup>\*1)</sup>, Maya Rini Handayani<sup>2)</sup>, Nur Cahyo Hendro Wibowo<sup>3)</sup>, Khothibul Umam<sup>4)</sup>

1. Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negri Walisongo, Semarang
2. Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negri Walisongo, Semarang
3. Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negri Walisongo, Semarang
4. Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negri Walisongo, Semarang

## Article Info

**Kata Kunci:** Opini; X; Patrick Kluivert; Naive Bayes; Klasifikasi

**Keywords:** Opinion; Application X; Patrick Kluivert; Naïve Bayes; Classification

## Article history:

Received 31 May 2025

Revised 15 June 2025

Accepted 5 July 2024

Available online 1 September 2025

## DOI :

<https://doi.org/10.29100/jipi.v10i3.8145>

\* Corresponding author.

Bagus Febrianto

E-mail address:

[2208096006@student.walisongo.ac.id](mailto:2208096006@student.walisongo.ac.id)

## ABSTRAK

Hasil dari percobaan ini digunakan untuk menentukan persepsi publik terhadap penunjukan Patrick Kluivert sebagai pelatih baru Tim Nasional Sepak Bola Indonesia menggunakan opini komentar Masyarakat Indonesia yaitu analisis sentimen. Dataset diperoleh dari media digital X, melalui teknik *crawling* dengan API X, dengan periode pengambilan data dari 8 Februari hingga 24 Maret 2025. Sebanyak 3.162 tweet yang relevan berhasil dihimpun dan selanjutnya dianalisis dengan metode algoritma Multinomial Naïve Bayes. Proses analisis ini mempunyai beberapa tingkat tahapan: praproses dataset teks, labelan sentimen, serta pemodelan klasifikasi. Sentimen dibagi ke dalam 3 kelas, yaitu Netral, positif, dan Negatif. Dengan uji coba ini kita tahu bahwa mayoritas opini publik bersifat netral (77,2%), disusul sentimen positif (11,8%) dan negatif (10,9%). Model mencapai akurasi sebesar 95,2% pada data latih dan 82,5% pada data uji. Namun, performa klasifikasi terhadap kelas minoritas (positif dan negatif) masih dapat ditingkatkan, khususnya dari aspek *recall*. Temuan ini mengindikasikan pentingnya upaya penyeimbangan data serta optimalisasi algoritma untuk meningkatkan akurasi model terhadap data yang tidak seimbang.

## ABSTRACT

The results of this experiment were used to determine public perception regarding the appointment of Patrick Kluivert as the new head coach of the Indonesian National Football Team by analyzing public opinion through sentiment analysis. The dataset was collected from the digital platform X using a crawling technique via the X API, with a data collection period spanning from February 8 to March 24, 2025. A total of 3,162 relevant tweets were gathered and subsequently analyzed using the Multinomial Naïve Bayes algorithm. The analysis process involved several stages, including text preprocessing, sentiment labeling, and classification modeling. The sentiments were categorized into three classes: neutral, positive, and negative. The findings revealed that the majority of public opinion was neutral (77.2%), followed by positive sentiment (11.8%) and negative sentiment (10.9%). The model achieved an accuracy of 95.2% on the training data and 82.5% on the test data. However, classification performance for the minority classes (positive and negative) remains an area for improvement, particularly in terms of recall. These results highlight the importance of data balancing efforts and further algorithm optimization to improve model accuracy when dealing with imbalanced datasets.

## I. PENDAHULUAN

PADA zaman digital ini, internet menjadi wadah paling dominan untuk menyampaikan opini dan perasaan Masyarakat terhadap berbagai isu, salah satunya tokoh publik dan peristiwa terkini. Salah satu tokoh yang belakangan menjadi sorotan adalah Patrick Kluivert, mantan pesepakbola internasional asal Belanda yang dikenal

luas berkat karier cemerlangnya di sejumlah klub besar Eropa. Keterlibatannya dalam berbagai aktivitas kepelatihan, termasuk penunjukannya sebagai pelatih Tim Nasional Indonesia, memicu beragam reaksi dari publik. [1]. Patrick Kluivert dikenal karena menggunakan pendekatan partisipatif dalam pelatihannya, sangat memungkinkan pemain dapat bekerja sama dalam mengambil keputusan strategi. Ini sangat mencerminkan kepemimpinan kontemporer yang sangat fleksibel di banding model autoritarian konvensional. Persepsi publik terhadap Kluivert memberikan gambaran menarik mengenai citra serta pengaruhnya di tengah Masyarakat, dan juga Masyarakat sangat berharap agar Timnas Indonesia lolos piala dunia. Dampak potensial dari penunjukkan ini bukan hanya pada aspek teknis di lapangan tapi juga pada pencitraan sepak bola nasional di mata dunia internasional[2].

Media sosial X merupakan salah satu platform berbasis microblogging, yaitu jenis media sosial yang memungkinkan pengguna untuk membagikan aktivitas dan opini mereka dalam bentuk tulisan singkat. Platform ini sering digunakan oleh pengguna untuk mengekspresikan perasaan terhadap berbagai hal, baik dalam bentuk pujian maupun kritik yang sering kali bernuansa emosional. Melalui X, pengguna bisa menyampaikan pendapat, berbagi informasi, memberi motivasi, berdiskusi tentang isu tertentu, hingga mengekspresikan kebahagiaan dan emosi lainnya melalui status dan tweet. Dalam penelitian ini, platform X dipilih sebagai sumber data karena memiliki beberapa keunggulan, seperti kemampuannya menampilkan opini secara real-time, fitur hashtag yang mempermudah pengelompokan topik, serta format teks yang ringkas sehingga lebih mudah diproses dan dianalisis dalam klasifikasi sentimen[3].

Untuk memahami opini publik tersebut, analisis sentimen menjadi salah satu pendekatan yang relevan, khususnya di bidang NLP. Namun, dalam konteks bahasa Indonesia, tantangan seperti penggunaan slang, singkatan, serapan dari bahasa daerah, serta ironi dan konteks tersembunyi sering kali melemahkan performa model terutama dalam mendeteksi sentimen negatif atau minoritas[4]. Penelitian ini bertujuan mengidentifikasi dan mengelompokkan opini atau perasaan seseorang terhadap suatu topik melalui data teks, seperti unggahan media sosial. Naïve Bayes menjadi pilihan utama untuk penelitian ini. Karena memiliki kemampuan yang sederhana namun efisien dalam memproses data dalam jumlah besar. Dengan mengandalkan Teori klasifikasi Bayes dan yakin disetiap fitur dalam data saling independen, algoritma ini dapat melakukan klasifikasi secara cepat dan cukup akurat.

Efektivitas algoritma Naïve Bayes dalam analisis sentimen telah banyak dibuktikan melalui berbagai penelitian. Kita ambil contoh dari penemuan yang dilakukan Priansyah dan Sutabri (2024), yang menerapkan algoritma ini untuk menganalisis sentimen publik di Twitter terkait hasil Pemilu Indonesia 2024. Hasil penelitian tersebut bisa kita lihat bahwa klasifikasi dengan Naïve Bayes memperoleh hasil yang cukup baik. Kemampuannya dalam menangani data yang besar dan beragam menjadikan algoritma ini sebagai salah satu pilihan yang tepat untuk menganalisis di media internet (sosial)[5].

Sejumlah hasil sebelumnya telah menunjukkan keberhasilan implementasi klasifikasi Naïve Bayes untuk menganalisis opini berbagai berita sosial. Sebagai contoh, studi oleh Ramadhani et al. (2021) meneliti respons rakyat Indonesia dengan adanya covid-19 di twitter dan melaporkan bahwa metode Naïve Bayes mampu mengklasifikasikan sentimen publik dengan akurasi mencapai 87% [6]. penelitian lain yang dilakukan oleh Sasmita et al. (2022) mengevaluasi komentar masyarakat terhadap kebijakan PPKM di Indonesia. Dalam studi tersebut, algoritma Naïve Bayes yang dikombinasikan dengan berbagai library pemrosesan teks, seperti NLTK dan kamus khusus, berhasil mencapai akurasi terbaik sebesar 77,2%. Percobaan menggunakan metode *5-fold cross-validation* menunjukkan hasil evaluasi yang cukup konsisten, di mana akurasi, presisi, recall, dan f-measure berada di kisaran 75–77%, tergantung pada library yang digunakan. Temuan ini menunjukkan bahwa pemilihan alat bantu pengolahan teks juga turut memengaruhi performa model dalam klasifikasi sentimen[7].

Dalam dunia pendidikan, Naive Bayes telah berhasil kinerjanya dalam menganalisis sentimen melalui berbagai penelitian. Salah satunya adalah studi oleh Zusrotun et al. (2022) yang menggunakan algoritma ini untuk mengkaji sentimen publik terhadap pembelajaran daring di Twitter. Hasil awal menunjukkan tingkat akurasi sebesar 74,08%. Untuk meningkatkan keandalan hasil, penelitian tersebut dengan penerapannya menggunakan K-fold Cross Validation menghasilkan nilai k yaitu 15, yang kemudian menghasilkan peningkatan akurasi menjadi 76,39%[8]. Sementara itu, penelitian oleh Putri et al. (2024) menganalisis opini masyarakat terhadap Gerakan Muhammadiyah melalui data media sosial. Menggunakan Naïve Bayes klasifikasi, mereka mencapai akurasi sebesar 87,5%, yang menegaskan efektivitas metode ini dalam klasifikasi sentimen terhadap topik sosial keagamaan. Namun demikian, tantangan tetap ada, terutama dalam mengolah teks yang tidak terstruktur, penggunaan bahasa gaul (*slang*), serta kompleksitas konteks dalam kalimat. Hal ini menunjukkan perlunya pengembangan lebih lanjut, baik dari sisi pre-processing maupun pemodelan, guna meningkatkan pemahaman konteks dan akurasi klasifikasi.[9].

Penelitian milik Apriyani . (2022) kita tahu bahwa Naïve dapat diterapkan secara efektif untuk analisis pendapat publik terhadap penyebaran berita hoaks. Dalam studi tersebut, algoritma ini digunakan untuk mengklasifikasikan opini publik terkait isu hoaks mengenai presiden tiga periode dengan pembagian sentimen positif dan negatif. Hasil pengujian model mampu mencapai akurasi sebesar 88%, dengan precision sebesar 94% dan recall sebesar 90%

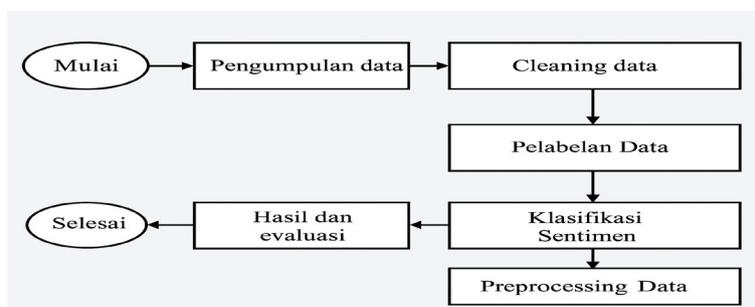
menggunakan skema pengujian data 50:50. Mayoritas sentimen publik dalam kasus tersebut cenderung positif terhadap isu yang beredar.[10]. Sementara itu, Pratama (2023) dalam sebuah kajian literatur terkait analisis sentiment menggunakan Naïve Bayes menyimpulkan bahwa algoritma ini secara umum mampu mengklasifikasikan teks dengan baik ke dalam kategori positif, negatif, atau netral. Temuan tersebut memperkuat bukti bahwa Naïve Bayes merupakan salah satu metode yang andal dalam pengolahan data teks dan opini publik secara efisien.

Berdasarkan kajian-kajian sebelumnya, terlihat bahwa metode Naïve Bayes memiliki potensi besar dalam memahami persepsi masyarakat terhadap berbagai isu sosial dan tokoh publik. Namun, hingga saat ini belum ditemukan banyak penelitian yang secara akurat membahas persepsi publik terhadap sosok Patrick Kluivert melalui media sosial. Dengan ini, penelitian hadir untuk mengisi kekosongan tersebut dengan menerapkan metode Naïve Bayes guna menganalisis sentimen masyarakat terhadap penunjukan Kluivert sebagai pelatih tim nasional Indonesia.

## II. METODE PENELITIAN

### A. Metode (Penelitian)

Penelitian ini mengadopsi metode analisis sentimen dengan memanfaatkan *Naïve Bayes Klasifikasi* sebagai teknik klasifikasi utama. Penelitian ini memiliki tujuan untuk mengetahui arah dan kecenderungan opini publik terhadap Patrick Kluivert setelah ia diumumkan sebagai pelatih baru Tim Nasional Sepak Bola Indonesia. Sentimen public dikelompokkan menjadi 3 kelas utama yaitu : sentimen Netral, positif, dan negatif.[11]. Ada beberapa tahapan untuk mencapai tujuan tersebut Sumber data, Pra-prosesing data, Pelabelan, Klasifikasi sentiment.



Gambar 1. Tahap Penelitian

### 1) Sumber Dataset

Dataset dikumpulkan dengan memanfaatkan Application Programming Interface (API) dari Twitter melalui bantuan library Tweepy yang berjalan pada bahasa pemrograman Python. Tahap ini bertujuan untuk memperoleh tweet yang secara spesifik memuat kata kunci "Patrick Kluivert", mengingat fokus utama penelitian adalah mengamati opini publik terhadap penunjukan Kluivert sebagai pelatih baru Timnas Indonesia. Proses crawling dilakukan secara bertahap selama periode 8 Februari hingga 24 Maret 2025, dengan mempertimbangkan volume dan kecepatan publikasi tweet yang muncul secara real-time di platform tersebut. Kriteria seleksi digunakan dalam proses *crawling* untuk memastikan relevansi dan kualitas data, yaitu hanya tweet orisinal yang diambil, sementara *retweet* dan *reply* dikecualikan. Hal itu dilakukan untuk menghindari duplikasi dan memastikan data merepresentasikan opini individu secara langsung, sehingga kualitas dataset lebih tinggi dan bebas noise[12]. Untuk menjaga kualitas dan relevansi data, hanya tweet yang menggunakan bahasa Indonesia yang dipertahankan dalam dataset. Sementara itu, tweet berbahasa asing atau yang tidak mengandung nilai analitis disaring dan dihapus[13].

Setelah proses *crawling* selesai, seluruh tweet yang terkumpul akan tersimpan dalam format CSV (Comma-Separated Values). Format csv dipilih karena fleksibel dan kompatibel dengan berbagai platform analisis data, serta memudahkan dalam proses *preprocessing* dan pemodelan klasifikasi. Data disusun secara sistematis dan mencakup beberapa atribut penting seperti isi teks, waktu unggahan (timestamp), dan metadata relevan lainnya. Seluruh data yang telah tersimpan kemudian digunakan dalam tahap *preprocessing*, yang mencakup pembersihan teks, normalisasi kata, serta pelabelan otomatis sebelum akhirnya dimasukkan ke dalam proses klasifikasi menggunakan algoritma Naïve Bayes. Penggunaan Twitter sebagai sumber data juga didukung oleh temuan dari penelitian Handayani et al. (2023), yang menyatakan bahwa Twitter merupakan platform yang kaya akan opini publik dan sangat efektif digunakan sebagai basis data untuk penelitian analisis sentimen berbasis teks[14].

### B. Pre-processing Data

Pada analisis ini, proses pengumpulan data dilakukan dengan mengekstraksi tweet yang memuat kata kunci "Patrick Kluivert" selama periode waktu tertentu. Melalui pencarian tersebut, berhasil dikumpulkan sebanyak 3.162

tweet, yang kemudian disimpan dalam format CSV agar memudahkan tahap analisis lanjutan. Sebelum masuk ke proses analisis sentimen, data mentah hasil crawling terlebih dahulu melalui tahap pra-pemrosesan. Tahap ini mencakup *cleaning* dataset dari karakter-karakter yang tidak sesuai, normalisasi teks, serta pelabelan data dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses ini penting untuk memastikan bahwa data yang dianalisis berada dalam kondisi bersih, terstruktur, dan siap untuk dimasukkan ke dalam model klasifikasi. Pendekatan ini sejalan dengan metodologi yang digunakan oleh Basir et al. (2025) dalam studi berjudul "Analysis of Electronic Wallet User Sentiment on Twitter (X) Social Media Using the Naïve Bayes Classifier Algorithm". Dalam penelitian tersebut, mereka juga memanfaatkan teknik crawling untuk mengumpulkan dataset dari Twitter dan menggunakan klasifikasi Naïve Bayes. Hasilnya menunjukkan tingkat akurasi yang memadai, yang memperkuat efektivitas pendekatan ini dalam analisis data teks dari media sosial[15].

### C. CaseFolding

Proses awal dalam prapemrosesan teks yang bertujuan untuk menyamakan format penulisan huruf dengan cara mengubah keseluruhan kalimat jadi huruf yang kecil. Langkah ini penting didalam dataset mentah, penggunaan huruf kapital sering kali tidak konsisten—misalnya, kata "Timnas" dan "timnas" dianggap berbeda jika tidak disera-gamkan. Dengan melakukan case folding, model analisis dapat membaca teks tanpa memperlmasalahakan perbedaan antara huruf besar dan kecil, sehingga meningkatkan akurasi pengolahan data.

TABEL 1.  
CASE FOLDING

Tweet	Case folding
"Pelatih Tim Nasional Indonesia Patrick Kluivert membuka peluang untuk melakukan rotasi pemain pada pertandingan putaran ketiga kualifikasi Piala Dunia 2026 zona Asia melawan Bahrain yang akan berlangsung di Stadion Utama Gelora Bung Karno <a href="https://t.co/g1LeqXHqW5">https://t.co/g1LeqXHqW5</a> "	"pelatih tim nasional indonesia patrick kluivert membuka peluang untuk melakukan rotasi pemain pada pertandingan putaran ketiga kualifikasi piala dunia 2026 zona asia melawan bahrain yang akan berlangsung di stadion utama gelora bung karno <a href="https://t.co/g1leqxhqw5">https://t.co/g1leqxhqw5</a> "
"Pelatih Timnas Indonesia Patrick Kluivert meminta agar anak asuhnya tampil menyerang sejak menit awal saat melawan Bahrain. #TimnasDay <a href="https://t.co/FLyuv8hobX">https://t.co/FLyuv8hobX</a> <a href="https://t.co/zqLcrJEI8h">https://t.co/zqLcrJEI8h</a> "	"pelatih timnas indonesia patrick kluivert meminta agar anak asuhnya tampil menyerang sejak menit awal saat melawan bahrain. #timnasday <a href="https://t.co/flyuv8hobx">https://t.co/flyuv8hobx</a> <a href="https://t.co/zqlcrjei8h">https://t.co/zqlcrjei8h</a> "

#### 1) Cleansing

Merupakan tahap penting dalam praproses untuk membersihkan dataset dari elemen-elemen yang tidak relevan. Pada proses ini, berbagai karakter seperti tautan, *username* atau *mention*, tagar, retweet, hingga simbol atau gambar dihapus dari teks. Pembersihan ini dilakukan agar data yang dianalisis menjadi lebih bersih, terstruktur, dan fokus pada isi pesan utama, sehingga tidak mengganggu proses klasifikasi sentimen di tahap selanjutnya

TABEL 2.  
CLEANSING

Tweet	Cleansing
"pelatih tim nasional indonesia patrick kluivert membuka peluang untuk melakukan rotasi pemain pada pertandingan putaran ketiga kualifikasi piala dunia 2026 zona asia melawan bahrain yang akan berlangsung di stadion utama gelora bung karno <a href="https://t.co/g1leqxhqw5">https://t.co/g1leqxhqw5</a> "	"pelatih tim nasional indonesia patrick kluivert membuka peluang untuk melakukan rotasi pemain pada pertandingan putaran ketiga kualifikasi piala dunia zona asia melawan bahrain yang akan berlangsung di stadion utama gelora bung karno"
"pelatih timnas indonesia patrick kluivert meminta agar anak asuhnya tampil menyerang sejak menit awal saat melawan bahrain. #timnasday <a href="https://t.co/flyuv8hobx">https://t.co/flyuv8hobx</a> <a href="https://t.co/zqLcrJEI8h">https://t.co/zqLcrJEI8h</a> "	"pelatih timnas indonesia patrick kluivert meminta agar anak asuhnya tampil menyerang sejak menit awal saat melawan bahrain"

#### 2) Tokenizing

Setelah tahap pembersihan selesai, proses dilanjutkan dengan *tokenizing*. Pada langkah ini, teks akan dipecah menjadi satuan-satuan kecil, seperti kata atau frasa, dengan cara menghilangkan tanda baca serta pemisah lainnya. Tujuannya adalah untuk mengubah kalimat menjadi kumpulan kata-kata yang berdiri sendiri, sehingga memudahkan dalam analisis lebih lanjut, terutama dalam klasifikasi atau penghitungan frekuensi kata. Dalam proses tokenizing menggunakan Library NLTK (*Natural Language Toolkit*) yang disesuaikan dengan karakteristik teks dari media sosial, termasuk slang, serapan Bahasa daerah, dan singkatan khas X. Seluruh proses ini dilakukan secara otomatis melalui pipeline preprocessing untuk memastikan data bersih, terstruktur, dan siap dimasukkan ke dalam model klasifikasi sentiment.

TABEL 3.  
TOKENIZING

Tweet	Tokenizing
"pelatih tim nasional indonesia patrick kluivert membuka peluang untuk melakukan rotasi pemain pada pertandingan putaran ketiga kualifikasi piala dunia 2026 zona asia melawan bahrain yang akan berlangsung di stadion utama gelora bung karno"	["pelatih", "tim", "nasional", "indonesia", "patrick", "kluivert", "membuka", "peluang", "untuk", "melakukan", "rotasi", "pemain", "pada", "pertandingan", "putaran", "ketiga", "kualifikasi", "piala", "dunia", "zona", "asia", "melawan", "bahrain", "yang", "akan", "berlangsung", "di", "stadion", "utama", "gelora", "bung", "karno"]
"pelatih timnas indonesia patrick kluivert meminta agar anak asuhnya tampil menyerang sejak menit awal saat melawan bahrain"	["pelatih", "timnas", "indonesia", "patrick", "kluivert", "meminta", "agar", "anak", "asuhnya", "tampil", "menyerang", "sejak", "menit", "awal", "saat", "melawan", "bahrain"]

### 3) Stopword Removal

Pada tahapan ini dilakukan penghilangan kata yang umum dan muncul tapi enggan memberikan kontribusi langsung terhadap makna teks, seperti “yang”, “pada”, “untuk”, dan sejenisnya. Kata-kata tersebut biasanya tidak memengaruhi konteks sentimen dan hanya menambah beban proses analisis. Proses ini dapat dilakukan dengan memilih proses algoritma berbasis *stoplist* dengan menghapus kata yang tidak berarti, atau sebaliknya digunakan *wordlist* supaya hanya kata yang penting untuk di perlukan analisis.

TABEL 4.  
STOPWORD REMOVAL

Tweet	Stopword Removal
["pelatih", "tim", "nasional", "indonesia", "patrick", "kluivert", "membuka", "peluang", "untuk", "melakukan", "rotasi", "pemain", "pada", "pertandingan", "putaran", "ketiga", "kualifikasi", "piala", "dunia", "zona", "asia", "melawan", "bahrain", "yang", "akan", "berlangsung", "di", "stadion", "utama", "gelora", "bung", "karno"]	["pelatih", "tim", "nasional", "indonesia", "patrick", "kluivert", "membuka", "peluang", "melakukan", "rotasi", "pemain", "pertandingan", "putaran", "ketiga", "kualifikasi", "piala", "dunia", "zona", "asia", "melawan", "bahrain", "berlangsung", "stadion", "utama", "gelora", "bung", "karno"]
["pelatih", "timnas", "indonesia", "patrick", "kluivert", "meminta", "agar", "anak", "asuhnya", "tampil", "menyerang", "sejak", "menit", "awal", "saat", "melawan", "bahrain"]	["pelatih", "timnas", "indonesia", "patrick", "kluivert", "meminta", "anak", "asuhnya", "tampil", "menyerang", "menit", "awal", "melawan", "bahrain"]

### 4) Stemming

Pada tahap **stemming**, setiap kata dalam teks diubah ke bentuk dasarnya dengan cara menghapus imbuhan, baik di awal, tengah, maupun akhir kata. Proses ini bertujuan untuk menyamakan berbagai variasi kata agar diperlakukan sebagai satu entitas yang sama. Pada proses stemming ini kita menggunakan library Sastrawi, yang sudah tersedia untuk proses morfologi bahasa Indonesia secara efisien. Library Sastrawi digunakan untuk mengembalikan kata ke bentuk dasar secara akurat.

TABEL 5.  
STEMMING

Tweet	Stemming
["pelatih", "tim", "nasional", "indonesia", "patrick", "kluivert", "membuka", "peluang", "melakukan", "rotasi", "pemain", "pertandingan", "putaran", "ketiga", "kualifikasi", "piala", "dunia", "zona", "asia", "melawan", "bahrain", "berlangsung", "stadion", "utama", "gelora", "bung", "karno"]	["latih", "tim", "nasional", "indonesia", "patrick", "kluivert", "buka", "luang", "laku", "rotasi", "main", "tanding", "putar", "tiga", "kualifikasi", "piala", "dunia", "zona", "asia", "lawan", "bahrain", "langsung", "stadion", "utama", "gelora", "bung", "karno"]
["pelatih", "timnas", "indonesia", "patrick", "kluivert", "meminta", "anak", "asuhnya", "tampil", "menyerang", "menit", "awal", "melawan", "bahrain"]	["latih", "timnas", "indonesia", "patrick", "kluivert", "minta", "anak", "asuh", "tampil", "serang", "menit", "awal", "lawan", "bahrain"]

## D. Pelabelan Sentimen

Tweet yang berhasil dikumpulkan dalam penelitian ini diberi label sentimen secara otomatis dengan pustaka (library) Python untuk dapat memberikan penilaian sentimen terhadap setiap tweet atau cuitan di platform X (Twitter). Tweet yang dikumpulkan dalam penelitian ini menggunakan model pre-trained berbasis Transformer melalui pustaka (library) Python. Model yang digunakan adalah *Indonesian RoBERTa Base Sentiment Classifier*, yang mampu mengklasifikasikan sentimen menjadi tiga kategori: negatif, netral, dan positif [16]. Untuk meningkatkan konsistensi klasifikasi dan meminimalkan potensi bias anotasi, Proses pelabelan ini turut melibatkan dua annotator, yang secara terpisah melakukan verifikasi terhadap hasil pelabelan otomatis. Dalam studi tersebut, hasil pelabelan otomatis kemudian diklasifikasikan lebih lanjut menggunakan algoritma Naïve Bayes [17]. Pada penelitian tersebut kita menggunakan metode ini untuk melakukan klasifikasi sentimen secara otomatis tanpa memerlukan intervensi manual, sekaligus menjaga efisiensi dan akurasi pemrosesan data dalam skala besar

## E. Algoritma Klasifikasi

Pada percobaan ini kita menggunakan algoritma Multinomial Naïve Bayes, salah satu jenis dari Naïve Bayes yang dianggap paling sesuai untuk analisis teks, khususnya pada data media sosial seperti tweet. Cara kerja algoritma ini ialah menghitung probabilitas kemunculan kata didalam masing-masing dari 3 sentimen (netral, positif, negatif), dengan keyakinan untuk fitur atau kata bersifat berdiri sendiri satu sama lain. Pendekatan seperti ini dikenal efisien untuk klasifikasi data teks yang bersifat singkat dan informatif, seperti unggahan Twitter. Dalam proses pelatihan model, sebanyak 80% data digunakan sebagai data latih, dan sisanya 20% digunakan sebagai data uji. Untuk mengevaluasi performa model, digunakan beberapa metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang mendapatkan gambaran komprehensif mengenai sejauh mana pemodelan mampu mengklasifikasikan dataset secara tepat. Penggunaan algoritma ini sejalan dengan temuan Lahase et al. (2021) yang mengaplikasikan Multinomial untuk menganalisis pada twitter. Penelitian tersebut juga mengandalkan data yang dikumpulkan melalui API Twitter dan diproses menggunakan pendekatan serupa, menunjukkan bahwa metode ini mampu mengklasifikasikan sentimen dengan cukup akurat dan efisien tanpa perlu intervensi manual [18].

Selain itu, hasil eksperimen dalam studi ini juga menunjukkan bahwa Naïve Bayes memberikan performa yang kompetitif dibandingkan dengan algoritma lain dalam klasifikasi sentimen, khususnya pada topik yang melibatkan opini publik seperti penunjukan pelatih tim nasional. Hal ini semakin memperkuat alasan pemilihan algoritma ini sebagai pendekatan utama dalam memahami respon masyarakat terhadap Patrick Kluivert di media sosial

Naïve Bayes didasarkan pada **Teorema Bayes**, yang secara umum dituliskan sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan:

- $P(C|X)$  adalah probabilitas dari kelas  $C$  yang diberikan fitur  $X$ ,
- $P(X|C)$  adalah probabilitas fitur  $X$  muncul dalam kelas  $C$ ,
- $P(C)$  adalah probabilitas awal dari kelas  $C$  (prior),
- $P(X)$  adalah probabilitas dari fitur  $X$  secara keseluruhan (konstan dan bisa diabaikan dalam perbandingan antar kelas).

Dalam konteks **klasifikasi teks**, rumus tersebut disesuaikan menjadi:

$$\hat{C} = \arg \max_{C_i \in C} P(C_i) \prod_{j=1}^n P(x_j|C_i)$$

Keterangan tambahan:

- $\hat{C}$  adalah kelas prediksi (positif, netral, atau negatif),
- $C_i$  adalah himpunan semua kelas,
- $x_j$  adalah fitur berupa kata dalam dokumen,
- $n$  adalah jumlah fitur dalam dokumen.

Gambar 2. Rumus Pemodelan

## F. Evaluasi Model

Evaluasi terhadap performa menggunakan metode confusion matrix, dengan tujuan memperoleh metrik yang penting seperti akurasi, presisi, recall, dan **F1-score**. Metrik evaluasi ini dapat digunakan mengukur seberapa jauh model mampu memberikan hasil klasifikasi dengan benar, terutama pada data uji. Confusion matrix memungkinkan kita menilai ketepatan model secara menyeluruh, termasuk kemampuannya dalam mendeteksi kelas sentimen tertentu seperti positif, negatif, dan netral. Akurasi digunakan untuk mengetahui proporsi prediksi yang sesuai dengan label sebenarnya, sedangkan presisi mengukur ketepatan model dalam mengidentifikasi kelas positif. Recall menilai seberapa jauh data yang di tangkap model yang termasuk dalam kelas tertentu, dan F1-score memberi keseimbangan antara presisi dengan recall, terutama berguna untuk konteks data yang tidak seimbang. Pendekatan ini didukung oleh penelitian yang dipublikasikan dalam Jurnal Informatika Teknologi dan Sains (JINTEKS), yang menekankan pentingnya confusion matrix dalam menilai kinerja model klasifikasi secara komprehensif. Studi tersebut menyatakan bahwa evaluasi menggunakan confusion matrix sangat membantu dalam

menyelesaikan persoalan klasifikasi, karena mampu mengukur performa model secara menyeluruh berdasarkan hasil prediksi setiap kelas [19].

Confusion matrix menyajikan empat elemen utama yang digunakan untuk mengevaluasi kinerja model klasifikasi: true positif (tp), true negative (tn), false positif (fp), dan false negative (fn)

1. Accuracy (Akurasi)

Mengukur keseluruhan prediksi yang benar dibandingkan seluruh data:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision (Presisi)

Mengukur seberapa akurat model dalam memprediksi kelas tertentu (misalnya positif):

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall (Sensitivitas)

Mengukur kemampuan model menemukan seluruh data positif yang sebenarnya:

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score

Rata-rata harmonis antara precision dan recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Gambar 3. Rumus ConfusionMatrix

Penggunaan *confusion matrix* beserta akurasi, precision, recall, dan F1-score memiliki peran sangat penting dalam menilai kinerja pengklasifikasian, terutama ketika berhadapan dengan data yang tidak seimbang. Dalam kasus di mana distribusi label tidak merata—misalnya ketika satu kelas mendominasi jumlah data—mengandalkan akurasi saja tidak cukup untuk menggambarkan performa model secara menyeluruh. Dalam situasi seperti ini, metrik tambahan seperti presisi dan recall menjadi sangat krusial karena mampu memberikan gambaran yang lebih detail mengenai bagaimana model menangani masing-masing kelas.

Sebagaimana dijelaskan oleh Dyllan Bagus Siswanto (2023) dalam penelitiannya, *confusion matrix* digunakan sebagai dasar dalam menghitung metrik-metrik tersebut. Hasil dari evaluasi ini memberikan wawasan mendalam mengenai kemampuan model dalam melakukan prediksi, baik dalam mengidentifikasi sentimen netral, positif, negatif, [20]. Evaluasi ini juga membantu peneliti dalam mengidentifikasi kekuatan dan kelemahan model secara spesifik, sehingga dapat menjadi dasar perbaikan model ke depan. Dalam konteks analisis sentimen publik, di mana opini masyarakat cenderung beragam dan tidak selalu terdistribusi secara merata, penggunaan metrik evaluasi yang komprehensif menjadi sangat relevan untuk memastikan bahwa model mampu bekerja secara adil dan akurat terhadap seluruh kelas sentiment

### III. HASIL PEMBAHASAN

#### 1. Pembahasan Dari Hasil

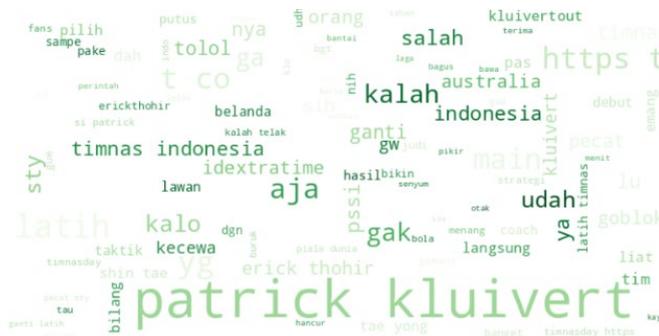
Dari percobaan ini kita tahu bahwa kecenderungan sentimen publik terhadap penunjukan Patrick Kluivert sebagai pelatih baru Tim Nasional Indonesia dengan menggunakan data dari platform media internet X. Pengumpulan dataset menggunakan teknik crawling dengan Twitter API selama periode waktu tertentu, dan menghasilkan sebanyak 3.162 tweet yang relevan dengan topik penelitian. Seluruh tweet yang berhasil dikumpulkan lalu tersimpan dengan format csv agar memudahkan dalam proses pengolahan data tahapan selanjutnya. Sebelum dianalisis, data mentah tersebut melalui serangkaian proses pra-proses (preprocessing). Tahapan ini mencakup pembersihan data yang di peroleh dari elemen yang berguna seperti simbol, URL, dan tanda baca, kemudian dilanjutkan dengan normalisasi teks serta pelabelan otomatis memperoleh 3 kategori sentimen, yaitu Netral, positif, dan, negatif. Pendekatan ini sejalan dengan penelitian yang dilakukan oleh Subasar (2024) dalam artikelnya berjudul "*Sentiment Analysis of Twitter Users Ahead of the 2024 Election Using the Naive Bayes Method*". Dalam penelitian tersebut, data dikumpulkan dengan teknik serupa dan dianalisis menggunakan algoritma Naive Bayes, yang dinilai memberikan hasil klasifikasi yang cukup baik dalam konteks analisis sentimen [21].



Gambar 5. WordCloudNetral

c. Word Cloud Negatif (Hijau)

Di sentimen negatif menunjukkan kata seperti: "kluivertout", "salah", "tolol", "kalah", "sty", "gagal", dan "kecewa". Dari sini dapat ditarik kesimpulan bahwa sebagian publik menyuarakan kekecewaan atau keraguan terhadap kemampuan Kluivert, baik karena hasil pertandingan yang buruk, perbandingan dengan pelatih sebelumnya, atau anggapan bahwa penunjukan ini tidak tepat. Kata-kata bernada kasar seperti "tolol" dan ekspresi frustrasi lainnya memperkuat persepsi tersebut[23]. Kata "kluivertout" dan "sty" menunjukkan adanya perbandingan dengan pelatih sebelumnya, serta tuntutan agar kluivert segera diganti. Sementara itu, kata "tolol" dan "kecewa" mencerminkan luapan emosi dan kritik keras dari public. Jika dibandingkan dengan word cloud positif yang lebih banyak mengandung kata bersifat harapan seperti "menang" dan "bagus", maka word cloud negatif lebih terfokus pada hasil yang tidak memuaskan dan pesimisme. Analisis komparatif ini menunjukkan bahwa sentimen public sangat dipengaruhi oleh performa pertandingan, persepsi terhadap reputasi pelatih, dan ekspektasi terhadap timnas.

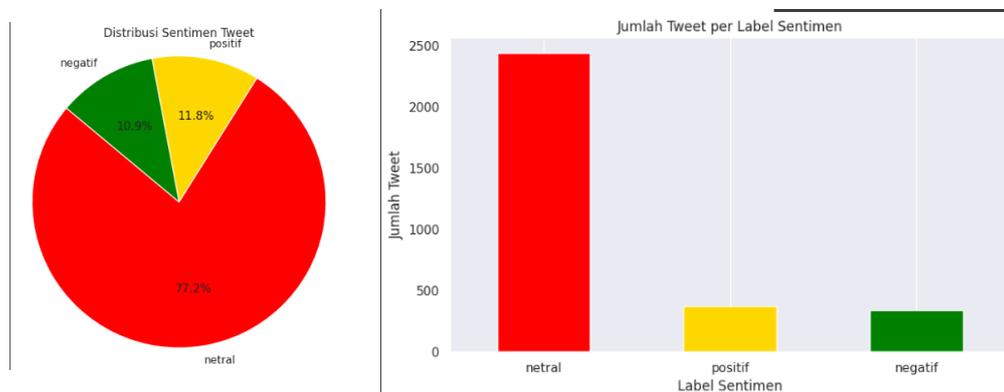


Gambar 6. Word Cloud Negatif

• **Distribusi Sentimen**

Berdasarkan proses preprocessing dan pelabelan otomatis, distribusi data sentimen terhadap Patrick Kluivert adalah sebagai berikut:

- Sentimen Netral: 77.2%
- Sentimen Positif: 11.8%
- Sentimen Negatif: 10.9%



Gambar 7. Grafik pie chart & Bar chart

Distribusi ini menunjukkan bahwa mayoritas masyarakat menunjukkan respon netral terhadap penunjukan Patrick Kluivert sebagai pelatih baru timnas, sementara opini positif sedikit lebih tinggi dibandingkan opini negatif. Berdasarkan grafik pie chart dan bar chart di atas dan informasi bahwa total data adalah 3.162 tweet, berikut adalah penjelasan dan perhitungan jumlah data per label sentimen secara akurat berdasarkan persentase yang tertera

TABEL 6. HASIL GRAFIK

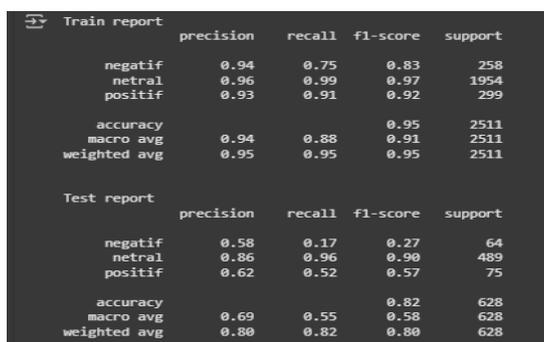
Label Sentimen	Persentase	Jumlah Tweet (≈)
Netral	77.2%	2.442 tweet (Merah)
Positif	11.8%	373 tweet (Kuning)
Negatif	10.9%	345 tweet (Hijau)
Total	100%	3.162 tweet

Berdasarkan visualisasi distribusi sentimen dari total 3.162 tweet, diketahui bahwa data didominasi oleh sentimen netral sebesar 77.2% atau sekitar 2.442 tweet. Dominasi ini dapat mencerminkan beberapa kemungkinan: (1) publik belum memiliki pendapat yang kuat, baik mendukung maupun menolak, terkait penunjukan Patrick Kluivert sebagai pelatih baru; atau

(2) model klasifikasi cenderung mengkategorikan tweet ambigu atau informatif sebagai netral. Ketidakseimbangan kelas ini juga dapat dipengaruhi oleh karakteristik bahasa Indonesia di media sosial, di mana banyak pernyataan bersifat informatif atau sekadar laporan tanpa muatan emosional yang jelas. Untuk memastikan bahwa dominasi sentimen netral bukan sekadar hasil bias model, validasi tambahan dengan data berlabel manual dapat dilakukan. Sementara itu, tweet dengan sentimen positif dan negatif masing-masing berjumlah sekitar 373 (11.8%) dan 345 (10.9%) tweet. Ketimpangan distribusi ini menunjukkan adanya class imbalance, yang penting untuk diperhatikan dalam proses pelatihan model klasifikasi karena dapat memengaruhi akurasi terhadap kelas minoritas.

Dalam penelitian ini, algoritma NaïveBayes digunakan agar pengelompokkan data menjadi ke dalam tiga jenis sentimen, yaitu Netral, positif, dan negatif. Proses klasifikasi dilakukan setelah model melalui tahap pelatihan dan pengujian, dengan hasil yang divisualisasikan melalui confusion matrix sebagai alat evaluasi kinerja model. Efektivitas penggunaan algoritma ini juga didukung oleh temuan Hidayah dan Dodiman (2022), yang kita lihat dengan NaiveBayes performanya lebih baik untuk tugas klasifikasi teks. Berdasarkan hasil studi mereka, algoritma ini mampu mencapai tingkat akurasi hingga 85%, menandakan bahwa Naïve Bayes cukup andal dalam menangani data teks yang kompleks, terutama dalam konteks analisis sentime[24].

### • Evaluasi Model Naïve Bayes



Train report				
	precision	recall	f1-score	support
negatif	0.94	0.75	0.83	258
netral	0.96	0.99	0.97	1954
positif	0.93	0.91	0.92	299
accuracy			0.95	2511
macro avg	0.94	0.88	0.91	2511
weighted avg	0.95	0.95	0.95	2511
Test report				
	precision	recall	f1-score	support
negatif	0.58	0.17	0.27	64
netral	0.86	0.96	0.90	489
positif	0.62	0.52	0.57	75
accuracy			0.82	628
macro avg	0.69	0.55	0.58	628
weighted avg	0.80	0.82	0.80	628

Gambar 8. Evaluasi Model

#### a. Hasil Pelatihan

Pada rumus yang sudah di jelaskan pada gambar di bagian Evaluasi Model mendapatkan hasil:

- Akurasi pelatihan: 95,2%
- Kinerja per kelas:

TABEL 7. PELATIHAN

Kelas	Precisi	Recall	F1
Netral	0.96	0.99	0.97
Positif	0.93	0.91	0.92
Negatif	0.94	0.75	0.83

#### b. Hasil Pengujian

Data uji ini, akurasi model mencapai 0.825 (82,5%). Namun, terdapat penurunan performa pada kelas negatif dan positif. Berikut detailnya:

TABEL 8. PENGUJIAN

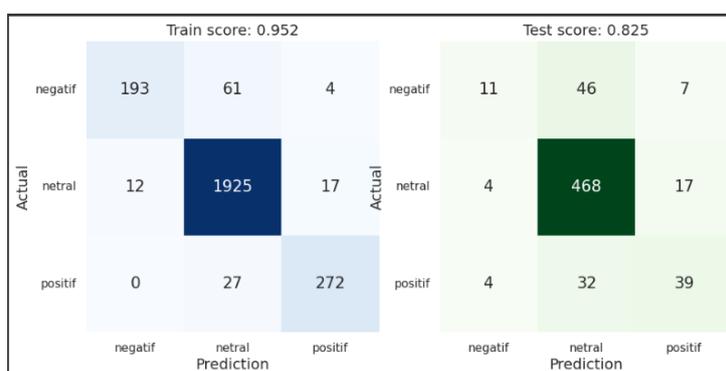
Kelas	Precisi	Recall	F1-Score
Netral	0.86	0.96	0.90
Positif	0.62	0.52	0.57
Negatif	0.58	0.17	0.27

Meskipun akurasi umum tinggi, recall untuk kelas negatif tergolong rendah 17% (0.17) yang menunjukkan model cenderung tidak sensitif terhadap deteksi sentimen negatif. Hal ini menunjukkan bahwa model cenderung tidak sensitive terhadap deteksi sentiment negatif. Salah satu penyebabnya adalah keterbatasan model Naïve Bayes dalam memahami konteks kalimat yang kompleks, seperti sarkasme atau ironi. Misalnya, kalimat “wah, mainnya bagus banget, sampai kebobolan 5 gol!” mungkin terklasifikasi positif karena adanya kata “bagus”, padahal konteks sebenarnya bernada negatif.

Dengan ini mengindikasikan bahwa model Naive Bayes sangat baik dalam mengklasifikasi data netral, namun perlu dilakukan penguatan untuk menangani ketidakseimbangan kelas (class imbalance), terutama pada data sentimen negatif. Penelitian oleh Humaidi dan Maulani (2024) menyatakan bahwa pada dataset yang tidak seimbang, model Naïve Bayes cenderung memiliki performa yang lebih rendah pada kelas minoritas, seperti sentimen negatif[25].

- **ConfusionMatrix**

Proses ini adalah salah satu cara evaluasi yang amat sangat berpengaruh dalam menilai kerja metode klasifikasi, khususnya dalam konteks pembelajaran mesin. Matriks ini menampilkan informasi dalam bentuk tabel yang memperlihatkan perbandingan antara hasil prediksi model dan label sebenarnya dari data uji. Pada struktur matriks tersebut, baris merepresentasikan nilai aktual (label sebenarnya), sedangkan kolom menunjukkan hasil prediksi dari model. Melalui confusion matrix, kita dapat mengidentifikasi dengan jelas jenis-jenis kesalahan yang dilakukan model, serta mengetahui seberapa sering kesalahan tersebut terjadi. Ini sangat membantu dalam menganalisis seberapa akurat model mengenali setiap kategori kelas yang ada. Berdasarkan hasil evaluasi, penggunaan ini sangat mempermudah perhitungan metrik evaluasi penting seperti akurasi, precision, recall, dan F1. Metrik seperti ini secara keseluruhan menampilkan gambaran yang lengkap untuk menemukan model mengklasifikasi data secara benar. Dalam konteks penelitian ini, confusion matrix terbukti menjadi alat evaluasi yang efektif, terlebih lagi ketika bekerja dengan data kompleks dan beragam seperti unggahan media social.



Gambar 9. Confusion Matrix

a. Data Latih (Train)

TABEL 9. DATA LATIH CONFUSION MATRIX

	Prediksi Negatif	Netral	Positif
Negatif	193	61	4
Netral	12	1925	17
Positif	0	27	272

b. Data Uji (Test)

TABEL 10. DATA UJI CONFUSION MATRIX

	Prediksi Negatif	Netral	Positif
Negatif	11	46	7
Netral	4	468	17
Positif	4	32	39

Berikut adalah confusion matrix dari hasil klasifikasi:

- Prediksi Netral paling akurat: 468 dari 489 tweet diklasifikasi dengan benar.
- Prediksi Positif masih perlu ditingkatkan: hanya 39 dari 75 yang berhasil diprediksi dengan benar.
- Prediksi Negatif masih rendah: hanya 11 dari 64 yang diklasifikasi secara tepat[26] .

#### • Analisis dan Implikasi

Ketimpangan jumlah data antar kelas (class imbalance) menjadi tantangan utama dalam penelitian ini, sebagaimana juga ditemukan oleh Sihombing & Simarmata yang menyatakan bahwa performa klasifikasi sentimen sangat dipengaruhi oleh proporsi data tiap kelas dalam dataset. Dalam studi mereka, dilakukan pendekatan resampling dan fine-tuning preprocessing untuk mengurangi bias prediksi terhadap kelas dominan.

Penelitian ini juga mengonfirmasi temuan Nugroho et al. (2022), di mana metode NaïveBayes menunjukkan prospek kerja yang amat baik untuk teks pendek dengan contoh tweet, khususnya dengan pengolahan bahasa Indonesia . Meskipun begitu, performa terhadap kelas minoritas seperti positif dan negatif dapat ditingkatkan melalui teknik seperti SMOTE atau penalized learning[27].

Temuan ini merefleksikan bahwa masyarakat belum memiliki pendapat yang cukup kuat, baik mendukung maupun menolak, terhadap Kluivert, dan isu ini masih berkembang. Berdasarkan Sartika et al. (2023), fluktuasi sentimen publik biasanya meningkat seiring dengan performa awal tokoh publik atau pejabat baru yang menimbulkan ekspektasi masyarakat[28].

#### IV. KESIMPULAN

Dengan penelitian ini berhasil melakukan analisis sentimen terhadap 3.162 tweet dengan kata kunci Patrick Kluivert menggunakan algoritma Naïve Bayes. Dari data yang dianalisis, diketahui bahwa mayoritas tweet bersentimen netral (77,2%), sementara tweet yang menunjukkan sentimen positif sebesar 11,8%, dan sentimen negatif sebanyak 10,9%. Distribusi data yang tidak seimbang ini memengaruhi performa model dalam klasifikasi, khususnya untuk sentimen negatif dan positif. Model Naïve Bayes bekerja dengan sangat baik pada data pelatihan, memiliki akurasi mencapai 95,2%, terutama dalam mengenali sentimen netral. Namun, saat diuji dengan data uji, akurasi menurun menjadi 82,5%, dan performa klasifikasi pada kelas negatif menunjukkan kelemahan, ditandai dengan nilai recall yang sangat amat rendah (0,17). Ini mengindikasikan bahwa model cenderung tidak sensitif dalam mengenali sentimen negatif.

Confusion matrix menunjukkan bahwa banyak tweet positif dan negatif diprediksi sebagai netral, memperkuat dugaan bahwa ketidakseimbangan kelas dalam data menyebabkan model menjadi bias terhadap label netral. Oleh karena itu, diperlukan metode penyeimbangan data seperti resampling atau pemberian bobot pada kelas minoritas untuk meningkatkan sensitivitas model terhadap sentimen minoritas, terutama negatif.

Hasil dari percobaan ini sangat sesuai dari yang sebelumnya yang menyatakan model NaïveBayes efektif dalam klasifikasi teks, namun perlu penyesuaian dalam kondisi data yang tidak seimbang. Pemanfaatan algoritma lain seperti Support Vector Machine (SVM) juga dapat menjadi alternatif dalam mengatasi kelemahan ini.

#### REFERENSI

- [1] M. Qadri, "Pengaruh Media Sosial Dalam Membangun Opini Publik," *Qaumiyah J. Huk. Tata Negara*, vol. 1, no. 1, pp. 49–63, 2020, doi: 10.24239/qaumiyah.v1i1.4.
- [2] M. Kepemimpinan, P. Kluivert, N. Menuju, P. Dunia, A. Strategi, and A. D. Aryasatya, "Manajemen Kepemimpinan Patrick Kluivert dalam Mempersiapkan Tim Nasional Menuju Piala Dunia : Analisis Strategi dan Implementasi," no. June, 2025.
- [3] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier," *SINTECH (Science Inf. Technol. J.)*, vol. 5, no. 2, pp. 227–233, 2022, doi: 10.31598/sintechjournal.v5i2.1241.
- [4] V. W. D. Thomas and F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218.
- [5] E. Priansyah and T. Sutabri, "IJM: Indonesian Journal of Multidisciplinary Analisis Sentimen Berbasis Naïve Bayes Pada Media Sosial Twitter Terhadap Hasil Pemilu Indonesia 2024," *IJM Indones. J. Multidiscip.*, vol. 2, pp. 128–138, 2024, [Online]. Available:

- <https://journal.csspublishing/index.php/ijm>
- [6] T. Ramadhani, Y. A. Sari, and E. Santoso, "Analisis Sentimen Masyarakat Indonesia terhadap Covid-19 pada Media Sosial Twitter menggunakan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5680–5686, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] A. B. Sasmita, B. Rahayudi, and L. Muflikhah, "Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1208–1214, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.
- [9] S. Azahra, N. Wangsa, and A. Khoirudin, "Analisis Sentimen Media Sosial X terhadap Gerakan Muhammadiyah Menggunakan Algoritma Naïve Bayes Analysis of Social Media Sentiment X towards the Muhammadiyah Movement Using the Naïve Bayes Algorithm," vol. 6, no. 1, pp. 9–17, 2025.
- [10] M. E. Apriyani *et al.*, "ANALISIS SENTIMEN BERITA HOAX MENGGUNAKAN NAÏVE," pp. 1–6, 2022.
- [11] I. Juventius, T. Gurning, P. P. Adikara, and R. S. Perdana, "Analisis Sentimen Dokumen Twitter menggunakan Metode Naïve Bayes dengan Seleksi Fitur GU Metric," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2169–2177, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12665>
- [12] S. Stefanni, Z. Zulfachmi, Z. Zulkipli, and A. Saputra, "Analisis Sentimen Pengguna X Terhadap Kebocoran Data Pribadi Menggunakan Algoritma Naïve Bayes Classifier," *J. Bangkit Indones.*, vol. 14, no. 1, pp. 32–40, 2025, doi: 10.52771/bangkitindonesia.v14i1.434.
- [13] A. Basir, T. Yudha, P. Muhamad, A. Abdillah, and A. Faizin, "Jurnal Informatika : Jurnal pengembangan IT Analysis of Electronic Wallet User Sentiment on Twitter ( x ) Social Media Using the Naïve Bayes Classifier Algorithm," vol. 10, no. 1, pp. 79–85, 2025, doi: 10.30591/jpit.v10i1.8180.
- [14] N. T. Luchia *et al.*, "Analysis of Twitter User Sentiments for the Aplikasi TikTok Application Using Naïve Bayes Classifier Algorithm Analisis Sentimen Pengguna Twitter Terhadap Aplikasi TikTok Menggunakan Algoritma Naïve Bayes Classifier," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, no. September 2016, pp. 100–104, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [15] J. A. Costales, C. M. De Los Santos, J. J. E. Catulay, and M. G. Albino, "Sentiment Analysis for Twitter Tweets: A Framework to Detect Sentiment Using Naïve Bayes Algorithm," *2022 4th Int. Conf. Comput. Commun. Internet, ICCCI 2022*, no. July, pp. 39–44, 2022, doi: 10.1109/ICCCI55554.2022.9850257.
- [16] T. D. Salma, M. F. Kurniawan, R. Darmawan, and A. Basri, "Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 757–764, 2025, doi: 10.35870/jtik.v9i2.3535.
- [17] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [18] A. Lahase, S. Maher, and S. Bhable, "Twitter Data analysis ( Classification ) for text using Multinomial Naive Bayes," vol. 9, no. 3, pp. 422–425, 2021.
- [19] W. A. Naseer, S. Sarwido, and B. B. Wahono, "Gradient Boosting Optimization with Pruning Technique for Prediction of Bmt Al-hikmah Permata Customer Data," *Jinteks*, vol. 6, no. 3, pp. 719–727, 2024.
- [20] D. B. Siswanto and D. Normawati, "Sistem Klasifikasi Monitoring dan Evaluasi Kelayakan Penerima Beasiswa UAD Menggunakan Algoritma Naïve Bayes," *J. SAINTEKOM*, vol. 13, no. 2, pp. 161–172, 2023, doi: 10.33020/saintekom.v13i2.428.
- [21] A. Subasar, "Sentiment Analysis of Twitter Users Ahead of the 2024 Election Using the Naive Bayes Method," vol. 04, pp. 1–8, 2024, doi: 10.31763/iota.v4i3.784.
- [22] U. Pembangunan, N. Veteran, and J. Timur, "1 , 2 , 3\*," vol. 3, 2024.
- [23] J. A. Rieuwpassa, S. Sugito, and T. Widiharh, "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Sentimen Ulasan Pengguna Aplikasi Netflix Pada Google Play," *J. Gaussian*, vol. 12, no. 3, pp. 362–371, 2024, doi: 10.14710/j.gauss.12.3.362-371.
- [24] N. Hidayah and Dodiman, "Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," *J. Akad. Pendidik. Mat.*, vol. 10, no. 1, pp. 8–15, 2024.
- [25] M. R. Humaidi and A. Maulani, "Klasifikasi Naïve Bayes Dan Confusion Matrix Pada Pengguna Aplikasi E-Commerce Di Play Store," vol. 8, no. 2, pp. 132–139, 2023.
- [26] T. D. Putra, E. Utami, and M. P. Kurniawan, "Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO)," *Explore*, vol. 13, no. 1, pp. 1–5, 2023, doi: 10.35200/ex.v11i2.13.
- [27] I. Ismail, M. A. Nugroho, and R. Sulistiyowati, "Sentiment Analysis Netizens on Social Media Twitter Against Indonesian Presidential Candidates in 2024 Using Naive Bayes Classifier Algorithm," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1611, 2023, doi: 10.30865/mib.v7i3.6536.
- [28] S. G. Alexander, A. T. Ananto, I. P. Adhitya, P. Mangku, and B. Liano, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Konten Deepfake Tokoh Publik," *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komputer)*, vol. 05, no. 02, pp. 95–102, 2023.