

KLASIFIKASI SENTIMEN NETIZEN TERHADAP PATRICK KLUIVERT DI PLATFORM X DENGAN METODE NAÏVE BAYES

Alif Al Fadhilla¹, Wahyu Pramusinto^{2*}, Hadidtyo Wisnu Wardani³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹aliefalfadhillah@gmail.com, ^{2*}wahyu.pramusinto@budiluhur.ac.id, ³hadidtyo.wardani@budiluhur.ac.id
(* : corresponding author)

Abstrak- Penunjukan Patrick Kluivert sebagai pelatih tim nasional Indonesia telah memicu beragam reaksi di media sosial X (*Twitter*), menjadikannya topik yang relevan untuk analisis sentimen. Penelitian ini bertujuan menganalisis dan memprediksi sentimen komentar netizen terhadap Patrick Kluivert menggunakan metode *text mining* dengan algoritma Naïve Bayes Classifier. Proses penelitian dimulai dengan pengumpulan 263 data komentar yang relevan dari platform X (*Twitter*). Data tersebut kemudian melalui tahap preprocessing yang terdiri dari *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*, serta pembobotan kata menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Model Naïve Bayes diuji menggunakan *Confusion Matrix* dengan pembagian data 80% latih dan 20% uji, dan berhasil mencapai tingkat akurasi sebesar 64,15%. Hasil analisis menunjukkan bahwa model memiliki performa baik dalam mengidentifikasi sentimen positif, namun kurang akurat pada sentimen negatif dan netral, yang diindikasikan karena adanya ketidakseimbangan jumlah data antar kelas sentimen. Penelitian ini menyimpulkan bahwa Naïve Bayes dapat diterapkan untuk memetakan sentimen publik, meskipun memiliki keterbatasan dalam menangani data yang tidak seimbang.

Kata Kunci: Analisis Sentimen, Klasifikasi Teks, *Text Mining*, *Naive Bayes*, X (*Twitter*), Timnas Indonesia

CLASSIFICATION OF NETIZEN SENTIMENT TOWARDS PATRICK KLUIVERT ON THE X PLATFORM USING THE NAÏVE BAYES METHOD

Abstract- The appointment of Patrick Kluivert as the coach of the Indonesian national team has sparked various reactions on social media platform X (*Twitter*), making it a relevant case study for sentiment analysis. This research aims to analyze and predict user comments' sentiment towards Patrick Kluivert using a text mining method with the Naïve Bayes algorithm. The research methodology began with collecting 263 relevant comment data from platform X (*Twitter*). The raw data then underwent a series of preprocessing phases consisting of *case folding*, *cleansing*, *tokenizing*, *stopword removal*, and *stemming* to standardize the text. Furthermore, the TF-IDF (*Term Frequency-Inverse Document Frequency*) method was used for word weighting to convert text data into a numerical representation. The Naïve Bayes Classifier algorithm was used to create the sentiment analysis classification model by splitting the data into training and testing sets. The model's testing and evaluation results using a *Confusion Matrix* showed that the Naïve Bayes classification model achieved an accuracy rate of 64.15% with an 80:20 ratio. Further analysis indicated that the model performed well in identifying positive sentiment but tended to be less accurate in classifying negative and neutral sentiments, which suggests an influence from the imbalanced amount of data between sentiment classes. This study concludes that the Naïve Bayes method can be applied to map public sentiment, despite its limitations in handling imbalanced data and the complexity of language on social media.

Keywords: Sentiment Analysis, Text Classification, Text Mining, Naive Bayes, X (*Twitter*), Timnas Indonesia

1. PENDAHULUAN

Pesatnya perkembangan teknologi informasi telah mengubah cara masyarakat menyampaikan opini. Media sosial, khususnya platform X (sebelumnya *Twitter*), kini menjadi kanal utama untuk mengemukakan pendapat terhadap isu-isu aktual, termasuk dalam ranah olahraga. Penunjukan Patrick Kluivert sebagai pelatih Timnas Indonesia menimbulkan reaksi publik yang bervariasi, mulai dari dukungan hingga kritik. Informasi ini penting untuk dipetakan sebagai bahan evaluasi bagi para pemangku kepentingan di dunia sepak bola nasional. Pendekatan tradisional seperti survei tidak lagi efisien untuk memahami opini publik dalam skala besar. Oleh karena itu, *Text Mining* menjadi sangat krusial sebagai proses ekstraksi informasi berharga dari data teks tidak terstruktur. Salah satu teknik utamanya adalah analisis sentimen, yang berfungsi mengidentifikasi emosi atau sikap penulis terhadap suatu topik. Meskipun penelitian mengenai analisis sentimen telah ada, studi yang berfokus pada figur spesifik seperti pelatih timnas di Indonesia masih minim. Algoritma Naive Bayes mampu memberikan performa yang memuaskan dalam mengklasifikasikan berita olahraga. Penelitian ini bertujuan menerapkan pendekatan *Text*

Mining dengan metode Naive Bayes untuk menganalisis dan memprediksi sentimen publik terhadap pelatih Timnas Indonesia, Patrick Kluivert [1]. Penelitian ini diharapkan dapat memberikan kontribusi akademik dan masukan strategis bagi para pemangku kepentingan di bidang olahraga nasional [2]. Metode naive bayes untuk mengklasifikasikan pelanggan dapat membantu pemilik memberikan bonus kepada pelanggan baru dan meningkatkan kualitas pelanggan [3]. Metode Naive Bayes Classifier menunjukkan presisi sebesar 95% dan recall sebesar 94% untuk kelas positif, yang menunjukkan bahwa algoritma ini efektif dalam mengklasifikasikan judul berita dengan sentimen positif. Secara keseluruhan, temuan ini membantu memahami analisis sentimen pada judul berita di situs media online, terutama di Kompas.TV di Kota Semarang [4].

2. METODE PENELITIAN

Penelitian ini menerapkan metode *Text Mining* yang melibatkan beberapa langkah sistematis untuk mencapai tujuan penelitian. Tahapan penelitian ini menggambarkan urutan logis untuk mendapatkan hasil sesuai harapan.

2.1 Pengumpulan Data

Tahap awal adalah pengumpulan data komentar netizen terhadap Patrick Kluivert dari platform X. Data dikumpulkan melalui teknik *scrapping* menggunakan Google Collab dengan kata kunci "Timnas Indonesia Patrick Kluivert", menghasilkan total 263 data setelah proses preprocessing. Data mentah tersebut kemudian disimpan dalam format .CSV.

2.2 Analisis Sentimen

Analisis sentimen adalah cabang dari *Natural Language Processing* (NLP) yang secara komputasional mengidentifikasi, mengekstrak, dan mengklasifikasikan opini subjektif dari data teks. Tujuan utamanya adalah untuk menentukan polaritas emosional apakah sentimen yang diekspresikan bersifat positif, negatif, atau netral. Dengan memproses informasi tekstual secara otomatis, teknik ini memungkinkan analisis data dalam skala besar, seperti komentar media sosial atau ulasan produk, untuk memahami persepsi dan pandangan publik secara efisien [5].

2.3 Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan mempersiapkan data agar lebih terstruktur. Dalam konteks pencarian informasi atau analisis data, proses ini menciptakan sebuah indeks yaitu representasi digital dari konten dari dokumen yang efisien dan akurat [6]. Oleh karena itu, *preprocessing* sering kali berbuah hasil yang signifikan dalam pengambilan keputusan dan pengembangan strategi basis data [7]

Langkah-langkah yang dilakukan yaitu:

- Case Folding* adalah langkah dalam *preprocessing* yang bertujuan untuk mengubah huruf kapital menjadi huruf kecil (*lowercase*), sehingga data menjadi lebih terstruktur dan lebih mudah untuk diproses lebih lanjut.
- Tokenization*, atau tokenisasi, berfungsi untuk menghapus tanda baca, tag, emotikon, dan memecah setiap kalimat menjadi kumpulan kata (kata tunggal). Proses tokenisasi ini disesuaikan dengan kebutuhan spesifik dari pemrosesan yang akan dilakukan.
- Normalization*, atau normalisasi, bertujuan untuk memperbaiki kesalahan penulisan kata, seperti typo atau penggunaan bentuk yang tidak baku, termasuk penghilangan perulangan yang digunakan untuk penekanan kata. Contohnya, kata "sangat" yang ditulis sebagai "sangaaat" akan diperbaiki.
- Filtering* adalah proses pemilihan kata-kata penting dari hasil tokenisasi, di mana kata-kata yang tidak relevan untuk identifikasi, seperti "yang", "kecuali", "tetapi", "dan", "dapat", dan sejenisnya, akan dihapus.
- Stemming* adalah proses yang digunakan untuk mengembalikan kata ke bentuk dasarnya (kata dasar), sehingga memudahkan analisis lebih lanjut terhadap data yang telah diproses.

2.4 TF-IDF

Pada tahap ini akan dilakukan pemberian bobot pada setiap kata yang ada pada data dengan menggunakan metode TF-IDF. TF-IDF merupakan suatu metode pembobotan kata yang dikenal baik dalam mengevaluasi pentingnya sebuah kata yang ada dalam dokumen [8]. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk pembobotan kata mengubah teks menjadi representasi numerik. TF mengukur frekuensi kemunculan kata dalam dokumen, sementara IDF mengukur tingkat kepentingan kata tersebut di seluruh dokumen. Selain itu, TF-IDF juga sering digunakan dalam pengembangan sistem rekomendasi dan analisis

sentimen, di mana pemahaman yang mendalam tentang kata-kata yang digunakan sangat penting untuk menghasilkan hasil yang akurat dan bermanfaat [9].

Perhitungan pada TF, IDF dan TF-IDF sebagai berikut:

a. TF (*Term Frequency*)

Frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai TF menunjukkan seberapa penting sebuah kata dalam dokumen tertentu. Semakin sering suatu kata muncul, semakin tinggi nilai TF-nya dalam dokumen tersebut.

$$TF(t, d) = \frac{\text{Jumlah kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d} \quad (2.1)$$

b. IDF (*Inverse Document Frequency*)

Ukuran pentingnya suatu kata terhadap seluruh kumpulan dokumen. IDF menurunkan bobot kata-kata yang terlalu sering muncul di banyak dokumen (seperti kata “dan”, “itu”, dll), karena kata seperti itu tidak memiliki nilai informatif yang tinggi.

$$IDF(t) = \log_{10}\left(\frac{N}{df(t)}\right) \quad (2.2)$$

Dimana:

N : jumlah total dokumen

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

c. TF-IDF (*Term Frequency-Inverse Document Frequency*)

TF-IDF merupakan gabungan dari TF dan IDF yang digunakan untuk menilai seberapa penting suatu kata dalam sebuah dokumen relatif terhadap seluruh dokumen.

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (2.3)$$

2.5 Klasifikasi Naïve Bayes

Naïve Bayes adalah sebuah metode yang berbasis pada probabilitas untuk menentukan kategori atau kelompok dari suatu data. Algoritma klasifikasi ini dirancang untuk mampu mengolah data dalam jumlah besar dengan tingkat akurasi yang tinggi. Metode ini bekerja dengan menghitung probabilitas kemunculan setiap fitur pada setiap kelas yang ada di dalam dataset [10]. Dengan kemampuannya untuk menangani data yang beragam dan besar, *Naïve Bayes Classifier* terus menjadi salah satu metode yang paling banyak digunakan dalam penelitian dan aplikasi praktis di bidang pemrosesan bahasa alami dan analisis data [11].

2.6 Pengembangan Dan Pengujian Model

Model klasifikasi sentimen dibangun menggunakan algoritma *Naïve Bayes Classifier*. Dataset dibagi menjadi 3 (tiga) jenis yaitu, 70% data latih 30% data uji, 80% data latih 20% data uji dan 90% data latih 10% data uji. Model dilatih menggunakan data latih dan kinerjanya dievaluasi menggunakan data uji dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung dari *Confusion Matrix*. Selain itu, *confusion matrix* juga memungkinkan para peneliti dan praktis untuk melakukan analisis lebih lanjut terhadap jenis kesalahan yang terjadi, sehingga dapat diambil langkah-langkah perbaikan yang tepat [12].

Berikut adalah cara menghitung akurasi berdasarkan tabel *confusion matrix*:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} * 100\% \quad (2.4)$$

Presisi merupakan rasio antara elemen penting yang dipilih terhadap semua elemen yang dipilih. Sementara itu, akurasi bisa diartikan sebagai tingkat kecocokan sebuah tanggapan dengan permintaan informasinya. Berikut adalah rumus presisi:

$$\text{Precision} = \frac{TP}{TP+FP} * 100\% \quad (2.5)$$

Recall adalah perbandingan antara jumlah item signifikan yang berhasil diambil dan jumlah total item signifikan yang tersedia. Berikut adalah rumus *recall*:

$$\text{Recall} = \frac{TP}{TP+FN} * 100\% \quad (2.6)$$

Nilai *precision* (presisi) dan *recall* dapat dinyatakan dalam dua bentuk, yaitu dalam bentuk persentase (0-100%) atau sebagai bilangan desimal dalam rentang 0 hingga 1. Kinerja sebuah sistem dinilai optimal apabila berhasil mencapai nilai yang tinggi untuk kedua skor tersebut.

$$\text{F1 Score} = \frac{2*(\text{recall}*\text{precision})}{(\text{recall}+\text{precision})} * 100\% \quad (2.7)$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi analisis, hasil implementasi ataupun pengujian serta pembahasan dari topik penelitian, yang bisa dibuat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya.

3.1 Crawling

Tahap pertama adalah pengumpulan data. Data yang dikumpulkan bersumber dari komentar netizen di platform X. Proses pengumpulan dataset berdasarkan *kata kunci* seperti “Timnas Indonesia”, “Timnas” dan “Patrick Kluivert” dari konten yang ditemukan di platform X. *Crawling Data* menjadi langkah awal untuk mengambil dataset dengan Bahasa pemrograman python terkoneksi *Auth Token X (Twitter)* yang kemudian dataset disimpan kedalam file CSV lalu import file ke dalam sistem.

Tabel 1. Crawling Data

<i>Tweets</i>	Label
peluang Timnas Indonesia PSSI berharap Patrick Kluivert bisa datang lebih cepat untuk merencanakan TC Timnas Indonesia jelang melawan China dan Jepang https://t.co/2sLqSEkOxC	POSITIF
Dengan segala kemungkinan peluang Timnas Indonesia PSSI berharap Patrick Kluivert bisa datang lebih cepat untuk merencanakan TC Timnas Indonesia jelang melawan China dan Jepang https://t.co/2sLqSEkOxC	NETRAL
2 Pemain Inti Timnas Indonesia Absen Lawan China Skuat Patrick Kluivert Bakal Pincang di Stadion GBK https://t.co/Fus8VL4SF0 lewat @tribunnews	NEGATIF

3.2 Preprocessing

Preprocessing diterapkan untuk membersihkan dataset untuk memastikan bahwa proses klasifikasi sentimen menghasilkan tingkat akurasi yang optimal. Berikut tahapan-tahapan dari *preprocessing*:

a. Case Folding

Mengkonversi teks menjadi huruf kecil (*lowercase*) secara konsisten. Tujuan dari proses ini adalah agar data dalam analisis teks dianggap seragam. Hasil *case folding* dapat dilihat pada tabel dibawah:

Tabel 2. Case Folding

Sebelum	Sesudah
Sekjen PSSI Yunus Nusi menyampaikan harapannya agar pelatih Timnas Indonesia Patrick Kluivert segera datang ke Tanah Air #patrickkluivert #yunusnusi wujud bakti nyata Baila Selasa Bank Dunia https://t.co/OquaKr5Tkr	sekjen pssi yunus nusi menyampaikan harapannya agar pelatih timnas indonesia patrick kluivert segera datang ke tanah air #patrickkluivert #yunusnusi wujud bakti nyata baila selasa bank dunia https://t.co/oquakr5tkr
Pelatih Timnas Indonesia Patrick Kluivert memantau Ole Romeny dan Marselino Ferdinan dengan hadir langsung di markas Oxford United bersama asisten pelatih	pelatih timnas indonesia patrick kluivert memantau ole romeny dan marselino ferdinan dengan hadir langsung di markas oxford united bersama asisten pelatih denny landzaat dan penasihat

Denny Landzaat dan Penasihat Teknis PSSI Jordi Cruyff. https://t.co/XXYQUxwjLi Sejak Ngikuti @FabrizioRomano ini kali kedua terkait dengan sepakbola Indonesia. Pertama: Spil Pengumuman Patrick Kluivert jadi Pelatih Timnas Indonesia Kedua: Premier League di Vidio extend 2025 - 2028 Here We Go	teknis pssi jordi cruyff. https://t.co/xyyquxwjli sejak ngikuti @fabrizioromano ini kali kedua terkait dengan sepakbola indonesia. pertama: spil pengumuman patrick kluivert jadi pelatih timnas indonesia kedua: premier league di vidio extend 2025 - 2028 here we go
---	--

b. *Cleaning*

Cleaning dataset adalah proses menghapus *noise* seperti simbol, emoji, *mention*, spasi berlebih, *hashtag*, dan tautan yang tidak relevan agar tidak mengganggu pemrosesan data. Hasil *cleansing* dapat dilihat pada tabel dibawah:

Tabel 3. *Cleansing*

Sebelum	Sesudah
Sekjen PSSI Yunus Nusi menyampaikan harapannya agar pelatih Timnas Indonesia Patrick Kluivert segera datang ke Tanah Air #patrickkluivert #yunusnusi wujud bakti nyata Baila Selasa Bank Dunia https://t.co/OquaKr5TKr Pelatih Timnas Indonesia Patrick Kluivert memantau Ole Romeny dan Marselino Ferdinan dengan hadir langsung di markas Oxford United bersama asisten pelatih Denny Landzaat dan Penasihat Teknis PSSI Jordi Cruyff. https://t.co/XXYQUxwjLi Sejak Ngikuti @FabrizioRomano ini kali kedua terkait dengan sepakbola Indonesia. Pertama: Spil Pengumuman Patrick Kluivert jadi Pelatih Timnas Indonesia Kedua: Premier League di Vidio extend 2025 - 2028 Here We Go	sekjen pssi yunus nusi menyampaikan harapannya agar pelatih timnas indonesia patrick kluivert segera datang ke tanah air wujud bakti nyata baila selasa bank dunia pelatih timnas indonesia patrick kluivert memantau ole romeny dan marselino ferdinan dengan hadir langsung di markas oxford united bersama asisten pelatih denny landzaat dan penasihat teknis pssi jordi cruyff sejak ngikuti ini kali kedua terkait dengan sepakbola indonesia pertama spil pengumuman patrick kluivert jadi pelatih timnas indonesia kedua premier league di vidio extend here we go

c. *Tokenizing*

Penerapan *Tokenizing* untuk memproses teks dengan membaginya menjadi beberapa unit yang lebih kecil dan terpisah. Bagian ini yang disebut ‘token’ merupakan kata-kata tunggal, namun mencakup potongan-potongan lain yang dianggap sebagai unit penting. Hasil *Tokenizing* dapat dilihat pada tabel dibawah:

Tabel 4. *Tokenizing*

Sebelum	Sesudah
Sekjen PSSI Yunus Nusi menyampaikan harapannya agar pelatih Timnas Indonesia Patrick Kluivert segera datang ke Tanah Air #patrickkluivert #yunusnusi wujud bakti nyata Baila Selasa Bank Dunia https://t.co/OquaKr5TKr Pelatih Timnas Indonesia Patrick Kluivert memantau Ole Romeny dan Marselino Ferdinan dengan hadir langsung di markas Oxford United bersama asisten pelatih Denny Landzaat dan Penasihat Teknis PSSI Jordi Cruyff. https://t.co/XXYQUxwjLi Sejak Ngikuti @FabrizioRomano ini kali kedua terkait dengan sepakbola Indonesia. Pertama: Spil Pengumuman Patrick Kluivert jadi Pelatih Timnas Indonesia Kedua: Premier League di Vidio extend 2025 - 2028 Here We Go	sekjen, pssi, yunus, nusi, menyampaikan, harapannya, agar, pelatih, timnas, indonesia, patrick, kluivert, segera, datang, ke, tanah, air, wujud, bakti, nyata, baila, selasa, bank, dunia pelatih, timnas, indonesia, patrick, kluivert, memantau, ole, romeny, dan, marselino, ferdinan, dengan, hadir, langsung, di, markas, oxford, united, bersama, asisten, pelatih, denny, landzaat, dan, penasihat, teknis, pssi, jordi, Cruyff sejak, ngikuti, ini, kali, kedua, terkait, dengan, sepakbola, indonesia, pertama, spil, pengumuman, patrick, kluivert, jadi, pelatih, timnas, indonesia, kedua, premier, league, di, vidio, extend, here, we, go

d. *Stopword*

Proses stopword removal merupakan tahap preprocessing yang bertujuan mengeliminasi kata-kata bantu (*stopword*) - istilah-istilah umum yang tidak memiliki makna signifikan namun frekuensinya tinggi. Contoh stop words antara lain "ini", "tetap", dan "nya". Output dari proses ini dapat diamati pada tabel berikut:

Tabel 5. Stopword

Sebelum	Sesudah
Sekjen PSSI Yunus Nusi menyampaikan harapannya agar pelatih Timnas Indonesia Patrick Kluivert segera datang ke Tanah Air #patrickkluivert #yunusnusi wujud bakti nyata Baila Selasa Bank Dunia https://t.co/OquaKr5Tkr Pelatih Timnas Indonesia Patrick Kluivert memantau Ole Romeny dan Marselino Ferdinan dengan hadir langsung di markas Oxford United bersama asisten pelatih Denny Landzaat dan Penasihat Teknis PSSI Jordi Cruyff. https://t.co/XXYQUxwjLi Sejak Ngikuti @FabrizioRomano ini kali kedua terkait dengan sepakbola Indonesia. Pertama: Spil Pengumuman Patrick Kluivert jadi Pelatih Timnas Indonesia Kedua: Premier League di Vidio extend 2025 - 2028 Here We Go	sekjen, pssi, yunus, nusi, menyampaikan, harapannya, pelatih, timnas, indonesia, patrick, kluivert, segera, datang, tanah, air, wujud, bakti, nyata, baila, selasa, bank, dunia pelatih, timnas, indonesia, patrick, kluivert, memantau, ole, romeny, marselino, ferdinan, hadir, langsung, markas, oxford, united, bersama, asisten, pelatih, denny, landzaat, penasihat, teknis, pssi, jordi, Cruyff sejak, ngikuti, kali, kedua, terkait, sepakbola, indonesia, pertama, spil, pengumuman, patrick, kluivert, jadi, pelatih, timnas, indonesia, kedua, premier, league, vidio, extend, here, we, go

e. *Stemming*

Pada proses ini menjadikan keseluruhan kata yang ada menjadi kata dasar saja. Hasil *stemming* dapat dilihat pada tabel dibawah:

Tabel 6. Stemming

Sebelum	Sesudah
Sekjen PSSI Yunus Nusi menyampaikan harapannya agar pelatih Timnas Indonesia Patrick Kluivert segera datang ke Tanah Air #patrickkluivert #yunusnusi wujud bakti nyata Baila Selasa Bank Dunia https://t.co/OquaKr5Tkr Pelatih Timnas Indonesia Patrick Kluivert memantau Ole Romeny dan Marselino Ferdinan dengan hadir langsung di markas Oxford United bersama asisten pelatih Denny Landzaat dan Penasihat Teknis PSSI Jordi Cruyff. https://t.co/XXYQUxwjLi Sejak Ngikuti @FabrizioRomano ini kali kedua terkait dengan sepakbola Indonesia. Pertama: Spil Pengumuman Patrick Kluivert jadi Pelatih Timnas Indonesia Kedua: Premier League di Vidio extend 2025 - 2028 Here We Go	sekjen, pssi, yunus, nus, sampai, harap, latih, timnas, indonesia, patrick, kluivert, segera, datang, tanah, air, wujud, bakti, nyata, baila, selasa, bank, dunia latih, timnas, indonesia, patrick, kluivert, pantau, ole, romeny, marselino, ferdinan, hadir, langsung, markas, oxford, united, sama, asisten, latih, denny, landzaat, nasihat, teknis, pssi, jordi, Cruyff sejak, ngikuti, kali, dua, kait, sepakbola, indonesia, pertama, spil, umum, patrick, kluivert, jadi, latih, timnas, indonesia, dua, premier, league, vidio, extend, here, we, go

3.3 TF-IDF

Langkah berikutnya adalah proses ekstraksi fitur menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik. TF-IDF mengukur pentingnya sebuah kata dalam sebuah dokumen relatif.

3.4 Hasil dan Pengujian

3.4.1 Confussion Matrix

Hasil dari perhitungan Naive Bayes sebelumnya kemudian diuji dengan menggunakan *Confussion matrix*. *Confussion matrix* digunakan untuk membandingkan hasil prediksi model dengan label sebenarnya dari data uji. Dari *Confussion matrix*, dapat dihitung berbagai metrik evaluasi seperti akurasi, presisi, Recall, dan F1-Score. Tabel 10 berikut adalah hasil dari pengujian *Confussion matrix*.

Tabel 7. *Confussion Matrix*

Aktual/Prediksi	Negatif	Netral	Positif
Negatif	2	4	2
Netral	1	14	7
Positif	0	5	18

Evaluasi model dilakukan melalui analisis laporan klasifikasi yang memuat berbagai parameter kinerja, termasuk *Precision*, *Recall*, *F1-Score* dan *Support* untuk masing-masing kategori. Data lengkap dari analisis klasifikasi tersebut tersaji dalam tabel berikut.

Tabel 8. Laporan Klasifikasi

Ukuran	Precision	Recall	F1-Score	Support
Negatif	0.67	0.25	0.36	8
Netral	0.61	0.64	0.62	22
Positif	0.67	0.78	0.72	23
Weighted Avg	0.64	0.64	0.63	53

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan dan mengevaluasi model klasifikasi sentimen menggunakan algoritma *Naive Bayes* untuk menganalisis opini netizen di platform X terhadap pelatih Timnas Indonesia, Patrick Kluivert. Dengan menggunakan 263 data komentar yang telah melalui tahapan *preprocessing* dan pembobotan TF-IDF, model yang dikembangkan mencapai tingkat akurasi keseluruhan sebesar 64,15%. Analisis kinerja yang lebih mendalam melalui *Confussion Matrix* mengungkap bahwa model menunjukkan performa yang bervariasi antar kelas sentimen. Model ini terbukti sangat efektif dalam mengidentifikasi sentimen positif dengan nilai *Recall* mencapai 0.78, namun performanya menurun secara signifikan saat mengklasifikasikan sentimen netral (*Recall* 0.64) dan sangat terbatas dalam mengenali sentimen negatif (*Recall* hanya 0.25). Temuan utama dari penelitian ini adalah adanya pengaruh kuat dari ketidakseimbangan data terhadap kinerja model. Dominasi data sentimen positif dalam dataset pelatihan menyebabkan model lebih cenderung memprediksi kelas mayoritas dan kurang terlatih untuk mengenali pola pada kelas minoritas (negatif dan netral). Keterbatasan ini menggarisbawahi tantangan umum dalam analisis sentimen pada data media sosial yang seringkali tidak rata. Untuk pengembangan di masa depan, beberapa perbaikan sangat direkomendasikan. Pertama, perluasan dataset menjadi prioritas utama, baik dari segi jumlah maupun keseimbangan antar kelas sentimen, untuk meningkatkan akurasi dan kemampuan model. Kedua, disarankan untuk melakukan studi komparatif dengan algoritma yang lebih kompleks seperti Support Vector Machine (SVM) atau model *deep learning* (misalnya LSTM), yang berpotensi lebih baik dalam menangani konteks dan bahasa yang sering ditemukan dalam komentar media sosial. Terakhir, untuk meningkatkan nilai praktis, penelitian dapat dikembangkan dengan menambahkan fitur visualisasi tren sentimen dari waktu ke waktu, yang akan memberikan wawasan yang lebih dinamis dan informatif bagi para pemangku kepentingan di industri olahraga nasional.

DAFTAR PUSTAKA

- [1]. F. R. Mohammad, “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine.,” *Jurnal Teknologi Dan Komunikasi Pemerintah*, Vol. 13, P. 13, 2022.
- [2]. W. P. Ari And D. S. Arief, “ Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Konten Berita Olahraga,” *Ejournal.Stmik*, Vol. 6, P. 6, 2020.
- [3]. H.F.Putro, R. T. Vlandari And W. L. Y. Saptomo, “Penerapan Metode Naïve Bayes Untuk Klasifikasi Pelanggan”, *Jurnal Tikomsin*, Vol. 8, No. 2, Oktober 2020, ISSN Online : 2620-7532, DOI: <https://doi.org/10.30646/Tikomsin.V8i2.500>.
- [4]. A. R. Wibowo, Et All, “Penerapan Metode Naïve Bayes Classifier untuk Klasifikasi Sentimen Pada Judul Berita”, *Logiclink : Journal Of Artificial Intelligence And Multimedia In Informatics* vol. 1 No. 1, June 2024, Pp. 1 –12
- [5]. S. Alfandi And H. N. Firman, “ANALISIS SENTIMEN MASYARAKAT TERHADAP AYAT LATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER.,” *Zonasi, Jurnal Sistem Informasi*, Vol. 70, P. 70, 2023.
- [6]. R. T. Muhammad, N. Dede And N. Hani, “Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier Dengan Fitur TF-IDF.,” *E-Proceeding Of Engineering*, Vol. 15, P. 15, 2023.
- [7]. R. Bagus And S. R. Ryan, “Komparasi Algoritma Naïve Bayes Dan Logistic Regression Untuk Analisis Sentimen Metaverse.,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, Vol. 12, P. 12, 2024.
- [8]. A. Ramadhan, “The Impact Of Features Extraction On The Sentiment Analysis. International Conference On Pervasive Computing Advances And Applications.,” *Procedia Computer Science*, Vol. 12, P. 12, 2019.
- [9]. N. Musfiroh, H. And A. F. Inda, “Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining.,” *Jurnal Informatika Mulawarman*, Vol. 4, P. 4, 2013.
- [10]. D. Dedi, S. Neri And A. Zaenal, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen.,” *Jurnal TEKNO Kompak*, Vol. 15, P. 15, 2021.
- [11]. P. D. Syahril, H. S. Shofa And N. Fitri, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes Dan KNN.,” *Jurnal Komtekinfo*, Vol. 7, P. 7, 2023.
- [12]. N. G. Zelin And N. , “Analisis Sentimen Opini Pengguna Twitter Pada Aplikasi Bibit Menggunakan Multinomial Naive Bayes.,” *JTS : Jurnal Teknik Dan Science*, Vol. 8, P. 8, 2023.