



Analisis Sentimen Publik terhadap Pemecatan Shin Tae-yong sebagai Pelatih Timnas Indonesia

¹Hibban Sani Muttaqin*, ²Rafi Sugara, ³Rafli Candra Nugraha

^{1,2,3} Universitas Siliwangi, Indonesia

Email : ¹227006073@student.unsil.ac.id*, ²227006063@student.unsil.ac.id,
³227006075@student.unsil.ac.id

Abstract

This research analyzes public sentiment regarding the dismissal of Shin Tae-yong (STY) as the Indonesian National Team coach through social media X using a text classification approach with the Multinomial Naive Bayes algorithm. A total of 3,377 Indonesian tweets underwent text preprocessing stages and were classified into three sentiment categories: positive, negative, and neutral. The results show that the majority of sentiments were negative (~64%), followed by positive (~24%) and neutral (~12%) sentiments. The classification model with TF-IDF representation (ngram 1-2, maximum 5,000 features) achieved 84.37% accuracy on training data and 82.54% on testing data, demonstrating stable performance with minimal overfitting. Performance analysis per class yielded an F1-score of 0.89 for negative sentiment, but lower scores for neutral (0.65) and positive (0.68) sentiments, indicating challenges in differentiating these two categories due to data distribution imbalance. WordCloud analysis confirmed distinct lexical patterns across sentiment categories, with words like "dismiss" and "not" dominating negative sentiments, while "support" and "proud" appeared in positive sentiments. This research proves the effectiveness of the Naive Bayes algorithm in analyzing public sentiment on sports issues in Indonesia, despite limitations in classifying neutral and positive sentiments that need to be addressed in future research.

Keywords : *sentiment analysis, naive bayes, TF-IDF.*

Abstrak

Penelitian ini menganalisis sentimen publik terhadap pemecatan Shin Tae-yong (STY) sebagai pelatih Tim Nasional Indonesia melalui media sosial X dengan pendekatan klasifikasi teks menggunakan algoritma Multinomial Naive Bayes. Sebanyak 3.377 tweet berbahasa Indonesia diproses melalui tahapan pra-pemrosesan teks dan diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa mayoritas sentimen bersifat negatif (~64%), diikuti oleh sentimen positif (~24%) dan netral (~12%). Model klasifikasi dengan representasi TF-IDF (ngram 1-2, maksimal 5.000 fitur) mencapai akurasi 84,37% pada data pelatihan dan 82,54% pada data pengujian, menunjukkan performa yang stabil dengan overfitting minimal. Analisis performa per kelas menghasilkan F1-score 0,89 untuk sentimen negatif, namun lebih rendah untuk sentimen netral (0,65) dan positif (0,68), yang mengindikasikan adanya tantangan dalam membedakan kedua kategori tersebut akibat ketidakseimbangan distribusi data. Analisis WordCloud mengonfirmasi pola leksikal yang berbeda antar kategori sentimen, dengan kata-kata seperti "pecat" dan "tidak" mendominasi sentimen negatif, sementara "dukungan" dan "bangga" muncul pada sentimen positif. Penelitian ini membuktikan efektivitas algoritma Naive Bayes dalam menganalisis sentimen

publik terhadap isu olahraga di Indonesia, meskipun terdapat limitasi dalam klasifikasi sentimen netral dan positif yang perlu diatasi dalam penelitian selanjutnya.

Kata kunci : analisis sentimen, naive bayes, TF-IDF .

Corresponding Author;
E-mail: 227006073@student.unsil.ac.id



Pendahuluan

Sepak bola merupakan salah satu olahraga paling populer di Indonesia dengan basis pendukung yang sangat besar dan loyal. Kinerja Tim Nasional Indonesia (Timnas) selalu menjadi sorotan publik, termasuk setiap keputusan strategis seperti pergantian pelatih. Perubahan tersebut seringkali menimbulkan berbagai respons publik, mulai dari dukungan hingga kecaman, yang banyak diekspresikan melalui media sosial. Sentimen publik terhadap keputusan seperti pemecatan pelatih menjadi penting untuk dianalisis karena dapat memengaruhi kepercayaan masyarakat terhadap federasi serta atmosfer dukungan terhadap tim (Nufus & Surapati, 2024).

Media sosial kini telah menjadi wadah utama bagi masyarakat untuk menyampaikan opini dan reaksi terhadap isu-isu terkini, termasuk dalam ranah olahraga. Di Indonesia, isu pemecatan pelatih timnas seperti Shin Tae-yong (STY) telah memicu perdebatan luas di platform seperti X (sebelumnya Twitter), Instagram, dan YouTube. Sentimen yang muncul tidak hanya mencerminkan preferensi publik, tetapi juga membentuk narasi kolektif tentang kepemimpinan dan performa timnas (Nufus & Surapati, 2024).

Analisis sentimen menjadi salah satu metode yang efektif untuk mengukur kecenderungan opini publik melalui data teks dari media sosial. Teknik ini memungkinkan pengelompokan sentimen menjadi kategori positif, negatif, atau netral dan memberikan gambaran umum tentang persepsi publik (Gunawan, Purnama, & Rohani, 2024). Dalam konteks olahraga, analisis semacam ini telah digunakan untuk memahami persepsi terhadap tim nasional maupun peristiwa olahraga tertentu (Nufus & Surapati, 2024).

Meskipun analisis sentimen telah banyak diterapkan dalam berbagai konteks sosial dan politik, urgensi penelitian ini muncul dari tingginya sensitivitas publik terhadap keputusan strategis dalam sepak bola nasional, khususnya terkait pemecatan pelatih Tim Nasional Indonesia. Sepak bola di Indonesia tidak hanya dipandang sebagai olahraga, tetapi juga sebagai simbol identitas nasional dan kebanggaan kolektif, sehingga setiap kebijakan federasi berpotensi memicu dinamika opini publik yang luas dan intens. Reaksi masyarakat yang diekspresikan melalui media sosial dapat berdampak pada legitimasi keputusan federasi, stabilitas dukungan terhadap tim, serta

kepercayaan jangka panjang terhadap pengelolaan sepak bola nasional. Oleh karena itu, analisis sentimen publik terhadap isu pemecatan pelatih menjadi penting untuk memberikan gambaran objektif mengenai kecenderungan opini masyarakat. Penelitian ini mendesak dilakukan agar pemangku kepentingan memperoleh dasar empiris dalam memahami respons publik, bukan sekadar bersandar pada asumsi atau persepsi subjektif.

Metode Naive Bayes merupakan salah satu algoritma klasifikasi yang banyak digunakan dalam analisis teks karena kesederhanaan dan efisiensinya. Algoritma ini mengasumsikan independensi antar fitur dan menghitung probabilitas suatu teks masuk ke dalam kategori tertentu berdasarkan distribusi data sebelumnya. Beberapa penelitian telah menunjukkan bahwa Naive Bayes efektif dalam mengklasifikasikan sentimen dalam konteks media sosial, termasuk pada isu-isu politik dan olahraga (Prasetyo & Fitriani, 2023); (Nufus & Surapati, 2024).

Penelitian terbaru menunjukkan adanya peningkatan penggunaan algoritma Naive Bayes untuk analisis sentimen dalam konteks olahraga di media sosial. Nufus dan Surapati (2024) menerapkan metode ini untuk menganalisis sentimen publik terhadap Timnas Indonesia U-23 dalam ajang AFC-23 Asian Cup. Hasil analisis menunjukkan bahwa mayoritas sentimen bersifat positif, mencerminkan dukungan kuat masyarakat terhadap performa tim tersebut. Dalam konteks yang berbeda, Prasetyo dan Fitriani (2023) menggunakan Naive Bayes untuk mengklasifikasikan sentimen terhadap tokoh publik dalam ranah pemilu, dan menemukan bahwa algoritma ini cukup efektif dalam menghasilkan akurasi klasifikasi yang memadai.

Sementara itu, Gunawan dan rekan-rekannya (2024) memanfaatkan pendekatan serupa untuk menganalisis sentimen publik terhadap kandidat presiden Indonesia 2024, dan berhasil menunjukkan bahwa model ini mampu membedakan opini positif, negatif, dan netral secara akurat. Dukungan terhadap efektivitas metode ini juga dikemukakan oleh Hakiki, Pambudi, dan Asriyanik (2024) yang menegaskan bahwa Naive Bayes sangat andal untuk klasifikasi opini publik di platform X, terutama dalam isu-isu politik seperti pemilu dan tokoh publik. Meskipun hingga saat ini belum banyak penelitian yang secara spesifik membahas kasus pemecatan Shin Tae-yong, berbagai studi tersebut menunjukkan bahwa pendekatan analisis sentimen dengan metode Naive Bayes sangat relevan dan efektif dalam menilai persepsi publik terhadap peristiwa besar di dunia olahraga.

Tinjauan terhadap penelitian terdahulu menunjukkan bahwa sebagian besar studi analisis sentimen dengan metode Naive Bayes berfokus pada isu politik, pemilu, tokoh publik, atau performa tim dalam turnamen tertentu. Penelitian yang secara spesifik mengkaji sentimen publik terhadap keputusan manajerial dalam sepak bola nasional, terutama terkait pemecatan pelatih Timnas Indonesia seperti Shin Tae-yong, masih sangat terbatas. Selain itu, studi-studi sebelumnya cenderung menitikberatkan pada

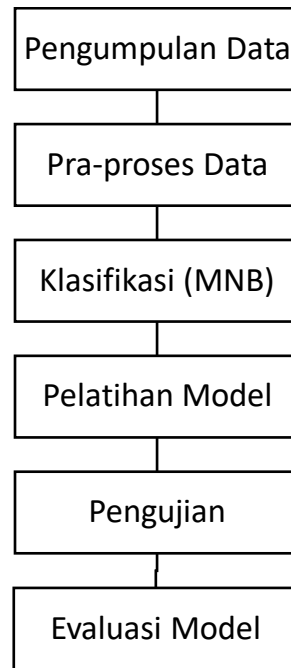
hasil klasifikasi sentimen tanpa mengaitkannya secara kontekstual dengan implikasi sosial dan institusional bagi federasi sepak bola. Kesenjangan ini menunjukkan perlunya penelitian yang secara khusus menganalisis persepsi publik terhadap keputusan strategis federasi olahraga dengan pendekatan komputasional yang sistematis. Dengan demikian, penelitian ini hadir untuk mengisi kekosongan tersebut dengan memadukan analisis sentimen berbasis Naive Bayes dan konteks manajemen olahraga nasional.

Penelitian ini bertujuan untuk menganalisis sentimen publik di media sosial terhadap isu pemecatan pelatih Tim Nasional Indonesia menggunakan metode Naive Bayes. Secara khusus, penelitian ini bertujuan untuk mengklasifikasikan opini publik ke dalam kategori positif, negatif, dan netral, serta mengidentifikasi kecenderungan dominan dari respons masyarakat. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi efektivitas metode Naive Bayes dalam mengolah data teks media sosial pada konteks isu olahraga nasional. Hasil penelitian diharapkan dapat memberikan gambaran empiris mengenai pola persepsi publik terhadap kebijakan strategis federasi sepak bola.

Secara teoretis, penelitian ini diharapkan dapat memperkaya khazanah kajian analisis sentimen dalam bidang olahraga, khususnya pada isu manajemen dan kebijakan sepak bola nasional. Secara metodologis, penelitian ini memberikan kontribusi dalam penerapan metode Naive Bayes untuk konteks isu olahraga yang masih relatif jarang diteliti dibandingkan isu politik. Secara praktis, hasil penelitian ini dapat menjadi bahan pertimbangan bagi federasi sepak bola dan pemangku kepentingan terkait dalam memahami respons publik terhadap keputusan strategis yang diambil. Selain itu, penelitian ini juga dapat menjadi rujukan bagi penelitian selanjutnya yang ingin mengembangkan analisis sentimen dengan pendekatan komputasional pada isu-isu olahraga dan kebijakan publik.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi teks berbasis pembelajaran mesin untuk menganalisis sentimen publik terhadap isu pemecatan Shin Tae-yong sebagai pelatih Tim Nasional Indonesia. Data yang digunakan diperoleh dari dataset publik di platform Kaggle yang terdiri atas 3.377 tweet berbahasa Indonesia. Dataset tersebut belum memiliki label sentimen, sehingga dilakukan pelabelan manual oleh peneliti untuk mengklasifikasikan setiap tweet ke dalam salah satu dari tiga kategori sentimen: positif, negatif, atau netral.



Gambar 1. Diagram Alur Penelitian

Sebelum dilakukan klasifikasi, data melalui proses pra-pemrosesan untuk meningkatkan kualitas dan konsistensinya. Tahapan ini meliputi normalisasi teks ke dalam huruf kecil (case folding), pembersihan karakter-karakter tidak relevan seperti URL, mention, dan simbol khusus, serta proses tokenisasi untuk memecah teks menjadi unit kata. Selanjutnya, dilakukan penghapusan stopword dan stemming menggunakan algoritma Nazief & Adriani melalui pustaka Sastrawi, yang telah terbukti efektif dalam pengolahan teks bahasa Indonesia (Purbolaksono, Reskyadita, & Suryani, 2020). Kata-kata tidak baku atau slang juga dinormalisasi menjadi bentuk baku agar makna teks lebih mudah diinterpretasikan oleh model klasifikasi.

Proses klasifikasi dilakukan menggunakan algoritma Multinomial Naive Bayes (MNB) yang dikenal efektif dalam tugas-tugas klasifikasi berbasis teks karena mempertimbangkan frekuensi kemunculan kata dalam dokumen. MNB dipilih karena memiliki kinerja yang baik dan efisien pada data teks berbahasa Indonesia, sebagaimana telah dibuktikan dalam sejumlah studi sebelumnya seperti dalam klasifikasi ujaran kebencian dan konten clickbait (Mahendra, Rahmat, & Mumpuni, 2024).

Secara matematis, probabilitas sebuah dokumen d termasuk dalam kelas c dihitung dengan rumus:

$$P(c | d) = \frac{P(c) \prod_{i=1}^n P(w_i | c)^{f_i}}{P(d)}$$

Keterangan:

- $P(c | d)$ = probabilitas tweet dikategorikan ke kelas sentimen c
- $P(c)$ = probabilitas awal dari kelas c (prior probability)
- $P(w_i | c)$ = probabilitas kemunculan kata w_i dalam kelas c

- f_i = frekuensi kemunculan kata w_i dalam tweet d
- $P(d)$ = probabilitas dokumen (normalizing constant, diabaikan dalam perbandingan antar kelas)

Untuk pelatihan dan pengujian model, dataset dibagi menjadi dua bagian menggunakan skema stratified split, yakni 80% data sebagai data latih dan 20% sisanya sebagai data uji. Pembagian ini menjaga distribusi kelas sentimen agar tetap seimbang di kedua subset data.

Evaluasi kinerja model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Metrik F1-score sangat penting terutama jika terdapat ketidakseimbangan distribusi antar kelas. Rumus-rumus yang digunakan adalah sebagai berikut:

- Akurasi

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

- F1-score

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Keterangan:

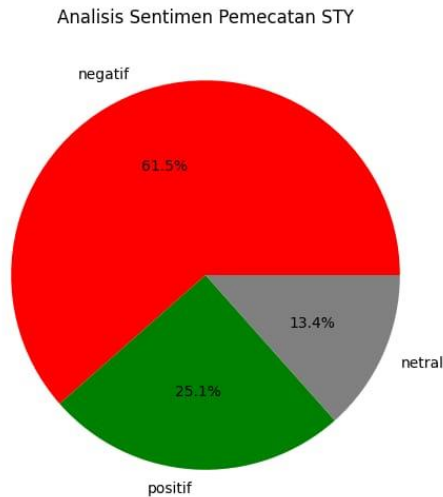
- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Selain itu, digunakan confusion matrix untuk menggambarkan secara visual distribusi prediksi benar dan salah dari masing-masing kelas. Evaluasi ini memberikan gambaran menyeluruh tentang efektivitas model dalam mengklasifikasikan sentimen publik secara tepat.

Seluruh proses penelitian diimplementasikan menggunakan bahasa pemrograman Python pada platform Google Colab, dengan dukungan berbagai pustaka yang umum digunakan dalam analisis data, pemrosesan bahasa alami, dan pembelajaran mesin.

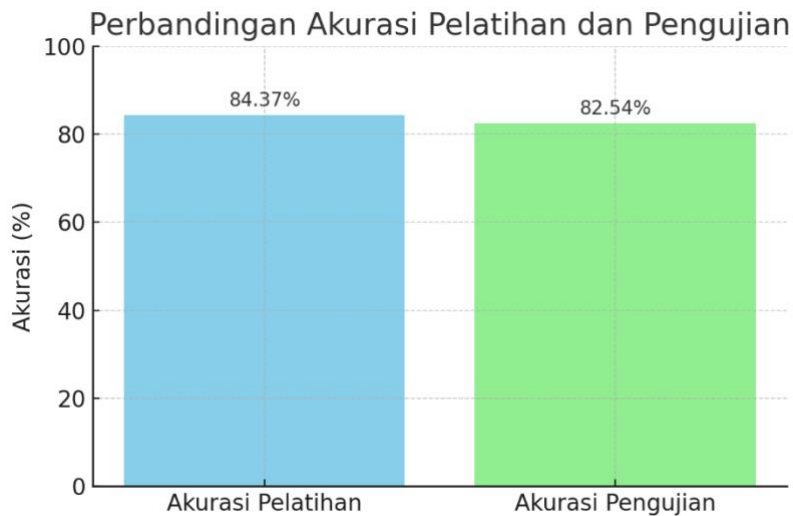
Hasil dan Pembahasan

Berdasarkan eksperimen pada data media sosial X tentang pemecatan Shin Tae-yong (STY) sebagai pelatih tim nasional Indonesia, diperoleh beberapa temuan utama. Pertama, distribusi sentimen menunjukkan dominasi sentimen negatif yang mencapai sekitar 64% dari total sampel, diikuti sentimen positif (~24%) dan netral (~12%) (Lihat Gambar 2). Hal ini mengindikasikan bahwa sebagian besar publik bereaksi secara kritis terhadap keputusan pemecatan tersebut.



Gambar 2. Diagram lingkaran “Distribusi Sentimen Publik terhadap Pemecatan STY”

Kedua, akurasi model Multinomial Naive Bayes dengan representasi TF-IDF (ngram 1-2, maksimal 5 000 fitur) mencapai 84,37% pada data pelatihan dan 82,54% pada data pengujian (Lihat Gambar 3), yang menunjukkan bahwa model tidak mengalami overfitting signifikan dan dapat digeneralisasi dengan cukup baik ($\Delta \approx 1,83\%$).



Gambar 3. Perbandingan akurasi pelatihan dan pengujian

Dalam implementasi penelitian ini, algoritma Multinomial Naive Bayes bekerja berdasarkan prinsip teorema Bayes dengan asumsi independensi fitur (kata) dalam teks. Mekanisme ini sangat efektif untuk klasifikasi teks karena kesederhanaan dan efisiensinya dalam menangani dataset besar dengan dimensi tinggi seperti data media sosial.

Proses klasifikasi sentimen menggunakan Naive Bayes pada penelitian ini meliputi beberapa tahapan utama:

1. Perhitungan Prior Probability: Model menghitung probabilitas awal (prior) untuk setiap kelas sentimen berdasarkan distribusi kelas dalam data latih. Dalam kasus ini, $P(\text{negatif}) = 0,64$, $P(\text{positif}) = 0,24$, dan $P(\text{netral}) = 0,12$. Prior probability ini mencerminkan distribusi awal sentimen dalam dataset.
2. Ekstraksi Fitur dengan TF-IDF: Setiap tweet direpresentasikan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dengan parameter ngram (1,2) dan maksimum 5000 fitur. Pendekatan ini memberikan bobot lebih tinggi pada kata-kata yang sering muncul pada kelas tertentu namun jarang pada kelas lain, meningkatkan kemampuan diskriminatif model.
3. Perhitungan Likelihood: Untuk setiap kata w dalam tweet baru, model menghitung probabilitas kemunculannya di setiap kelas sentimen menggunakan formula:

$$P(w|c) = (\text{count}(w,c) + \alpha) / (\sum(\text{count}(w',c) + \alpha))$$

Dimana α adalah parameter Laplace smoothing yang digunakan untuk mengatasi masalah zero probability pada kata-kata yang tidak muncul dalam data latih. Pendekatan ini mencegah model menghasilkan probabilitas nol ketika menemui kata baru.

4. Penerapan Teorema Bayes: Untuk mengklasifikasikan tweet baru, model menghitung probabilitas posterior untuk setiap kelas sentimen menggunakan perkalian dari prior probability dan likelihood dari setiap kata dalam tweet:

$$P(c|d) \propto P(c) \times \prod P(w_i|c)^{f_i}$$

Dimana f_i adalah frekuensi kemunculan kata w_i dalam tweet d . Logaritma natural sering diterapkan untuk menghindari underflow numerik akibat perkalian nilai-nilai probabilitas yang sangat kecil.

5. Pengambilan Keputusan: Tweet diklasifikasikan ke dalam kelas dengan nilai probabilitas posterior tertinggi:

$$c^* = \text{argmax } P(c|d)$$

Keunggulan pendekatan Multinomial Naive Bayes dalam penelitian ini terletak pada kemampuannya mempertimbangkan frekuensi kemunculan kata dan efisiensi komputasinya, yang sangat bermanfaat untuk klasifikasi real-time pada data media sosial yang berjumlah besar. Meskipun memiliki asumsi independensi yang "naif", model ini tetap memberikan performa klasifikasi yang baik dengan akurasi di atas 82%.

Lebih lanjut, laporan klasifikasi mengungkapkan performa per kelas:

1. Negatif: precision 0,80; recall 1,00; F1-score 0,89.
2. Netral: precision 0,91; recall 0,51; F1-score 0,65.
3. Positif: precision 0,93; recall 0,53; F1-score 0,68.

Skor macro F1 sebesar 0,74 dan weighted F1 0,81 menunjukkan bahwa model sangat andal dalam mengenali sentimen negatif (recall = 100%) tetapi memiliki tantangan pada kelas netral dan positif—kemungkinan disebabkan imbalance jumlah sampel dan kemiripan leksikal antara komentar netral dan positif.

- <https://doi.org/10.61132/neptunus.v2i3.249>
- Gunawan, Y., Purnama, I., & Rohani, R. (2024). Sentiment Analysis of Twitter towards the 2024 Indonesian Presidential Candidates Using the Naïve Bayes Algorithms. *International Journal of Science, Technology & Management*, 5(4), 953–961. <https://doi.org/10.46729/ijstm.v5i4.1154>
- Nufus, R. H., & Surapati, U. (2024). Analisis Sentimen Persepsi Masyarakat Terhadap Timnas Indonesia U-23 dalam AFC-23 Asian Cup 2024 Pada Media Sosial X Menggunakan Metode Naïve Bayes Classifier. *Jurnal Indonesia Manajemen Informatika Dan Komunikasi*, 5(3), 2647–2657.
- Prasetyo, H., & Fitriani, A. S. (2023). Sentiment Analysis Before Presidential Election 2024 Using Naïve Bayes Classifier Based On Public Opinion In Twitter (Analisa Sentimen Jelang Pilpres 2024 Menggunakan Naïve Bayes Classifier Berdasarkan Opini Publik Di Twitter). *Procedia of Engineering and Life Science*, 4(June 2023), 1–6.
- Purbolaksono, M. D., Reskyadita, F. D., & Suryani, A. A. (2020). Indonesian-text-classification-using-back-propagation-and-sastrawi-stemming-analysis-with-information-gain-for-selection-feature2020International-Journal-on-Advanced-Science-Engineering-and-Information-TechnologyOp.pdf, 10(1), 234–238.
- Ramdhan Hakiki, Pambudi, A., & Asriyanik. (2024). Classification of Public Sentiment Toward 2024 Presidential Candidates on Social Media Platform X Using Naïve Bayes Algorithm. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 3(2), 551–556. <https://doi.org/10.59934/jaiea.v3i2.422>