

Analisis Sentimen: Perbandingan Performa Algoritma Naive Bayes, Support Vector Machine, Random Forest, dan K-Nearest Neighbor Dalam Pemecatan Shin Tae Yong pada Media X

Agung Prasatya, Nirwana Hendrastuty*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹Agung_prasatya@teknokrat.ac.id, ^{2,*}nirwanahendrastuty@teknokrat.ac.id

Email Penulis Korespondensi: nirwanahendrastuty@teknokrat.ac.id

Email Penulis Korespondensi: emailpenuliskorespondensi@email.com

Submitted: 13/02/2025; Accepted: 16/03/2025; Published: 17/03/2025

Abstrak—Pemecatan Shin Tae Yong sebagai pelatih timnas Indonesia telah memicu reaksi yang beragam, mulai dari kekecewaan hingga kelegaan, di kalangan penggemar sepak bola Indonesia. Faktor-faktor seperti hasil pertandingan yang tidak memuaskan dan konflik internal dalam tim, serta tekanan dari penggemar dan media, menjadi penyebab utama keputusan ini. Meskipun perubahan ini membuka peluang bagi pelatih baru untuk meningkatkan kinerja timnas Indonesia, namun juga menimbulkan kontroversi dan perdebatan. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naive Bayes, *Random Forest*, dan *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) dalam analisis sentimen terkait pemecatan tersebut. Data penelitian diperoleh dari platform *Twitter* dengan total 4.345 *tweet* yang dikumpulkan menggunakan teknik *crawling*. Data tersebut kemudian melalui tahap pra-pemrosesan untuk menghasilkan data yang bersih. Pengujian dilakukan guna mengevaluasi tingkat akurasi masing-masing model dalam memprediksi sentimen publik. Hasil pengujian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan akurasi sebesar 78%, diikuti oleh *Random Forest* dengan akurasi 77%, dan Naive Bayes dengan akurasi 63% serta KNN 74% sebelum penerapan *Synthetic Minority Oversampling Technique* (SMOTE). Setelah optimasi menggunakan SMOTE, algoritma SVM tetap menunjukkan performa terbaik dengan akurasi 80%, diikuti oleh *Random Forest* dengan akurasi 79%, dan Naive Bayes serta KNN dengan akurasi 72%. Berdasarkan hasil tersebut, SVM terbukti menjadi algoritma paling efektif dalam mengklasifikasikan sentimen terkait pemecatan Shin Tae Yong. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam upaya memahami opini masyarakat terkait keputusan pemecatan Shin Tae Yong sebagai pelatih timnas Indonesia.

Kata Kunci: Shin Tae Yong; Analisis Sentimen; Algoritma *Naive Bayes*; SMOTE; Sepak Bola Indonesia

Abstract—The dismissal of Shin Tae Yong as the coach of the Indonesian national team has triggered a variety of reactions, ranging from disappointment to relief, among Indonesian football fans. Factors such as unsatisfactory match results and internal conflicts within the team, as well as pressure from fans and the media, were the main reasons for this decision. Although this change opens up opportunities for a new coach to improve the performance of the Indonesian national team, it also raises controversy and debate. This study aims to compare the performance of Naive Bayes, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) algorithms in analyzing sentiment related to this dismissal. The research data were obtained from the Twitter platform with a total of 4,345 tweets collected using crawling techniques. The data then underwent pre-processing stages to produce clean data. Testing was conducted to evaluate the accuracy of each model in predicting public sentiment. The test results showed that the SVM algorithm performed best with an accuracy of 78%, followed by Random Forest with an accuracy of 77%, and Naive Bayes with an accuracy of 63% and KNN 74% before the application of Synthetic Minority Oversampling Technique (SMOTE). After optimization using SMOTE, the SVM algorithm still showed the best performance with an accuracy of 80%, followed by Random Forest with an accuracy of 79%, and Naive Bayes and KNN with an accuracy of 72%. Based on these results, SVM proved to be the most effective algorithm in classifying sentiment related to the dismissal of Shin Tae Yong. It is hoped that the results of this study can contribute to understanding public opinion regarding the decision to dismiss Shin Tae Yong as coach of the Indonesian national team.

Keywords: Shin Tae Yong; Sentiment Analysis; Naive Bayes Algorithm; SMOTE; Indonesian Football

1. PENDAHULUAN

Perluasan media sosial dan platform online telah menghasilkan sejumlah besar data, termasuk opini dan perasaan masyarakat. Dalam penelitian ini, platform sosial yang digunakan adalah *Twitter*. Data yang dirilis oleh *We Are Social* menunjukkan bahwa, pada Oktober 2023, jumlah pengguna *Twitter* di seluruh dunia mencapai 666,2 juta, menempatkan platform tersebut di peringkat ke-12. *Twitter* juga sering digunakan untuk melakukan analisis data tentang tren dan opini publik[1]. *Twitter* atau X merupakan platform media sosial yang memungkinkan pengguna membagikan informasi, berbagi pengalaman, dan menyampaikan pandangan pribadi secara langsung dan interaktif[2].

Sepakbola adalah salah satu olahraga paling populer di dunia.[3] Di Indonesia, sepakbola memiliki makna yang lebih luas daripada sekadar hiburan, karena juga merupakan bagian penting dari budaya dan identitas nasional[4]. Sepak bola Indonesia dikelola dan dikembangkan oleh PSSI (Persatuan Sepakbola Seluruh Indonesia)[5]. Shin Tae-Yong yang merupakan pelatih yang ditunjuk oleh PSSI untuk memimpin Tim Nasional Sepak Bola Indonesia sejak tahun 2019[6]. Kinerja Timnas Indonesia di bawah asuhan Shin Tae Yong telah menjadi topik perbincangan hangat di media sosial X [2]. Pergantian pelatih timnas senior dan perubahan kepemimpinan telah memicu polemik dalam beberapa bulan terakhir. Hal ini memicu munculnya berbagai opini dan pendapat di media sosial, terutama di *Twitter*[7].

Dalam memahami reaksi masyarakat terhadap keputusan pemecatan Shin Tae Yong, analisis sentimen menjadi pendekatan yang dapat digunakan. Analisis sentimen memungkinkan pengidentifikasian dan pengklasifikasian opini publik yang tersebar melalui berbagai *platform digital*, seperti media sosial, forum diskusi, dan artikel berita[8]. Analisis sentimen merupakan teknik komputasi yang digunakan untuk menganalisis teks *digital* dan mengidentifikasi emosi atau sentimen yang terkandung di dalamnya. Hasil analisis ini kemudian dikelompokkan ke dalam dua kategori utama, yaitu positif dan negatif [4].

Beberapa algoritma klasifikasi teks yang populer adalah SVM, *Naive Bayes*, KNN, dan *Random Forest*. SVM bekerja dengan mencari batas *hyperplane* untuk memisahkan kelas, sementara *Naive Bayes* menggunakan probabilitas bersyarat untuk menentukan kelas. KNN mencari pola terdekat berdasarkan jarak[9], sedangkan dalam konteks algoritma *ensemble*, *Random Forest* merupakan salah satu contoh yang menggunakan kombinasi model *Decision Tree* untuk menghasilkan keputusan akhir yang lebih akurat[10].

Penelitian sebelumnya telah memanfaatkan analisis sentimen untuk memahami opini masyarakat terkait vaksinasi COVID-19. Penelitian tersebut menggunakan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN) untuk menganalisis sentimen masyarakat terhadap vaksin *Sinovac* berdasarkan data opini dari *Twitter* dengan keyword "*Sinovac*" yang diberi label secara otomatis menggunakan *TextBlob*. Hasilnya menunjukkan bahwa algoritma SVM memiliki performa yang lebih unggul dibandingkan KNN, dengan akurasi masing-masing 0,7 dan 0,56, sehingga menegaskan bahwa analisis sentimen dapat menjadi alat yang efektif untuk memahami opini masyarakat terkait vaksinasi COVID-19[11]. Selain itu, penelitian lainnya juga telah menggunakan algoritma *Random Forest* dan *Naive Bayes* untuk klasifikasi sentimen postingan media sosial. Penelitian tersebut menggunakan dataset besar dari *Twitter* dan menunjukkan bahwa algoritma *Naive Bayes* mencapai akurasi sebesar 90,41%, sedangkan algoritma *Random Forest* mencapai akurasi sebesar 39,74%. Hasil penelitian ini menunjukkan bahwa analisis sentimen dapat digunakan untuk memahami sentimen di media sosial, yang berguna bagi organisasi ketika merespons opini publik dan mengambil keputusan strategis[12].

Analisis sentimen masyarakat terkait Penelitian analisis sentimen masyarakat terkait PSSI di era Erick Thohir juga pernah di uji melalui media sosial menggunakan metode *Support Vector Machine* (SVM). Dalam penelitian ini, 3.900 *tweet* dikumpulkan melalui teknik data *crawling* dan dianalisis menggunakan metodologi *CRISP-DM*. Hasilnya menunjukkan bahwa sentimen positif mendominasi dengan 66%, diikuti sentimen netral 27%, dan sentimen negatif 7,1%. Penelitian ini juga menemukan bahwa skenario terbaik untuk pengujian model adalah dengan rasio 80:20 dan menggunakan bahasa Inggris, yang menghasilkan akurasi 76,6% [4]. Penelitian tambahan juga telah dilakukan untuk menganalisis sentimen masyarakat terhadap *gadget* Samsung Galaxy Z Flip 3 melalui komentar di *Youtube*. Dengan menggunakan model *CRISP-DM* dan membandingkan metode klasifikasi *Naive Bayes* (NB), *Support Vector Machine* (SVM), dan *k-Nearest Neighbor* (k-NN), penelitian ini menemukan bahwa model klasifikasi SVM menunjukkan hasil terbaik dengan rata-rata akurasi sebesar 96,43%. Sementara itu, algoritma *Naive Bayes* (NB) memiliki akurasi sebesar 85,12%, dan algoritma *k-Nearest Neighbor* (k-NN) memiliki akurasi sebesar 88,56%[13]. Penelitian sebelumnya juga telah menganalisis sentimen publik terkait pemindahan ibu kota Indonesia ke IKN Nusantara di Kalimantan. Penelitian ini menggunakan pendekatan *Natural Language Processing* untuk menguji akurasi tiga algoritma, yaitu *Naive Bayes Classifier*, *K-Nearest Neighbor*, dan *Random Forest*. Hasilnya menunjukkan bahwa sentimen negatif mencapai 38,1%, sentimen netral 25%, dan sentimen positif 36,8%, dengan akurasi algoritma yang berbeda-beda, yaitu *Naive Bayes Classifier* (65,26%), *K-Nearest Neighbor* (58,25%), dan *Random Forest* (45,05%). Berdasarkan hasil penelitian, kesimpulan yang memiliki tingkat akurasi tertinggi adalah *Naive Bayes Classifier* dengan akurasi sebesar 65,26%[14].

Penelitian ini berbeda dengan penelitian terdahulu yang membahas tentang analisis sentimen terkait vaksinasi COVID-19, PSSI di era Erick Thohir, *gadget* Samsung Galaxy Z Flip 3, dan pemindahan ibu kota Indonesia. Meskipun topik penelitian ini sama dengan beberapa penelitian terdahulu, yaitu analisis sentimen, namun penelitian ini memiliki perbedaan signifikan dalam hal akurasi dan penerapan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan kinerja algoritma. Selain itu, penelitian ini juga fokus pada perbandingan performa keempat algoritma, yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Naive Bayes*, dan *Random Forest*, dalam menganalisis sentimen terkait pemecatan Shin Tae-yong.

Berdasarkan Tujuan penelitian ini adalah untuk mengevaluasi kinerja empat algoritma berbeda, yaitu *Random Forest*, *Naive Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN), dalam menganalisis sentimen terkait Pemecatan Shin Tae-yong. Penelitian ini juga bertujuan untuk mengidentifikasi tren sentimen positif dan negatif serta membandingkan akurasi keempat algoritma dalam klasifikasi sentiment.

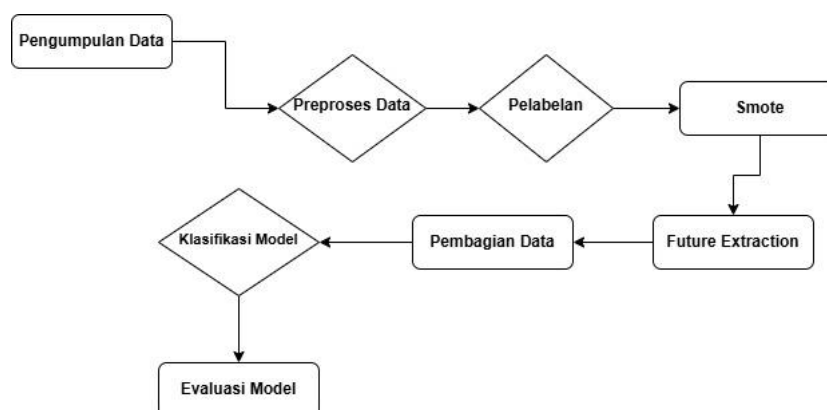
2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Proses penelitian ini terdiri dari 8 tahap, dimulai dengan pengumpulan data, kemudian *preprocessing*, *labeling* data untuk memperoleh sentimen negatif dan positif, teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi masalah ketidakseimbangan data. SMOTE bekerja dengan menghasilkan contoh sintesis dari kelas minoritas, sehingga meningkatkan jumlah contoh dalam kelas tersebut dan mengurangi ketidakseimbangan data. Ekstraksi fitur menggunakan *TF-IDF* dan pemisahan data menjadi set pelatihan dan pengujian. Penelitian ini

menggunakan empat algoritma, yaitu *Random Forest*, *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbor (KNN)*, untuk melakukan klasifikasi sentimen, di mana masing-masing algoritma memiliki kegunaan unik, seperti *Naïve Bayes* yang menawarkan kecepatan dan efisiensi, *SVM* yang efektif dalam mengatasi data berdimensi tinggi, *Random Forest* yang meningkatkan akurasi, dan *KNN* yang efektif dalam mengatasi data tidak terstruktur.

Dengan membandingkan keempat algoritma ini, penelitian ini bertujuan untuk memberikan pemahaman yang lebih baik tentang performa masing-masing algoritma dan berkontribusi secara praktis dalam membantu PSSI merancang strategi ke depan, terutama dalam pengambilan keputusan terkait pelatih Timnas Indonesia kedepan, menggunakan data Twitter dan melakukan evaluasi model dengan menggunakan matriks untuk *accuracy*, *precision*, *recall*, dan *F1-Score* menggunakan matriks untuk *accuracy*, *precision*, *recall*, dan *F1-Score*. Gambar 1 merupakan alur penelitian



Gambar 1. Tahapan Penelitian

2.3 Pengumpulan Data

Pengumpulan data dilakukan dengan melakukan crawling pada *Twitter* menggunakan metode *Tweet Harvest* pada platform *Google Colab* dengan bahasa pemrograman *Python*. *Crawling* dilakukan dengan mengakses *API Twitter* untuk mengumpulkan *tweet* yang mengandung kata kunci "Shin Tae Yong", "PSSI", "Timnas Indonesia", "Pemecatan Shin Tae Yong", dan kata kunci terkait lainnya pada periode tertentu. Hasil *crawling* ini kemudian diolah dan disimpan dalam format file CSV, dengan total jumlah data sebesar 4365, untuk selanjutnya diolah dalam tahap preprocessing.

Crawling data di *Twitter* merupakan proses pengambilan data yang memanfaatkan *API Twitter* untuk mengumpulkan informasi tentang pengguna atau *tweet* tertentu. Proses ini dapat dilakukan dengan membuat program khusus yang dirancang untuk mencari data berdasarkan kata kunci yang telah ditentukan [15]. *Tweet-harvest* adalah sebuah perpustakaan *open-source* yang memungkinkan pengguna mengumpulkan *tweet* berdasarkan kata kunci spesifik dan jumlah *tweet* yang diinginkan, sehingga memudahkan pengguna untuk mengumpulkan data *tweet* yang relevan [15].

2.4 Pre-processing

Pre-Proses Data Bahasa *Twitter* merupakan bahasa yang memiliki banyak karakteristik yang tidak penting, sehingga perlu dihilangkan untuk memperkecil ruang fitur dan meningkatkan efisiensi analisis sentimen [15]. Tujuan dari proses *preprocessing* data adalah untuk meningkatkan kualitas data sehingga dapat meningkatkan akurasi model prediksi [16]. Dalam penelitian ini, beberapa metode preprocessing yang digunakan yaitu

2.4.1 Cleaning

Data *cleaning* adalah proses penghilangan atau pembersihan data dari komponen-komponen yang tidak relevan, tidak penting, atau yang dapat mempengaruhi kualitas data, seperti *URL*, nama pengguna, *retweet*, kode HTML, dan hashtag [14].

2.4.2 Case Folding

Dalam dokumen teks, metode *Case folding* adalah proses mengonversi semua teks dalam dokumen menjadi huruf kecil untuk memudahkan pengolahan data [17].

2.4.3 Tokenizing

Tokenizing adalah proses pemecahan kalimat menjadi kata-kata individu. Setelah proses cleansing dan *case folding*, kalimat *tweet* kemudian dipecah menjadi kata-kata yang terpisah [18].

2.4.4 Filtering / Stopword removal

Stopword removal adalah teknik preprocessing yang digunakan untuk menghilangkan kata-kata umum yang tidak memiliki makna penting dalam konteks analisis teks, seperti "ke", "dari", "dan", atau "atau". Tujuan utamanya adalah

untuk menghindari pengaruh kata-kata tersebut pada hasil analisis dan meningkatkan efisiensi proses Information Retrieval (IR) dan text mining[19].

2.4.5 Stemming

Stemming adalah proses mengubah kata-kata yang memiliki imbuhan menjadi bentuk kata dasarnya dengan menghilangkan awalan dan akhiran. Tujuan utama stemming adalah untuk memperoleh kata dasar yang murni, sehingga dapat memudahkan proses analisis dan pengolahan teks dalam sistem[20].

2.5 Pelabelan Data

Pelabelan adalah proses membagi data ke dalam beberapa kategori atau kelas sentimen tertentu[21]. Pelabelan dalam penelitian ini digunakan untuk mengklasifikasikan data menjadi dua kelas sentimen, yaitu positif dan negatif, menggunakan pustaka TextBlob, yaitu sebuah pustaka Python untuk analisis sentimen dan pengolahan bahasa alami yang bekerja dengan cara menganalisis teks dan memberikan skor sentimen berdasarkan kata-kata dan frasa yang digunakan, yang menghasilkan pelabelan data sebanyak 3462 komentar, terdiri dari 2465 komentar negatif dan 996 komentar positif.

2.6 SMOTE

SMOTE (*Synthetic Minority Oversampling Technique*) adalah metode yang digunakan untuk mengatasi ketidakseimbangan kelas dalam dataset. Ketika jumlah data pada satu kelas jauh lebih sedikit daripada kelas lain, model pembelajaran mesin dapat menjadi bias. SMOTE mengatasi masalah ini dengan menciptakan sampel sintetis untuk kelas minoritas melalui interpolasi data yang sudah ada. Dengan demikian, proporsi kelas minoritas dalam dataset meningkat, sehingga model pembelajaran mesin dapat memahami pola kelas tersebut lebih baik dan menghasilkan prediksi yang lebih akurat[22]. Dalam kasus data pada penelitian ini yang berjumlah 3462, dengan komposisi 2465 komentar negatif dan 996 komentar positif, SMOTE dapat diterapkan dengan cara memilih sampel acak dari kelas minoritas (komentar positif), mencari sampel yang paling mirip, membuat sampel sintetis baru melalui interpolasi, dan mengulangi proses hingga jumlah sampel sintetis yang diinginkan tercapai, sehingga total data menjadi 4962 (2465 komentar negatif + 996 komentar positif + 1469 sampel sintetis), dan model pembelajaran mesin dapat dilatih dengan dataset yang telah diperbarui untuk menghasilkan prediksi yang lebih akurat

2.7 Ekstraksi Fitur

Karena mesin pembelajaran (*machine learning*) tidak dapat memahami makna kata dan karakter secara langsung, maka dilakukan ekstraksi fitur (*feature extraction*) untuk mengubah teks menjadi bentuk yang dapat dipahami oleh mesin. Dalam hal ini, metode *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) digunakan untuk menghitung bobot kata dalam teks berdasarkan frekuensinya dalam dokumen[23].

a. Term Frequency (IDF)

$$TF(d, t) = \frac{\text{jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{total jumlah term dalam dokumen } d} \quad (1)$$

TF (*Term Frequency*) mengukur seberapa sering sebuah kata muncul dalam sebuah teks dibandingkan dengan total jumlah kata dalam teks tersebut.

b. Inverse Document Frequency (IDF)

$$IDF(t, d) = \log \left(\frac{\text{jumlah total dokumen } d}{\text{jumlah dokumen dalam } D \text{ yang mengandung term } t+1} \right) \quad (2)$$

c. TF-IDF

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

Dengan menggabungkan nilai TF dan IDF, maka dihasilkan TF-IDF. TF-IDF memberikan skor yang lebih akurat dengan mempertimbangkan dua faktor, yaitu seberapa sering sebuah istilah muncul dalam sebuah dokumen (TF) dan seberapa luas istilah tersebut digunakan di seluruh koleksi dokumen (IDF).

2.8 Pembagian Data

Dalam proses pembuatan model, membagi data menjadi dua kategori, yaitu data pelatihan dan data uji, sangatlah penting. Data pelatihan digunakan untuk melatih model agar dapat membedakan antara kategori positif dan negatif, sedangkan data uji digunakan untuk mengevaluasi kinerja model tersebut. Dalam penelitian ini, pembagian data dilakukan dengan alokasi 80% untuk pelatihan dan 20% untuk pengujian, yang diharapkan dapat menghasilkan hasil yang lebih optimal.

2.9 Klasifikasi Model

Model ini belajar dari data pelatihan untuk mengidentifikasi pola-pola tertentu, lalu menerapkan pengetahuan tersebut untuk memprediksi kategori atau label dari data yang belum pernah dilihat sebelumnya. Keakuratan

model ini sangat tergantung pada kualitas data pelatihan, efektivitas algoritma yang digunakan, serta kemampuan model untuk menggeneralisir pola-pola yang dipelajari ke dalam data baru:

2.9.1 Naïve Bayes

Naïve Bayes adalah salah satu metode pembelajaran mesin yang menggunakan prinsip probabilitas dan statistik yang dikembangkan oleh Thomas Bayes. Metode ini bekerja dengan memprediksi kemungkinan kejadian di masa depan berdasarkan data atau pengalaman sebelumnya. Dasar dari *Naïve Bayes* adalah rumus *Bayes*, yang digunakan sebagai landasan dalam pemrograman[24]. Bentuk umum teorema bayes ditunjukkan pada Persamaan 3.

$$P(Y\setminus X) = \frac{p(X\setminus Y).p(y)}{p(x)} \tag{4}$$

Dalam persamaan (1), variabel X mewakili data yang belum diklasifikasikan, sedangkan Y mewakili kelas tertentu. Probabilitas P(Y|X) menggambarkan kemungkinan kelas Y diberikan kondisi X, sementara P(X|Y) menunjukkan kemungkinan awal dari suatu kelas berdasarkan hipotesis yang ada. Selain itu, P(Y) dan P(X) masing-masing mewakili probabilitas hipotesis Y dan probabilitas data X.[25]

2.9.2 Support Vector Machine (SVM)

Merupakan salah satu teknik klasifikasi yang populer saat ini. Prinsip utama dari metode ini adalah mencari garis pemisah (*hyperplane*) yang optimal untuk membedakan antara kelas-kelas data yang berbeda, dengan cara memaksimalkan jarak antara garis pemisah dan data-data yang paling dekat dengannya[26]. Persamaan 4 dan 5 menyajikan *hyperlane* ideal pada *kernel linear*.

$$(w \cdot x_i + b) \leq 1, y_i = -1 \tag{5}$$

$$(w \cdot x_i + b) \geq 1, y_i = 1 \tag{6}$$

Dalam persamaan, x_i melambangkan data ke- i , $w \cdot x_i$ mewakili bobot atau koefisien untuk kelas data ke- i , b adalah konstanta bias, dan y_i adalah label atau kelas dari data ke- i [27]

2.9.3 Random Forest

Model *Random Forest* menggabungkan dua teknik, yaitu bootstrap aggregating (*bagging*) dan seleksi fitur acak. Model ini merupakan kombinasi dari beberapa teknik pohon keputusan yang digabungkan untuk meningkatkan akurasi prediksi. Terdapat tiga langkah utama dalam metode *Random Forest*, yaitu: melakukan pengambilan sampel acak dengan metode bootstrap untuk membangun pohon prediksi, setiap pohon keputusan menggunakan prediktor acak untuk memprediksi hasil, dan menggabungkan hasil prediksi dari setiap pohon keputusan melalui pemungutan suara mayoritas untuk klasifikasi atau perhitungan rata-rata untuk regresi.[27] Persamaan 6 dapat digunakan untuk menentukan kelas yang paling banyak dipilih oleh model *Random Forest*.

$$f(x) = Average(f_1(x), f_2(x), \dots, f_n(x)) \tag{6}$$

Persamaan ini menjelaskan bahwa hasil prediksi ($f(x)$) adalah rata-rata dari hasil prediksi dari setiap pohon keputusan ($f_1-n(x)$) yang diberikan inputan data (x).

2.9.4 KNN (K-Nearest Neighbor)

Algoritma klasifikasi terakhir yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN). KNN adalah algoritma yang mengklasifikasikan data baru berdasarkan kesamaannya dengan data-data terdekat yang sudah ada. Cara kerjanya adalah dengan mencari sejumlah data terdekat (k) dari data masukan, lalu menentukan kelasnya berdasarkan kelas yang paling banyak muncul di antara data-data tersebut. Rumus jarak yang sering digunakan dalam algoritma KNN adalah:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \tag{7}$$

dimana $d(x, y)$ adalah jarak antara objek x dan y , dan n adalah jumlah atribut yang digunakan

2.10 Evaluasi Model

Salah satu cara untuk menilai kinerja model pembelajaran mesin adalah dengan melakukan evaluasi menggunakan Confusion Matrix. Matriks ini berbentuk tabel yang memuat informasi tentang akurasi dan kesalahan prediksi model. Tabel ini memuat empat elemen yang menjelaskan kinerja model, seperti yang ditunjukkan pada Tabel 1

Tabel 1. Confusion Matrix

	Prediction	
Actual	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Keterangan:

True Positive (TP) : Banyaknya data yang berhasil diidentifikasi dengan tepat sebagai positif oleh model.

False Positive (FP) : Banyaknya data yang keliru diidentifikasi sebagai positif oleh model.

True Negative (TN) : Banyaknya data yang berhasil diidentifikasi dengan tepat sebagai negatif oleh model.

False Negative (FN) : Banyaknya data yang keliru diidentifikasi sebagai negatif oleh model.

Persamaan 7 dapat digunakan untuk mengukur akurasi sistem

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{7}$$

Nilai presisi, recall, dan skor F1 dapat ditentukan dengan menggunakan rumus yang terdapat dalam persamaan 8, 9, dan 10. Rasio TP terhadap semua data positif yang diantisipasi dikenal sebagai nilai presisi, dan rasio TP terhadap semua data aktual dikenal sebagai nilai recall. Nilai presisi dan recall yang lebih tinggi ditunjukkan oleh nilai skor F1 yang lebih tinggi.

$$Precision = \frac{TP}{TF+FP} \tag{8}$$

$$Recall = \frac{TF}{TF+FN} \tag{9}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Rumus tersebut dapat digunakan sebagai acuan untuk membandingkan kinerja model yang telah dikembangkan.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan Dataset diperoleh dari platform X berdasarkan opini pengguna X tentang Pemecatan Shin Tae Yong. Pengumpulan data dilakukan melalui metode crawling dengan kata kunci seperti "PSSI", "Shin Tae Yong", "Timnas Indonesia", "Pemecatan Shin Tae Yong", dan lain-lain, menghasilkan 4345 data tweets. Hasil crawling kemudian disimpan dalam format csv untuk dianalisis sentiment pada Tabel 1

Tabel 1. Hasil Pengumpulan Data

Username	Tweet
ranggavega	Yakin Pak @erickthohir ? Jangan hanya mengikuti politik dan ego ya Pak. Mohon amat sangat dipertimbangkan. Saya rasa mayoritas penggemar Timnas mendukung STY STAY! Makasih Pak #STYSTAY
MafiaWasit	Ga sih Gua tetap dukung Timnas Indonesia Bukan dukung pemain Bukan dukung pelatih Bukan dukung Ketum PSSI Bahkan ketika ada Timnas KPSi dengan skuad mewahnya gua tetap dibarisan Timnas Indonesia meskipun pemainnya ga jelas dan dibantai Bahrain. https://t.co/cntzqrog05

3.2 Preprocessing

Proses *preprocessing* data melibatkan beberapa tahap, yaitu membersihkan data (*Cleaning*), mengubah huruf menjadi lowercase (*Casefolding*), memecah teks menjadi kata-kata (*Tokenizing*), menghilangkan kata-kata yang tidak penting (*Filtering/Stopword Removal*), dan mereduksi kata-kata ke bentuk dasarnya (*Stemming Data*)

Tabel 2. Hasil Preprocessing

Tahapan	hasil
<i>Dataset</i>	Yakin Pak @erickthohir ? Jangan hanya mengikuti politik dan ego ya Pak. Mohon amat sangat dipertimbangkan. Saya rasa mayoritas penggemar Timnas mendukung STY STAY! Makasih Pak #STYSTAY
<i>Cleaning</i>	Yakin Pak Jangan hanya mengikuti politik dan ego ya Pak Mohon amat sangat Saya rasa mayoritas penggemar Timnas mendukung STY STAY Makasih Pak STYSTAY
<i>Casefolding</i>	yakin pak jangan hanya mengikuti politik dan ego ya pak mohon amat sangat saya rasa mayoritas penggemar timnas mendukung sty stay makasih pak stystay
<i>Tokenizing</i>	['yakin', 'pak', 'jangan', 'hanya', 'mengikuti', 'politik', 'dan', 'ego', 'ya', 'pak', 'mohon', 'amat', 'sangat', 'saya', 'rasa', 'mayoritas', 'penggemar', 'timnas', 'mendukung', 'sty', 'stay', 'makasih', 'pak', 'stystay']
<i>Filtering/Stopword Removal</i>	['mengikuti', 'politik', 'ego', 'ya', 'mohon', 'mayoritas', 'penggemar', 'timnas', 'mendukung', 'sty', 'stay', 'makasih', 'stystay']

Stemming

ikut politik ego ya mohon mayoritas gemar timnas dukung sty stay makasih stystay

Setelah melakukan beberapa tahap preprocessing data, dapat dilihat pada tabel 2. yaitu perubahan dari setiap tahapan dan jumlah data dalam dataset berkurang secara signifikan. Seperti yang terlihat pada Tabel 2, dari total 4345 data awal, hanya 3462 data yang berhasil lolos proses pembersihan data. Dengan demikian, terdapat pengurangan sebanyak 883 data. Data bersih yang diperoleh ini kemudian akan digunakan untuk pengujian selanjutnya

3.3 Pelabelan Data

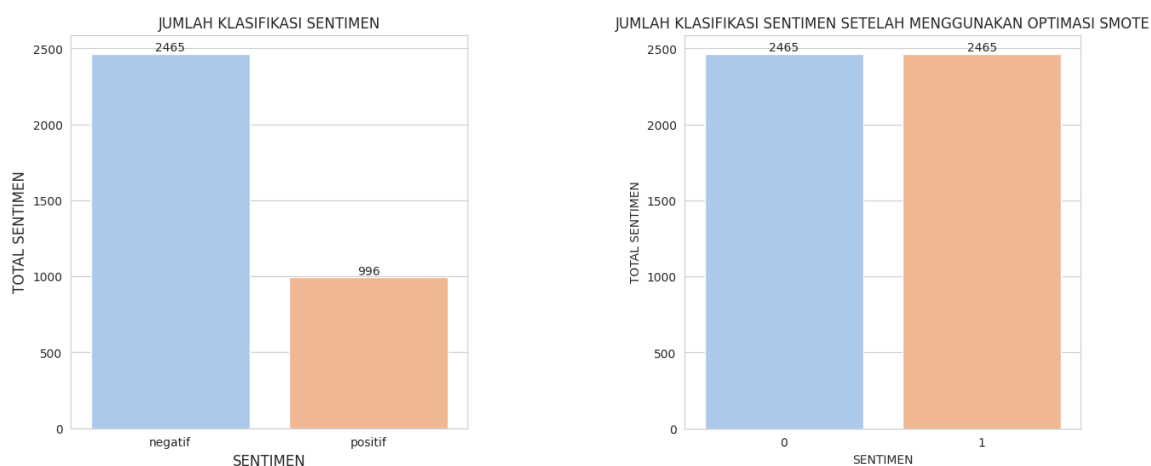
Proses pelabelan data menggunakan pustaka TextBlob menghasilkan dua kelas sentimen, yaitu negatif dan positif. Dari 3462 data tweets, 2465 data memiliki sentimen negatif dan 996 data memiliki sentimen positif, seperti yang ditunjukkan pada Gambar 2. Namun, data yang dihasilkan cenderung berlabel negatif, menyebabkan ketidakseimbangan yang mempengaruhi kinerja model dalam mengklasifikasikan sentimen. Oleh karena itu, penelitian ini menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan data, dengan perbandingan sebelum dan sesudah implementasi teknik SMOTE yang dapat dilihat pada Gambar 2 dan Gambar 3.

Tabel 3. Hasil Pelabelan Data

Tweet	Label
ikut politik ego ya mohon mayoritas gemar timnas dukung sty stay makasih stystay	Negatif
ga sih gua dukung timnas indonesia dukung main dukung latih dukung tum pssi timnas kpsi skuad mewah gua baris timnas indonesia main ga bantai bahrain	Positif

3.4 Perbandingan Akurasi

Hasil pre-processing data menunjukkan bahwa jumlah data dengan sentimen negatif (2465) jauh lebih banyak daripada data dengan sentimen positif (996). Ketidakseimbangan ini berpotensi mengganggu kinerja model dalam mengklasifikasikan emosi positif. Oleh karena itu, penelitian ini menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan tersebut. SMOTE (Synthetic Minority Oversampling Technique) adalah metode yang digunakan untuk mengatasi ketidakseimbangan kelas dalam dataset. Dalam kasus data Anda yang berjumlah 3462, dengan komposisi 2465 komentar negatif dan 996 komentar positif, SMOTE dapat diterapkan dengan cara memilih sampel acak dari kelas minoritas (komentar positif), mencari sampel yang paling mirip, membuat sampel sintesis baru melalui interpolasi, dan mengulangi proses hingga jumlah sampel sintesis yang diinginkan tercapai, sehingga total data menjadi 4962 (2465 komentar negatif + 996 komentar positif + 1469 sampel sintesis), dan model pembelajaran mesin dapat dilatih dengan dataset yang telah diperbarui untuk menghasilkan prediksi yang lebih akurat. Perbandingan hasil sebelum dan sesudah implementasi SMOTE dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Sebelum dan sesudah SMOTE.

Berdasarkan Gambar 2, dapat dilihat bahwa perbandingan sebelum dan sesudah implementasi SMOTE. Untuk mengatasi ketidakseimbangan data pada kelas minoritas, ditambahkan sebanyak 1469 data untuk kelas positif. Penambahan data ini bertujuan agar model dapat mempelajari sentimen positif dan negatif secara seimbang. Setelah penerapan teknik SMOTE ini, data untuk sentimen positif dan negatif menjadi seimbang, yaitu 2465 data untuk masing-masing kelas.

Setelah data berhasil diseimbangkan, langkah selanjutnya adalah melatih dan menguji model algoritma. Dalam penelitian ini, algoritma Naïve Bayes, Random Forest, dan SVM dilatih dan diuji menggunakan data yang sama, dengan pembagian 80% untuk pelatihan dan 20% untuk pengujian. Data yang digunakan mencakup data sebelum dan sesudah penerapan SMOTE, sehingga hasilnya dapat dibandingkan untuk menentukan algoritma yang memiliki

pendekatan performa terbaik. Hasil dari model algoritma sebelum dan sesudah penerapan SMOTE dapat dilihat pada Tabel 9 dan Tabel 10.

Tabel 4. Hasil Akurasi Before SMOTE

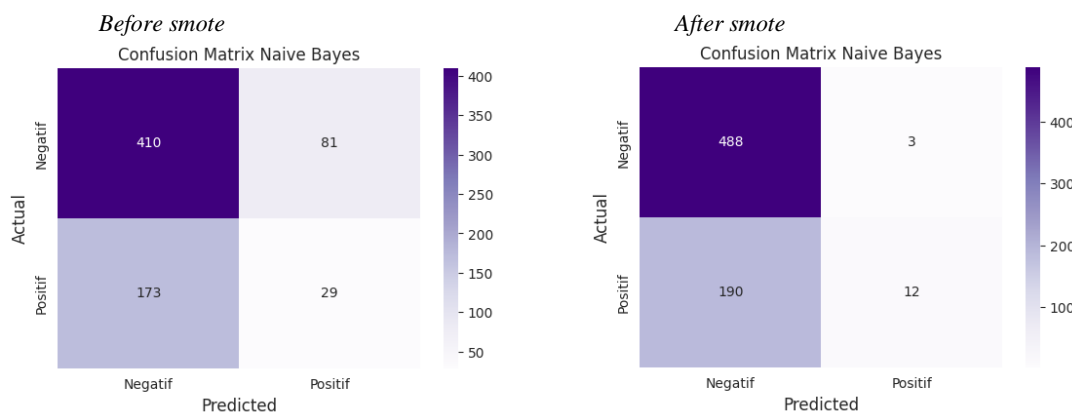
Model	Class	Accuracy	Precision	Recall	F1-Score
NBC	Negatif	63%	70%	84%	76%
	Positif		26%	14%	19%
RF	Negatif	77%	77%	96%	86%
	Positif		77%	32%	45%
SVM	Negatif	78%	80%	93%	86%
	Positif		69%	40%	50%
KNN	Negatif	74%	75%	95%	84%
	Positif		67%	23%	34%

Hasil pengujian model sebelum penerapan SMOTE dengan rasio pembagian data 80:20 untuk proses pengujian menunjukkan bahwa algoritma SVM memiliki tingkat akurasi tertinggi sebesar 78%, diikuti oleh Random Forest dengan akurasi 77%, KNN dengan akurasi 74%, dan Naïve Bayes dengan akurasi 63%. Dengan demikian, SVM merupakan algoritma yang paling akurat di antara keempat algoritma yang diuji..

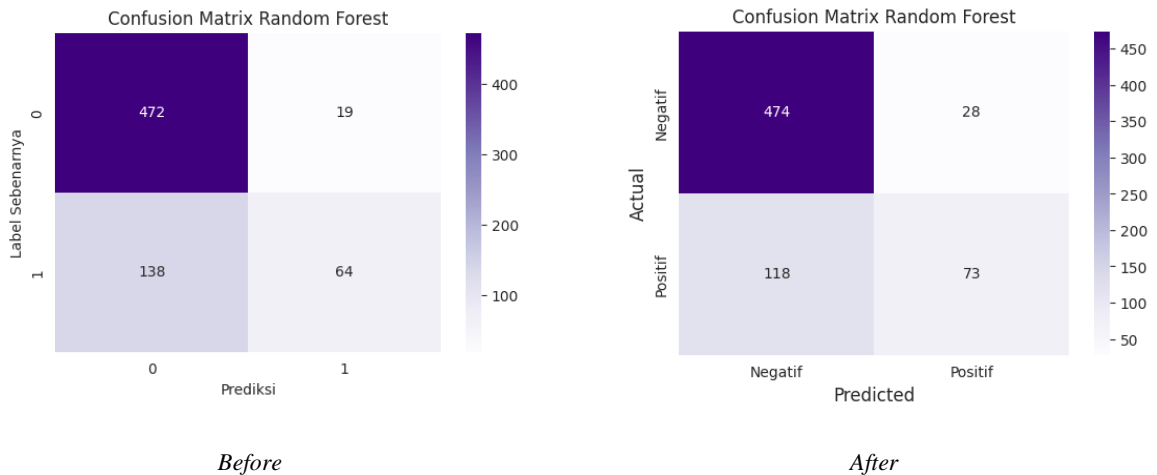
Tabel 5. Hasil Akurasi After SMOTE

Model	Class	Accuracy	Precision	Recall	F1-Score
NBC	Negatif	72%	72%	99%	83%
	Positif		80%	06%	11%
RF	Negatif	79%	80%	94%	87%
	Positif		72%	38%	50%
SVM	Negatif	80%	86%	85%	86%
	Positif		62%	64%	63%
KNN	Negatif	72%	72%	100%	84%
	Positif		91%	5%	9%

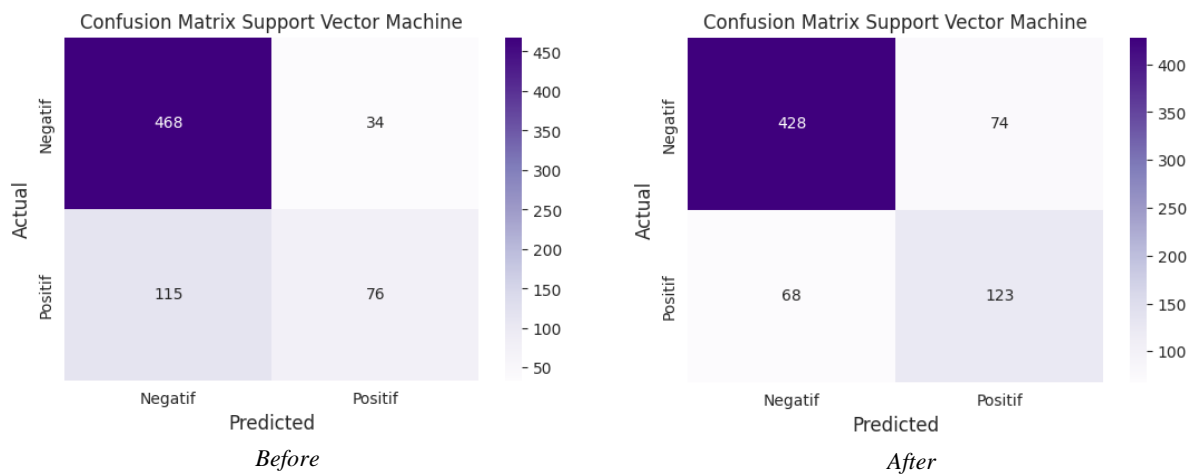
Penerapan SMOTE menyebabkan perubahan pada nilai evaluasi keempat algoritma. Algoritma Naïve Bayes, Random Forest, dan SVM menunjukkan peningkatan akurasi, namun algoritma KNN mengalami penurunan akurasi dari 74% menjadi 72%. Untuk menentukan pendekatan algoritma yang paling efektif, dilakukan perbandingan menggunakan matriks kebingungan (confusion matrix) di antara keempat algoritma.



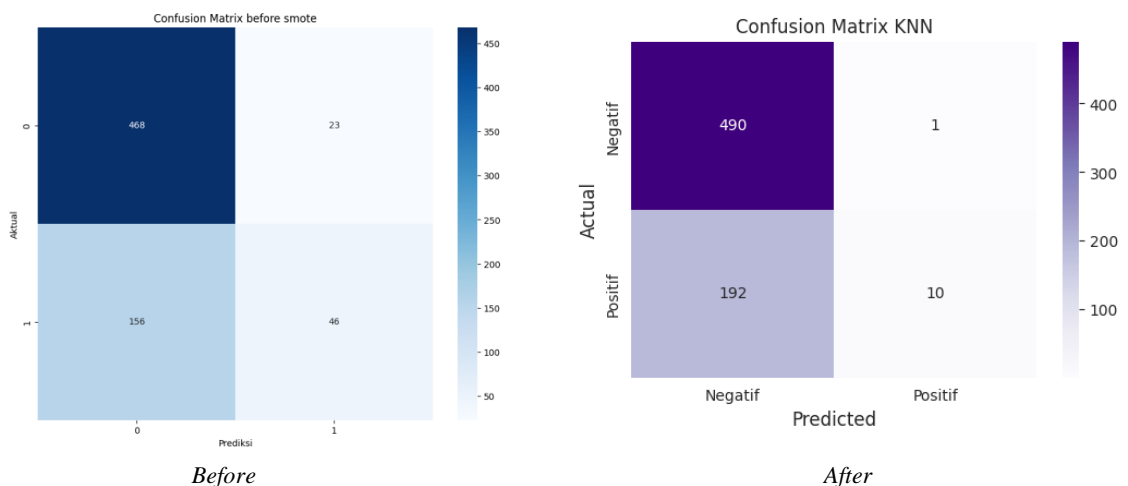
Gambar 3. Perbandingan Confusion Matrix Algoritma Naïve Bayes sebelum dan Sebelum SMOTE



Gambar 4. Perbandingan Confusioan Matrix Algoritma Random Ferest Sebelum dan Sesudah SMOTE



Gambar 5. Perbandingan Confusion Matrix Algoritma SVM Sesudah dan Sebelum SMOTE



Gambar 6. Perbandingan Confusion Matrix Algoritma KNN Sesudah dan Sebelum SMOTE

Penerapan SMOTE memiliki dampak signifikan terhadap kinerja algoritma dengan hasil yang bervariasi antara peningkatan dan penurunan. Pada Naïve Bayes, True Positif menurun dari 29 menjadi 12, sementara True Negatif meningkat dari 410 menjadi 488 dengan peningkatan kesalahan pada kelas negatif. Random Forest menunjukkan penurunan True Negatif dari 474 menjadi 472 dan True Positif dari 73 menjadi 64, dengan penurunan akurasi True Negatif dan True Positif setelah SMOTE. SVM mencatat peningkatan True Negatif dari 428 menjadi 468 dengan penurunan True Positif dari 123 menjadi 76. KNN menunjukkan peningkatan True Negatif dari 468 menjadi 490 dan penurunan True Positif dari 46 menjadi 10. Keempat model lebih efektif dalam mendeteksi kelas negatif, namun kurang akurat pada kelas positif. Naïve Bayes memiliki kinerja terendah, sementara SVM paling unggul. Setelah

SMOTE, akurasi SVM meningkat signifikan dalam mendeteksi kelas positif. Sedangkan KNN mengalami penurunan akurasi. SVM mencapai hasil akurasi hampir tanpa kesalahan pada True Negatif, Random Forest mendapatkan hasil akurasi yang lumayan, namun masih di bawah model SVM. Naïve Bayes masih lemah dalam mengklasifikasikan kelas. Secara keseluruhan, SMOTE efektif meningkatkan kinerja model pada data tidak seimbang, terutama untuk kelas positif.

3.5 Visualisasi Wordcloud

Wordcloud merupakan suatu teknik visualisasi data yang memungkinkan pengidentifikasian kata-kata yang paling sering digunakan dalam suatu dokumen. Dalam penelitian ini, wordcloud dibuat menggunakan pustaka matplotlib Python. Hasil visualisasi wordcloud dapat dilihat pada Gambar 3, yang menampilkan kata-kata yang paling banyak muncul dalam dataset.

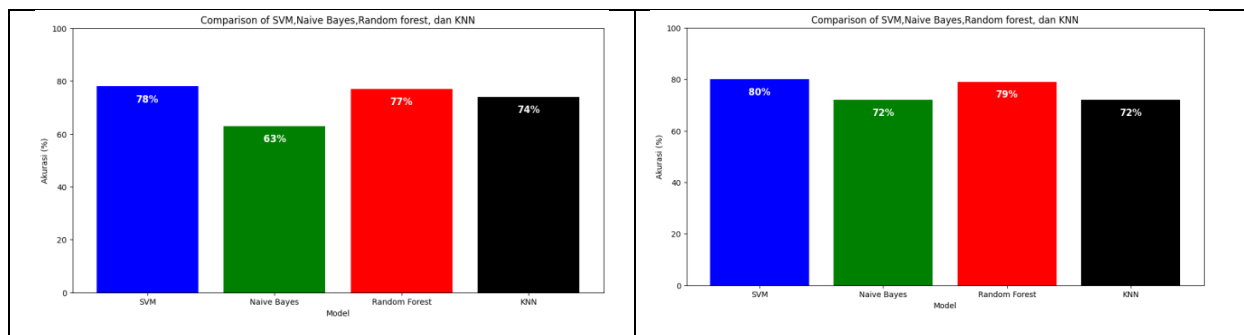


Gambar 7. Worldcloud

Gambar 7 memperlihatkan *wordcloud* yang merupakan gabungan dari sentimen negatif dan positif terkait pemecatan Shin Tae-yong. Kata-kata yang paling menonjol adalah "Shin Tae-yong", "Patrick Kluivert", "Timnas Indonesia", "Erick Thohir", "Latih Timnas", "STY", "Piala Dunia", dan "Pecat". Istilah-istilah ini mencerminkan berbagai reaksi dan asumsi masyarakat terhadap keputusan pemecatan Shin Tae-yong, yang mencakup sentimen positif dan negatif.

3.5 Visualisasi Perbandingan Model

Visualisasi model dalam penelitian ini memperlihatkan perbandingan akurasi antara empat model klasifikasi, yaitu Naïve Bayes, SVM, KNN, dan Random Forest. Penerapan SMOTE menyebabkan variasi kinerja pada setiap metode. Naïve Bayes, Random Forest, dan SVM menunjukkan peningkatan akurasi setelah penerapan SMOTE, yaitu dari 63% menjadi 72%, dari 77% menjadi 79%, dan dari 78% menjadi 80%. Namun, KNN mengalami penurunan akurasi dari 74% menjadi 72%. Secara keseluruhan, SMOTE terbukti efektif dalam meningkatkan akurasi pada tiga model, meskipun mengalami penurunan akurasi pada KNN. Perbandingan akurasi model dapat dilihat pada Gambar 8.



Gambar 8. Visualisasi sebelum dan sesudah smote

Dalam penelitian ini, kami menggunakan empat algoritma berbeda, yaitu *Random Forest*, *Naive Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (KNN)*, untuk menganalisis sentimen terkait Pemecatan Shin Tae-yong. Hasil analisis menunjukkan bahwa algoritma SVM memiliki akurasi tertinggi Setelah penerapan *Synthetic Minority Oversampling Technique (SMOTE)*. Setelah optimasi menggunakan SMOTE sebesar 80%, diikuti oleh algoritma Random Forest 79%, Terakhir Naïve Bayes, dan KNN yang sama-sama memiliki akurasi sebesar 72% Berdasarkan penelitian yang telah dilakukan, didapatkan kesimpulan bahwa penerapan SMOTE pada algoritma K-Nearest Neighbor (K-NN) untuk klasifikasi sentimen terhadap pemecatan shin tae-yong menyebabkan penurunan performa model. Terjadi penurunan dalam accuracy sebesar 2%.

Hasil ini berbeda dengan beberapa penelitian terdahulu, seperti "Analisis Sentimen Masyarakat Terkait Vaksinasi COVID-19" yang menunjukkan bahwa algoritma SVM memiliki performa yang lebih unggul, serta "Analisis Sentimen Masyarakat Terkait PSSI" yang menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi tertinggi. Namun, perbedaan hasil ini dapat dijelaskan oleh beberapa faktor, seperti dataset yang berbeda, preprocessing data yang berbeda, dan parameter algoritma yang berbeda. Hasil yang sama dengan beberapa penelitian terdahulu seperti penerapan SMOTE pada algoritma K-Nearest Neighbor (K-NN) untuk klasifikasi sentimen terhadap pertandingan Indonesia vs Uzbekistan pada Piala Asia U-23 menyebabkan penurunan performa model. Terjadi penurunan dalam accuracy sebesar 10%, precision sebesar 16%, recall sebesar 3%, dan F1-Score sebesar 11%. Dari analisis ini juga terungkap bahwa sentimen masyarakat terhadap hasil pertandingan cenderung negatif. Kesimpulan: Penerapan SMOTE pada algoritma K-NN tidak selalu meningkatkan kinerja model, bahkan dapat menyebabkan penurunan kinerja.[28] Dalam konteks penelitian ini, hasil analisis sentimen menggunakan algoritma SVM dapat digunakan untuk memahami opini masyarakat terkait Pemecatan Shin Tae-yong. Hasil ini juga dapat digunakan sebagai acuan untuk pendekatan pengambilan keputusan strategis dalam konteks yang lebih luas

4. KESIMPULAN

Penelitian ini berhasil mengumpulkan 4.345 data tweet melalui proses crawling, yang kemudian diproses dan diberi label untuk menghasilkan 3.462 data bersih. Sebelum melakukan teknik SMOTE, hasil analisis menunjukkan bahwa 2465 entri tweet memiliki sentimen negatif, sedangkan 996 entri tweet memiliki sentimen positif. Untuk mengatasi masalah ketidakseimbangan data, teknik SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan untuk meningkatkan jumlah contoh pada kelas minoritas (sentimen positif). Setelah melakukan SMOTE, jumlah data negatif dan positif menjadi seimbang, yaitu 2465 entri tweet untuk sentimen negatif dan 2465 entri tweet untuk sentimen positif. Dalam hal performa, algoritma *Support Vector Machine (SVM)* menunjukkan akurasi tertinggi sebesar 80% dalam mengkategorikan *tweet* terkait opini tentang pemecatan Shin Tae Yong, diikuti oleh Random Forest dengan akurasi 79%, dan *Naïve Bayes* serta KNN dengan akurasi 72%. Hasil ini menunjukkan bahwa SVM adalah algoritma yang paling efektif dalam mengidentifikasi sentimen terkait topic. Dari temuan ini, dapat disimpulkan bahwa pemilihan algoritma yang tepat dalam analisis sentimen sangat penting untuk memahami opini publik secara akurat. Hasil ini dapat dimanfaatkan oleh PSSI untuk menyempurnakan strategi Timnas Indonesia ke depan, terutama dalam mencapai target bermain di Piala Dunia. Dengan memahami sentimen negatif yang lebih dominan terkait keputusan pemecatan Shin Tae Yong, PSSI dapat mengambil langkah-langkah yang tepat untuk memperbaiki citra, meningkatkan kepuasan publik, dan meningkatkan performa tim untuk mencapai target yang diinginkan. Untuk penelitian selanjutnya, disarankan agar pendekatan ensemble learning diterapkan untuk meningkatkan akurasi prediksi sentimen. Selain itu, penggunaan '*Human Annotation*' dapat membantu meningkatkan akurasi karena Text Blob memiliki keterbatasan dalam mengenali sentimen yang kompleks. Dengan demikian, hasil penelitian dapat menjadi lebih akurat dan representatif. Penelitian lanjutan dapat memanfaatkan algoritma deep learning untuk meningkatkan akurasi analisis sentimen terhadap kebijakan pemerintah, PSSI, dan Timnas Indonesia. Dengan demikian, diharapkan pengambilan keputusan strategis oleh PSSI dan pemerintah dapat dilakukan untuk meningkatkan performa sepakbola Timnas Indonesia.

REFERENCES

- [1] R. R. Pratama and R. R. Suryono, "Performance Comparison Of Naive Bayes, Support Vector Machine And Random Forest Algorithms For Apple Vision Pro Sentiment Analysis," *J. Tek. Inform. Vol.*, vol. 6, no. 1, pp. 31–39, 2025, doi: <https://doi.org/10.52436/1.jutif.2025.6.1.4035>.
- [2] T. A. Anastasya, A. Diani, P. Saka, M. Juventus, D. Deke, and A. M. Rizki, "Optimasi Algoritma Svm Dengan Pso Untuk Analisis Sentimen Pada Media Sosial X Terhadap Kinerja Timnas Di Era Shin Tae-Yong," (*Jurnal Mhs. Tek. Inform.*), vol. 9, no. 1, pp. 384–391, 2025, doi: <https://doi.org/10.36040/jati.v9i1.12298>.
- [3] H. F. Wibowo and D. Nugroho, "Analisis Momen Pola Menyerang, Bertahan dan Transisi pada Timnas Indonesia U-19 Vs Bulgaria," *J. Sport Sci. Educ. /*, vol. 6, no. 1, p. 57, 2021, [Online]. Available: <http://journal.unesa.ac.id/index.php/jossae/indexhttps://doi.org/10.26740/jossae.v6n2>
- [4] S. Imaddudin, I. Astuti, and S. Ruhama, "Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan Algoritma Support Vector Machine (SVM) pada Media Sosial X," *J. Penelit. MULTIDISIPLIN BANGSA*, vol. 1, no. 8, pp. 1003–1013, 2025, doi: <https://doi.org/10.59837/jpnmb.v1i8.193>.
- [5] A. C. M. Pratama and A. T. Laksana, "Lahirnya kembali PSSI (Persatuan Sepak Bola Seluruh Indonesia) Tahun 1950-1954," *AVATARA, e-Journal Pendidik. Sej.*, vol. 10, no. 1, pp. 1–9, 2020
- [6] S. K. Fadhillah and Y. S. Isma, "Analisis Komentar Netizen di Instagram pada Pemberitaan Pemecatan Shin Tae-Yong Universitas Malikussaleh , Indonesia Universitas Malikussaleh , Indonesia Analisis Komentar Netizen di Instagram pada Pemberitaan Pemecatan Shin Tae-Yong PSSI untuk melatih Ti," *TUTURAN J. Ilmu Komunikasi, Sos. dan Hum.*, vol. 3, no. 1, p. 205, 2025, doi: <https://doi.org/10.47861/tuturan.v3i1.1566>.
- [7] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: [10.52985/insyst.v1i1.36](https://doi.org/10.52985/insyst.v1i1.36).
- [8] D. R. P. Jaya and S. Lestari, "Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma Naïve Bayes," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3262–3277, 2024, [Online].



- Available: <https://journal.stmiki.ac.id/index.php/jimik/>
- [9] F. S. Pamungkas and I. Kharisudin, “Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” *Pros. Semin. Nas. Mat.*, vol. 4, no. 628–634, pp. 1–7, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/45038>
 - [10] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
 - [11] A. Baita, Y. Pristyanto, and N. Cahyono, “Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (Svm) Dan K-Nearest Neighbor (Knn),” *Inf. Syst. J.*, vol. 4, no. 2, pp. 42–42, 2021, doi: <https://doi.org/10.24076/infosjournal.2021v4i2.687>.
 - [12] T. D. Putra and D. Oktafiani, “Klasifikasi Sentimen Postingan Sosial Media Menggunakan Machine Learning Random Forest dan Naïve Bayes,” *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 2338–2347, 2025, doi: <https://doi.org/10.31004/innovative.v5i1.17935>.
 - [13] J. W. Iskandar and Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: [10.29207/resti.v5i6.3588](https://doi.org/10.29207/resti.v5i6.3588).
 - [14] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: [10.25126/jtiik.0813944](https://doi.org/10.25126/jtiik.0813944).
 - [15] M. Rafli Ghufuron *et al.*, “<https://creativecommons.org/licenses/by-sa/4.0/> 307 J-INTECH (Journal of Information and Technology) Analisis Sentimen Pengguna Twitter Terhadap Pemilu 2024 Berbasis Model XLM-T,” *Inf. adn Technol.*, vol. 11 no 2, no. 204, pp. 307–315, 2023, doi: <https://doi.org/10.32664/j-intech.v11i2.1013>.
 - [16] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, “Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Analyzing Twitter Sentiment Regarding Public Opinion on the Removal of Honorary Employees,” *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, 2022.
 - [17] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm),” *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: [10.14710/j.gauss.v9i3.28932](https://doi.org/10.14710/j.gauss.v9i3.28932).
 - [18] T. Gori, A. Sunyoto, and H. Al Fatta, “Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: [10.25126/jtiik.20241118074](https://doi.org/10.25126/jtiik.20241118074).
 - [19] L. A. Andika, P. A. N. Azizah, and R. Respatiwalan, “Analisis Sentimen Masyarakat Terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: [10.13057/ijas.v2i1.29998](https://doi.org/10.13057/ijas.v2i1.29998).
 - [20] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: [10.47080/simika.v5i1.1411](https://doi.org/10.47080/simika.v5i1.1411).
 - [21] W. G. S. Parwita, “Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 14, no. 1, p. 27, 2019, doi: [10.30872/jim.v14i1.1272](https://doi.org/10.30872/jim.v14i1.1272).
 - [22] M. Syarifuddin, “Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: [10.33480/inti.v15i1.1433](https://doi.org/10.33480/inti.v15i1.1433).
 - [23] P. I. Purnama and R. R. Suryono, “Analisis Sentimen Acara Clash of Champions dengan Algoritma Naïve Bayes dan Support Vector Machine,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, pp. 2277–2287, 2025, doi: [10.47065/bits.v6i4.6575](https://doi.org/10.47065/bits.v6i4.6575).
 - [24] L. Apriani and N. Hendrastuty, “Analisis Sentiment Terhadap Diabetes Menggunakan Algoritma Naïve Bayes , Random Forest , SVM Pada Media Sosial X,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, pp. 2469–2479, 2025, doi: [10.47065/bits.v6i4.6941](https://doi.org/10.47065/bits.v6i4.6941).
 - [25] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: [10.33365/jtk.v15i1.744](https://doi.org/10.33365/jtk.v15i1.744).
 - [26] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: [10.31311/ji.v6i2.5129](https://doi.org/10.31311/ji.v6i2.5129).
 - [27] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: [10.26877/jiu.v7i1.7099](https://doi.org/10.26877/jiu.v7i1.7099).
 - [28] P. G. Dewata, A. Rizky, and H. Irsyad, “Analisis Sentimen Terhadap Boikot Produk Israel Menggunakan Algoritma Naive Bayes Dan SMOTE,” *J. Rein (Rekayasa Inform.)*, vol. 1, no. 1, pp. 7–15, 2024, [Online]. Available: <https://rekayasainformatika.com/index.php/JREIN/article/view/1>