

Analisis Sentimen Tweet Netizen Terhadap Timnas Sepak Bola Indonesia di Era Shin Tae-Yong Menggunakan Algoritma Support Vector Machine (SVM)

Muhammad Rifqi Fauzan^a, Siska Anraeni^b, Lutfi Budi Imawan^c

Universitas Muslim Indonesia, Makassar, Indonesia

^a13020190207@umi.ac.id; ^bsiska.anraeni@umi.ac.id; ^clutfibudi.imawan@umi.ac.id

Received: xx xx xxxx | Revised: xx xx xxxx | Accepted: xx xx xxxx | Published: xx xx xxxx

Abstrak

Era kepelatihan Shin Tae-Yong menjadi momen penting bagi sepak bola Indonesia. Performa Timnas Indonesia di bawah asuhan Shin Tae-Yong menuai beragam respons di media sosial X. Penelitian ini bertujuan untuk menganalisis sentimen netizen terhadap Timnas Indonesia di era Shin Tae-Yong dengan menggunakan algoritma *Support Vector Machine (SVM)*. Dataset yang digunakan dalam penelitian ini terdiri dari 398 tweet yang dikumpulkan dari platform media sosial X, namun setelah prosedur preprocessing, jumlah tersebut berkurang menjadi 391 tweet yang merepresentasikan opini dan pandangan netizen tentang Timnas Indonesia di era Shin Tae-Yong. Dari data yang dianalisis, 200 data (51,15%) menunjukkan sentimen positif terhadap Timnas Indonesia di era Shin Tae-Yong, sementara 191 data (48,85%) menunjukkan sentimen negatif. Proses pengumpulan data melibatkan pemilihan tweet yang relevan dengan menggunakan keyword terkait Timnas Indonesia di era Shin Tae-Yong, diikuti dengan tahap *preprocessing* untuk membersihkan data. Evaluasi dilakukan dengan mengukur akurasi, presisi, *recall*, dan *F1-score* setiap algoritma dalam mengklasifikasikan sentimen tweet sebagai positif atau negatif. Penerapan metode *SVM* dalam analisa sentimen tentang Timnas Indonesia di era Shin Tae-Yong berhasil mencapai tingkat *accuracy* sebesar 63,29%, *precision* sebesar 65%, *recall* sebesar 65% dan *F1-Score* sebesar 63%. Hasil penelitian menunjukkan bahwa algoritma *SVM* secara signifikan relatif baik dalam mengenali dan mengklasifikasikan sentimen terkait Timnas Indonesia di era Shin Tae-Yong. Penelitian ini dapat memberikan wawasan penting mengenai pemilihan algoritma machine learning yang optimal untuk analisis sentimen dan dapat membantu peneliti dalam memilih metode yang tepat untuk analisis sentimen di berbagai topik.

Kata kunci: *Support Vector Machine (SVM), Analisis Sentimen, X, Shin Tae-Yong*

Pendahuluan

Salah satu olahraga terpopuler di seluruh dunia adalah sepak bola. Di Indonesia, sepak bola sangat disukai dan memiliki banyak penggemar [1]. Menurut survei yang dilakukan oleh Ipsos (2022), Indonesia memiliki 69% penggemar sepak bola terbesar di dunia [2]. Sepak bola tidak hanya sekedar hiburan, tetapi juga sudah menjadi budaya pembentuk jati diri bangsa. Popularitas tersebut terlihat di berbagai ajang internasional seperti Piala AFF yang menarik perhatian jutaan penggemarnya setiap tahunnya [3].

Era kepelatihan Shin Tae-Yong menjadi momen penting bagi sepak bola Indonesia. Performa Timnas Indonesia di bawah asuhan Shin Tae-Yong menuai beragam respons di media sosial X [4]. Berbagai perubahan signifikan dilakukan mulai dari strategi permainan hingga manajemen tim yang membuahkan peningkatan performa, termasuk mengamankan posisi kedua Piala AFF 2020 [5]. Namun, tidak semua keputusan pelatih itu mendapat sambutan positif [3]. Strategi dan pemilihan pemain masih menjadi sasaran kritik di media sosial. Saat Timnas meraih hasil positif, banyak komentar bermunculan yang menggambarkan kebahagiaan dan pujian. Sebaliknya jika hasil yang diperoleh buruk, maka akan muncul komentar-komentar yang berisi kritikan kasar bahkan ejekan [6].

Perkembangan media baru yang pesat telah menyebabkan perubahan besar dalam media konvensional, termasuk peralihan dari radio, koran, majalah dan televisi ke format online [7]. Di era digital, media sosial telah menjadi platform penting bagi orang untuk bertukar pikiran, opini, dan emosi mereka tentang berbagai topik, termasuk olahraga [8]. Menurut laporan statistik *We Are Social 2024*, pengguna X di Indonesia sebanyak 57,5% dari jumlah populasi, menjadikannya salah satu platform yang paling populer untuk menyampaikan pendapat. Twitter atau yang sekarang dikenal sebagai X menjadi platform yang efektif bagi netizen untuk menyampaikan pendapat dan perasaan mereka terhadap Timnas. Salah satu fitur X yang paling banyak digunakan adalah fitur tweet, yang memungkinkan pengguna menuliskan pemikiran, opini, dan pandangan mereka [9].

Analisis sentimen menjadi salah satu teknik yang banyak digunakan untuk memahami opini dan perasaan seseorang berdasarkan pada data teks [10]. Analisis sentimen dilakukan untuk melihat perspektif atau cara seseorang berpendapat tentang suatu isu atau objek, apakah mereka cenderung memiliki pandangan atau pendapat positif atau negatif [11]. Positif berarti ada aspek apresiasi, persetujuan, dan sebagainya mengenai suatu topik. Sebaliknya, negatif berarti tidak suka, tidak setuju, kecewa, dan sebagainya, sedangkan netral berarti berada di antaranya [12]. Pada penelitian yang dilakukan oleh Pramudya Nugraha, Rasiban, Frensis Matheos Sarimole, dan Tundo tentang analisis sentimen kepuasan publik terhadap masa kepemimpinan Shin Tae-Yong, sentimen pada penelitian tersebut menghasilkan sentimen positif terhadap Shin Tae-Yong [13].

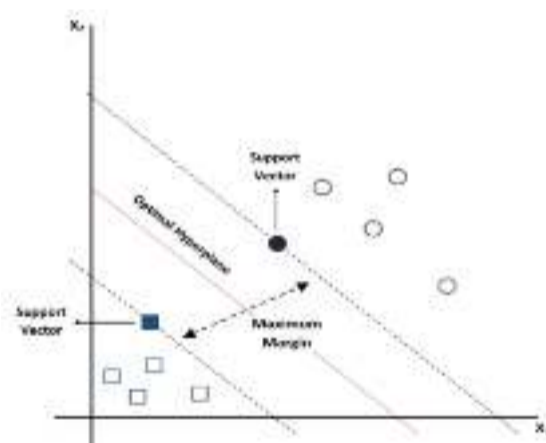
Untuk memahami sentimen publik dan dampaknya terhadap kinerja timnas sepak bola Indonesia di era kepelatihan Shin Tae-Yong, diperlukan metode analisis yang efektif, salah satunya adalah *Support Vector Machine (SVM)*. Dengan menggunakan *SVM*, masalah klasifikasi dapat diselesaikan dengan mengidentifikasi fungsi pemisah margin maksimum (*hyperplane*) yang dapat memisahkan dua kelas secara optimal [4]. *SVM* berfungsi untuk menemukan proses atau fungsi pemisah (*hyperplane*) yang memaksimalkan jarak antar kelas di setiap kategori sentimen. Algoritma *SVM* terbukti memiliki tingkat akurasi yang lebih tinggi dalam klasifikasi teks, dibandingkan dengan algoritma lain seperti *K-Nearest Neighbors (KNN)*, *C4.5*, dan *Naïve Bayes* [14].

Penelitian sebelumnya yang dilakukan oleh Galang Sumantri dan Bambang Sumarno Hadi Marwoto, telah menerapkan algoritma *Support Vector Machine* untuk analisis sentimen di twitter terkait tim nasional sepak bola Indonesia menggunakan metode *Support Vector Machine*. Data yang digunakan berjumlah 1.621 data bersih, yang mewakili 1.073 data sentimen positif dan 548 data negatif, sehingga dapat disajikan 66% data positif dan 34% data negatif. Berdasarkan pengujian yang dilakukan, performa akurasi sistem analisis sentimen terkait timnas sepak bola Indonesia dengan pendekatan *machine learning* dengan metode *Support Vector Machine* menghasilkan nilai *accuracy* 84%, *precision* 82%, *recall* 81% dan *F1 score* sebesar 81% [1].

Berdasarkan penjelasan beberapa penelitian sebelumnya, maka proses analisis klasifikasi sentimen juga dapat dilakukan terhadap opini publik di media sosial X. Proses analisis dilakukan dengan metode *SVM* menggunakan konsep *Natural Language Processing (NLP)* untuk menentukan pola sentimen positif dan negatif pada dataset penelitian. Penelitian ini bertujuan untuk melihat apa yang terjadi di tengah proses masyarakat mengenai timnas Indonesia era Shin Tae-Yong, lalu bagaimana tren opini publik terhadap timnas Indonesia yang tercermin di media sosial X. Kontribusi yang disajikan dalam penelitian ini dapat memberikan wawasan baru tentang klasifikasi menggunakan metode *Support Vector Machine* dan melihat bagaimana topik timnas Indonesia berkembang di media sosial X di Indonesia.

Metode

Support Vector Machine (SVM) adalah teknik pembelajaran yang menggunakan fungsi *linear* dan *non-linear* untuk membagi data menjadi dua atau lebih kategori [15]. Tujuan metode ini adalah menemukan *hyperplane* atau garis potong terbaik yang memiliki margin optimal. Konsep *hyperplane* disajikan pada gambar 1.



Gambar 1. Konsep *hyperplane* pada *SVM* [16]

Margin dalam *SVM* adalah jarak antara *hyperplane* dengan titik terdekat pada setiap kelas [17]. *Support Vector* juga dapat dianggap sebagai objek-objek data terluar yang paling dekat dengan *hyperplane*. *Support Vector* inilah yang dihitung oleh *SVM*. Pada dasarnya, *SVM* merupakan suatu *linear classifier*, namun *SVM* dapat diperluas ke *non-linear classifier*. Konsep *kernel trick* digunakan dalam ruang dimensi yang lebih tinggi untuk menangani *non-linear classifier*. Beberapa fungsi *kernel trick* pada *SVM* yang sering digunakan, yaitu:

1. *Kernel Linear*

Kernel linear sering digunakan untuk data dua kelas atau data yang diklasifikasikan secara *linear*.

$$K(x_i, x_j) = x_i \cdot x_j \quad (1)$$

2. *Kernel Radial Bias Function (RBF)*

Kernel RBF atau *Gaussian* digunakan untuk menyelesaikan masalah data yang tidak terpisah secara *linear* atau data dengan banyak kelas.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

3. *Kernel Polynomial*

Kernel polynomial merupakan *kernel trick* lainnya yang digunakan untuk menyelesaikan masalah data yang tidak terpisah secara *linear*.

$$K_{x_i x_j} = (x_i \cdot x_j)^d \quad (3)$$

Keterangan:

K = Fungsi kernel yang menghitung kemiripan antara dua buah vektor fitur x_i dan x_j dalam dataset.

x_i = Vektor fitur data point i dalam dataset.

x_j = Vektor fitur data point j dalam dataset.

d = Parameter kernel yang menentukan derajat polynomial transformasi fitur yang akan digunakan.

γ = Parameter kernel yang menentukan seberapa cepat nilai kernel menurun seiring dengan jarak antara kedua vektor fitur.

Diagram tahapan penelitian yang dibuat adalah rencana yang akan dilakukan oleh peneliti dalam melakukan analisis sentimen. Pada diagram tahapan penelitian, terdapat rincian dalam proses pencarian awal data sampai dengan akhir. Setiap proses yang ditulis pada diagram tahapan akan dijelaskan satu persatu guna memperjelas diagram yang diolah tersebut. Alur penelitian secara menyeluruh dapat dilihat pada visualisasi gambar 2 berikut ini:



Gambar 2. Tahapan Penelitian

1. Identifikasi Masalah

Tahap pertama dalam penelitian ini adalah identifikasi masalah. Peneliti menganalisis opini netizen terhadap Timnas sepak bola Indonesia di era kepelatihan Shin Tae-Yong pada sosial media X.

2. Studi Literatur

Tahapan selanjutnya adalah studi literatur. Tahapan ini dilakukan untuk memahami penelitian sejenis yang telah dilakukan sebelumnya guna memperoleh teori dan informasi yang mendukung penelitian literatur yang diteliti, khususnya mengenai klasifikasi risiko rendah, risiko sedang, dan risiko tinggi.

3. Pengumpulan Data

Pengumpulan data dilakukan menggunakan kaggle untuk memperoleh data yang diperlukan, yaitu data mengenai Timnas sepak bola di era Shin Tae-Yong.

4. Analisis Data

Analisis data dilakukan dengan menggunakan algoritma *Support Vector Machine (SVM)*.

5. Kesimpulan

Tahap ini merupakan tahapan yang dimana penulis mengambil kesimpulan dari hasil menganalisis sentimen tweet netizen terhadap Timnas sepak bola Indonesia di era Shin Tae-Yong dengan menggunakan algoritma *Support Vector Machine (SVM)*.

Perancangan

Penelitian ini dilakukan secara bertahap, dimulai dengan pengumpulan data, preprocessing, transformation, text mining dan evaluation untuk menganalisis sentimen tweet netizen terhadap Timnas sepak bola Indonesia di era Shin Tae-Yong. Gambar 3 menunjukkan tahapan metode tersebut.



Gambar 3. Tahapan Metode

A. Pengumpulan Data

Data untuk penelitian ini diambil dari sumber kaggle.com [18]. Kumpulan data ini terdiri dari 398 data, dengan dua atribut utama, yaitu label dan komentar. Atribut komentar berisi teks komentar

yang ditulis oleh pengguna terhadap Timnas Indonesia di era Shin Tae-Yong, sedangkan atribut label menunjukkan klasifikasi sentimen komentar tersebut, yakni positif atau negatif. Seluruh proses dilakukan dengan menggunakan bahasa pemrograman *Python* dan *Google Colab* sebagai *interface notebook*.

B. Preprocessing

Preprocessing merupakan langkah penting dalam proses analisis sentimen yang bertujuan untuk mengatasi data yang kurang sempurna, memperbaiki data yang tidak konsisten, dan mengubah data mentah menjadi data bersih yang siap untuk dianalisis [19]. Data mentah yang terkumpul pada tahap preprocessing akan diolah dan diproses untuk menghasilkan data yang siap digunakan pada langkah berikutnya [20]. Tahapan-tahapan dalam preprocessing data adalah sebagai berikut:

1. Cleansing

Cleansing merupakan tahapan yang berfungsi untuk menghilangkan atau menghapus data dari kalimat yang mengandung simbol atau elemen kata yang tidak diperlukan, misalnya nama pengguna (@username), hashtag (#), emoji, URL, email, dan kata kunci [16]. Selain itu, pada fase ini juga dilakukan pembersihan data duplikat dan nilai yang kosong [21]. Proses cleansing ini penting untuk memastikan bahwa hanya data relevan dan bermakna yang disimpan, sehingga tidak mengganggu analisis lebih lanjut. Pembersihan data dilakukan untuk menghilangkan gangguan yang dapat memengaruhi kinerja model dalam klasifikasi sentimen.

Tabel 1. Proses *Cleansing*

SEBELUM CLEANSING	SETELAH CLEANSING
<username> gue curiga ne anak biar dicoret masuk list senior ma STY jadi masih peluang masuk Timnas Belanda..ceeh malah dibawa juga di Piala Asia ma STY	gue curiga ne anak biar dicoret masuk list senior ma STY jadi masih peluang masuk Timnas Belandaceeh malah dibawa juga di Piala Asia ma STY
A untuk timnas! Statistik Thom Haye di Eropa sungguh mengesankan, pantas jadi andalan kita. Semoga proses naturalisasi berjalan lancar dan Haye bisa memberikan kontribusi besar untuk kebangkitan sepak bola Indonesia!	A untuk timnas Statistik Thom Haye di Eropa sungguh mengesankan pantas jadi andalan kita Semoga proses naturalisasi berjalan lancar dan Haye bisa memberikan kontribusi besar untuk kebangkitan sepak bola Indonesia

2. Case Folding

Case folding adalah proses mengubah semua huruf kapital pada dataset menjadi huruf kecil, untuk menormalkan karakter pada dataset tersebut sehingga proses tokenizing dapat dilakukan dengan lebih mudah [22].

Tabel 2. Proses *Case Folding*

SEBELUM CASE FOLDING	SETELAH CASE FOLDING
WKWWKK LEBIH MILIH DENDY DIBANDIN ADAM ALIS pasrah aja kadang emang sty	wkwwkk lebih milih dendy dibandin adam alis pasrah aja kadang emang sty
STY OUT JIKA TAHUN DEPAN KALAH DARI TIMOR TIMUR	sty out jika tahun depan kalah dari timor timur

3. Normalization

Normalization merupakan proses pembetulan kesalahan kata, seperti salah eja atau penggunaan kata-kata slang, sehingga kata tersebut kembali menjadi kata sebenarnya menurut EYD. Pada tahap ini, koreksi kata akan dilakukan menggunakan kamus slag [23] yang di dalamnya berisi data kata baku dan tidak baku yang telah disiapkan sebelumnya [19].

Tabel 3. Proses *Normalization*

SEBELUM NORMALIZATION	SETELAH NORMALIZATION
yang saya heran dari sty ini sudah ada saran dari expert tapi kok ttp ga didengar ya	yang saya heran dari sty ini sudah ada saran dari expert tapi kok tetap tidak didengar ya
kalo gua jadi sty gua coret edo panggil lilipaly atau si fajardi posisi bek kiri ampe nama gitu ada shaynearhan sama edo	kalau aku jadi sty aku coret edo panggil lilipaly atau sih fajardi posisi bek kiri sampai nama begitu ada shaynearhan sama edo

4. Stemming

Stemming bertujuan untuk menghilangkan kata imbuhan. Proses ini menggunakan library Sastrawi. Sastrawi digunakan untuk mengubah kata berimbuhan bahasa Indonesia menjadi bentuk dasarnya [24]. Stemming digunakan untuk mengurangi variasi kata-kata yang berasal dari akar kata yang sama, misalnya "kualitasnya" akan diganti dengan akar kata "kualitas". Dengan demikian, model dapat lebih fokus pada makna inti kata tanpa terpengaruh oleh variasi morfologi yang tidak relevan.

Tabel 4. Proses *Stemming*

SEBELUM STEMMING	SETELAH STEMMING
iya neneknya orang sulawesi itu sudah memenuhi syarat fifa kok untuk di naturalisasi half blood sampai pihak kakeknenek	iya nenek orang sulawesi itu sudah penuh syarat fifa kok untuk di naturalisasi half blood sampai pihak kakeknenek
cuma berharap sty di piala asia nanti bisa lebih maksimal	cuma harap sty di piala asia nanti bisa lebih maksimal

5. Stopword Removal

Stopword removal bertujuan untuk mengurangi beban pemrosesan data, karena kata-kata ini tidak berkontribusi secara signifikan terhadap makna atau sentimen yang akan dianalisis. Tahap stopwords removal melibatkan penghapusan kata-kata dari daftar stopwords NLTK, seperti konjungsi "dan", "atau", "yang", dll. Selain itu, daftar kata lainnya dibuat secara manual untuk mencakup kata-kata termasuk kata benda yang tidak relevan dengan penelitian ini dan tidak ada dalam daftar stopwords NLTK [24].

Tabel 5. Proses *Stopword Removal*

SEBELUM STOPWORD REMOVAL	SETELAH STOPWORD REMOVAL
kemarin adalah kamufase sty supaya kuat benar tidak tampak tunggu saat jepang akan	kemarin kamufase sty kuat tunggu jepang
banyak yang denial abang kalau sty itu jauh lebih buruk banding luis milla	denial abang sty buruk banding luis milla

6. Tokenization

Tokenization merupakan proses memecah teks menjadi kalimat atau paragraf menjadi bagian-bagian yang disebut token, yang kemudian akan dianalisis [25].

Tabel 6. Proses *Tokenization*

SEBELUM TOKENIZATION	SETELAH TOKENIZATION
lord dendy jaman sty auto main bro	['lord', 'dendy', 'jaman', 'sty', 'auto', 'main', 'bro']

lihat sty ya sosok dendy mas	['lihat', 'sty', 'ya', 'sosok', 'dendy', 'mas']
------------------------------	---

C. *Bag of Words*

Bag of words adalah pola proses dalam *Natural Language Processing (NLP)* dan sebagian besar digunakan untuk mendapatkan nilai kata yang telah diperlakukan dalam model *machine learning* [26]. *Bag of words* menawarkan kesederhanaan dan fleksibilitas, memungkinkan teks sebagai fungsi model kata dalam dokumen, yang berkinerja untuk aplikasi analisis teks [27]. Cara kerja *bag of words* yaitu mempelajari satu kata kemudian menafsirkan setiap kata pada sebuah dokumen dengan menghitung jumlah munculnya tiap kata dari setiap dokumen [26].

D. Klasifikasi Metode dan Pengujian Data

Setelah melalui tahap preprocessing, tweet akan diekstraksi fiturnya untuk kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine (SVM)* dalam penelitian ini. Tujuan dari proses klasifikasi ini adalah untuk mengidentifikasi sentimen yang terkandung dalam kumpulan tweet yang telah dikumpulkan [28]. Pada tahap pengujian, data tweet akan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, guna melatih dan menguji kinerja model secara efektif. Evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score yang dihitung berdasarkan *confusion matrix* untuk mengukur keakuratan dan efektivitas klasifikasi yang dilakukan.

E. *Evaluation*

Tahapan terakhir dalam proses ini adalah mengidentifikasi pola atau informasi yang diperoleh dari proses data mining untuk menilai apakah pola atau informasi tersebut konsisten dengan hipotesis atau fakta yang ada [29]. Proses ini tidak hanya memastikan validitas data yang ditemukan, namun juga memberikan peluang untuk menyempurnakan atau memperbaiki hipotesis jika hasilnya menunjukkan adanya penemuan baru. Pada tahap ini, kinerja algoritma *SVM* akan diukur menggunakan *confusion matrix*. *Confusion matrix* adalah metode prediksi untuk menghitung akurasi matriks dalam data mining. Dengan menggunakan metode ini, dimungkinkan untuk melihat efektivitas model dalam klasifikasi sentimen.

		PREDICTED VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP (True Positive)	FN (False Negative)
	NEGATIVE (0)	FP (False Positive)	TN (True Negative)

Gambar 4. *Confusion Matrix*

Keterangan:

1. *True Positive (TP)*, merupakan jumlah data yang dengan benar diklasifikasikan sebagai kelas positif.
2. *True Negative (TN)*, merupakan jumlah data yang benar diklasifikasikan sebagai kelas negatif.
3. *False Positive (FP)*, merupakan jumlah data yang seharusnya adalah kelas negatif, namun salah diklasifikasikan sebagai kelas positif.
4. *False negative (FN)*, merupakan jumlah data yang seharusnya adalah kelas positif, namun salah diklasifikasikan sebagai kelas negatif.

Pengujian dengan menggunakan confusion matrix menghasilkan perhitungan yang menghasilkan empat keluaran, yaitu:

1. Akurasi, digunakan untuk mengukur ketepatan suatu model dalam melakukan klasifikasi data secara keseluruhan, berikut rumus menghitung nilai akurasi.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (6)$$

2. Presisi, digunakan untuk mengukur banyaknya prediksi positif yang benar terhadap total prediksi positif, serta meminimalkan false positive. berikut rumus menghitung nilai presisi.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

3. Recall, digunakan untuk mengukur banyaknya prediksi positif yang benar terhadap total data aktual yang positif, serta meminimalkan false negative, berikut rumus menghitung nilai recall.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

4. F1 Score, adalah harmonic mean dari presisi dan recall yang berguna untuk mengevaluasi model dengan presisi dan recall yang tinggi, berikut rumus menghitung nilai F1 score.

$$F1\ Score = 2 \frac{(precision \times recall)}{precision + recall} \quad (9)$$

Keterangan:

TP: Jumlah prediksi *true positive*

TN: Jumlah prediksi *true negative*

FP: Jumlah prediksi *false positive*

FN: Jumlah prediksi *false negative*

Pemodelan

A. Wordcloud

Wordcloud digunakan untuk menampilkan gambar visual dari teks tertulis dan dapat digunakan sebagai alat untuk membantu menganalisis teks. Kata-kata yang ada ditampilkan dalam tampilan yang ukuran fontnya bergantung pada frekuensi kemunculan kata tersebut dalam teks. Singkatnya, kata yang paling besar adalah kata yang paling sering muncul dalam data.

1. *Wordcloud* sentimen positif

Wordcloud untuk sentimen positif menampilkan kata-kata yang paling banyak digunakan dalam komentar yang mendukung timnas sepak bola Indonesia di era Shin Tae-Yong. Kata-kata seperti "main", "timnas", "sty" (singkatan dari Shin Tae-Yong), dan "naturalisasi" menunjukkan bahwa komentar-komentar tersebut sangat berfokus pada gaya bermain tim dan pelatih timnas Indonesia, serta mendukung kebijakan naturalisasi pemain asing dalam skuad timnas Indonesia. Komentar – komentar positif tersebut mencerminkan harapan dan keyakinan banyak penggemar terhadap masa depan timnas Indonesia, meski masih terdapat kekurangan dan tantangan.

Gambar 5. *Wordcloud* sentimen positif

2. *Wordcloud* sentimen negatif

Wordcloud untuk sentimen negatif menggambarkan ketidakpuasan publik terhadap tim nasional sepak bola Indonesia dan Shin Tae-Yong. Kata-kata seperti "main", "timnas", "sty" (singkatan dari Shin Tae-Yong), dan "naturalisasi" menunjukkan kritik terhadap keputusan yang diambil pelatih dan performa timnas Indonesia, serta kritik terhadap kebijakan naturalisasi pemain. Kata "dendey" dan "kalah" menunjukkan kekecewaan terhadap kinerja pemain tertentu dan keputusan yang dianggap tidak efektif. Komentar negatif ini mencerminkan kekecewaan penggemar terhadap kinerja tim dan kebijakan yang diambil di bawah kepelatihan Shin Tae-Yong.

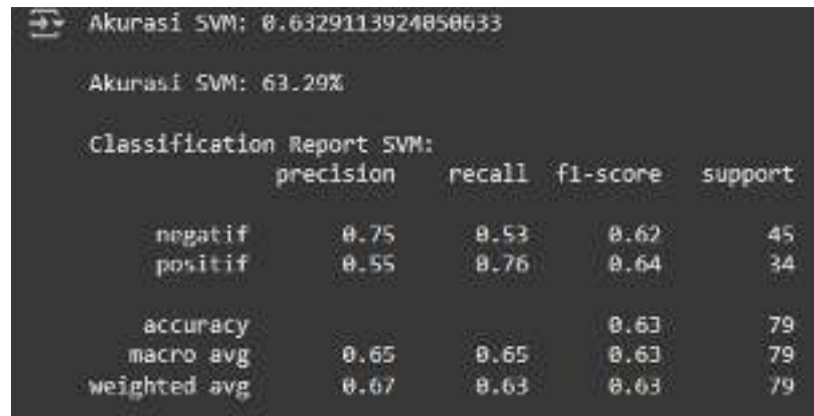
Gambar 6. *Wordcloud* sentimen negatif

B. Pembagian Data

Dataset yang telah didapatkan dan telah melalui proses *preprocessing* serta labeling, selanjutnya dataset dibagi menjadi 20% data *testing* dan 80% data *training*.

C. Evaluasi Model

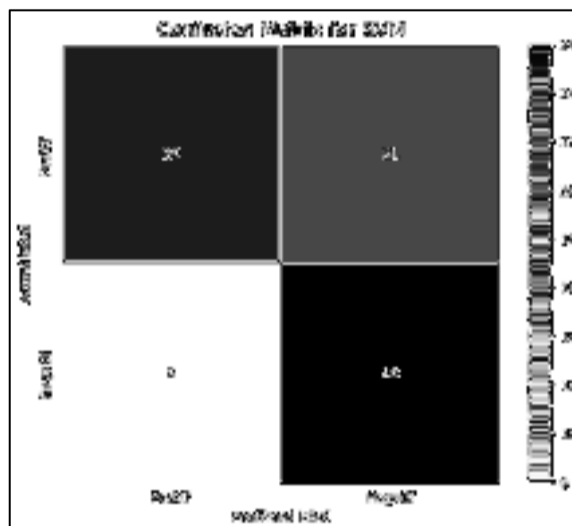
Pada tahap ini dilakukan pengujian pada *Support Vector Machine* dengan pembagian 80% data *training* dan 20% data *testing*. Tahapan pembagian data dilakukan untuk mengevaluasi performa model dari *SVM*.



Gambar 7. Classification Report SVM

Pada gambar 7, mengindikasikan bahwa nilai akurasi SVM dengan pembagian 80% data *training* dan 20% data *testing* adalah 63,29%. Pada kelas “negatif” memiliki precision 0.75, recall 0.53, dan F1-score 0.62. Sedangkan kelas positif memiliki precision 0.55, recall 0.76, dan F1-score 0.64.

Setelah memperoleh hasil klasifikasi sentimen dengan SVM, pengujian harus dilakukan dengan *confusion matrix* untuk menentukan informasi yang diharapkan dan hasil aktual. Perhitungan dari *confusion matrix* adalah *accuracy*, *recall*, *precision* dan *F1-score*. Berikut merupakan hasil pengujian menggunakan *Confusion Matrix*:



Gambar 8. Pengujian Confusion Matrix for SVM

Dari gambar 8, dapat dijelaskan bahwa memiliki 24 data *true positive*, 26 data *true negative*, 8 data *false positive*, dan 21 data *false negative*. Perhitungan manual *accuracy*, *recall*, *precision* dan *F1-score* dari *confusion matrix* diatas sebagai berikut:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} = \frac{24 + 26}{(24 + 26 + 8 + 21)} = 0,6329 = 63,29 \%$$

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

$$Label\ negatif = \frac{TP}{TP+FN} = \frac{24}{24+21} = 0,53$$

$$Label\ positif = \frac{TN}{TN+FP} = \frac{26}{26+8} = 0,76$$

$$Average\ Recall = \frac{label\ negatif+label\ positif}{jumlah\ label} = \frac{0,53+0,76}{2} = 0,65 = 65\%$$

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

$$\text{Label negatif} = \frac{TP}{TP+FP} = \frac{24}{24+8} = 0,75$$

$$\text{Label positif} = \frac{TN}{TN+FN} = \frac{26}{26+21} = 0,55$$

$$\text{Average Precision} = \frac{\text{label negatif} + \text{label positif}}{\text{jumlah label}} = \frac{0,75+0,55}{2} = 0,65 = 65\%$$

$$F1 \text{ Score} = 2 \frac{(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$$

$$\text{Label negatif} = 2 \frac{\text{label negatif precision} \times \text{label negatif recall}}{\text{label negatif precision} + \text{label negatif recall}} = 2 \frac{0,75 \times 0,53}{0,75 + 0,53} = 0,62$$

$$\text{Label positif} = 2 \frac{\text{label positif precision} \times \text{label positif recall}}{\text{label positif precision} + \text{label positif recall}} = 2 \frac{0,55 \times 0,76}{0,55 + 0,76} = 0,64$$

$$\text{Average F1 Score} = \frac{\text{label negatif} + \text{label positif}}{\text{jumlah label}} = \frac{0,62+0,64}{2} = 0,63 = 63\%$$

Kesimpulan

Berdasarkan penelitian yang menggunakan algoritma *Support Vector Machine* dalam analisis sentimen terhadap Timnas Indonesia di era Shin Tae-Yong, dapat disimpulkan bahwa jumlah data awal adalah 398 tweet. Namun setelah prosedur preprocessing, jumlah tersebut berkurang menjadi 391 tweet. Dari data yang dianalisis, 200 data (51,15%) menunjukkan sentimen positif terhadap Timnas Indonesia di era Shin Tae-Yong, sementara 191 data (48,85%) menunjukkan sentimen negatif. Mayoritas netizen Indonesia tampaknya memiliki pandangan yang mendukung terhadap perkembangan Timnas Indonesia di era Shin Tae-Yong. Penerapan metode SVM dalam analisa sentimen tentang Timnas Indonesia di era Shin Tae-Yong berhasil mencapai tingkat *accuracy* sebesar 63,29%, *precision* sebesar 65%, *recall* sebesar 65% dan *F1-Score* sebesar 63%. Secara keseluruhan, hasil ini mengindikasikan bahwa model klasifikasi analisis sentimen menggunakan metode SVM memiliki performa yang relatif baik dalam mengenali dan mengklasifikasikan sentimen terkait Timnas Indonesia di era Shin Tae-Yong.

Daftar Pustaka

- [1] G. Sumantri and S. H. Marwoto, "Analisis Sentimen Di Twitter Terkait Tim Nasional Sepak Bola Indonesia Menggunakan Metode Support Vector Machine," *J. Kaji. dan Terap. Mat.*, vol. 10, no. 2, pp. 96–104, 2024, [Online]. Available: <http://journal.student.uny.ac.id/ojs/index.php/jktm>:
- [2] R. Grimm and C. Jackson, "Attitudes Towards the Fifa World Cup 2022 in Qatar," pp. 1–23, 2022.
- [3] Muhammad Ilham Fadila, Hanafi, and Anggit Dwi Hartanto, "Sentiment Analysis of the Indonesian National Team in the 2020 AFF Cup Using Naïve Bayes and K-Nearest Neighbor Algorithms," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 1–6, 2023, doi: 10.25139/inform.v8i1.5222.
- [4] T. A. Anastasya, A. D. P. Saka, M. J. 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," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 384–391, 2025.
- [5] D. A. Prabowo and S. Sudianto, "Analisis Sentimen Sepak Bola Indonesia pada Twitter menggunakan K-Nearest Neighbors dan Random Forest," *JSAI J. Sci. Appl. Informatics*, vol. 06, no. 02, pp. 217–227, 2023.
- [6] A. E. Pratama, A. Ariesta, and G. Gata, "Analisis Sentimen Masyarakat Terhadap Tim Nasional Indonesia Pada Piala AFF 2020 Menggunakan Algoritma K-Nearest Neighbors," *J. TICOM Technol. Inf. Commun.*, vol. 10, no. 3, pp. 187–196, 2022, [Online]. Available: <https://jurnal-ticom.jakarta.aptikom.org/index.php/Ticom/article/view/33/47>
- [7] M. D. Taufiqurahman, S. Anraeni, and H. Darwis, "Analisis Sentimen Komentar Konten Kreator Gaming Menggunakan Metode Naive Bayes dan KNN," *LINIER (Literatur Inform. Komputer)*, vol. 1, no. 4, pp. 317–327, 2024.
- [8] R. H. Nufus and U. Surapati, "Analisis Sentimen Persepsi Masyarakat Terhadap Timnas Indonesia U-23 dalam AFC-23 Asian Cup 2024 Pada Media Sosial X Menggunakan Metode Naïve Bayes Classifier," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 2647–2657, 2024.
- [9] Y. Akbar and A. N. Ihsan, "Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games

- Kamboja 2023 Menggunakan Algoritma Support Vector Machine,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 814–821, 2023, doi: 10.31539/intecom.v6i2.7670.
- [10] N. Leelawat *et al.*, “Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning,” *Heliyon*, vol. 8, no. 10, p. e10894, 2022, doi: 10.1016/j.heliyon.2022.e10894.
- [11] 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.
- [12] A. W. Siswanto and Neneng Nurhasanah, “Analisis Fenomena Childfree di Indonesia,” *Bandung Conf. Ser. Islam. Fam. Law*, vol. 2, no. 2, pp. 64–70, 2022, doi: 10.29313/bcsifl.v2i2.2684.
- [13] P. Nugraha, Rasiban, F. M. Sarimole, and Tundo, “Analisis Sentimen Kepuasan Publik Terhadap Masa Kepemimpinan Shin Tae Yong Menggunakan Algoritma Naïve Bayes,” vol. 9, no. March, pp. 149–158, 2025.
- [14] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [15] R. Mukarramah, D. Atmajaya, and L. B. Ilmawan, “Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter,” *Ilk. J. Ilm.*, vol. 13, no. 2, pp. 168–174, 2021, doi: 10.33096/ilkom.v13i2.851.168-174.
- [16] A. D. Dayani, Yuhandri, and G. W. Nurcahyo, “Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine,” *J. KomtekInfo*, vol. 11, no. 1, 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [17] E. Ronando, M. Yasa, and E. Indasyah, “Sistem Prediksi Kepribadian Manusia Berdasarkan Status Media Sosial Menggunakan Support Vector Machine,” vol. 17, no. 1, pp. 13–22, 2021.
- [18] F. Ramadhan, “Analisis Sentimen Timnas Sepakbola di Era STY,” Kaggle.com. [Online]. Available: <https://www.kaggle.com/datasets/pajarbebek/analisis-sentimen-timnas-sepakbola-di-era-sty/data>
- [19] A. F. Nur, Y. Salim, and Ramdaniah, “Analisis Sentimen Pengguna X Terhadap Perkembangan Artificial Intelligence (AI) Menggunakan Algoritma Machine Learning,” *LINIER (Literatur Inform. Komputer)*, vol. 1, no. 4, pp. 347–357, 2024.
- [20] M. D. Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola pada Google Play Store Menggunakan Algoritma Support Vector Machine,” *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 2621–3249, 2022.
- [21] S. Ramdani, Azahari, and T. Bustomi, “Analisis Sentimen Terhadap Pemain Diaspora Timnas Indonesia pada Media Sosial Instagram Menggunakan Metode Support Vector Machine (SVM),” vol. 28, no. 2, pp. 1–8, 2024, doi: 10.46984/sebatik.v28i2.0000.
- [22] N. F. Az-haari, D. Juardi, and A. Jamaludin, “Analisis Sentimen Terhadap Boikot Brand Pro-Israel pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus: Starbucks),” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4256–4261, 2024.
- [23] A. Sentiment, “Kamus slag,” Kaggle.com. [Online]. Available: <https://www.kaggle.com/datasets/fornigulo/kamus-slag>
- [24] F. M. A. Sofyan, N. Sulistiyowati, and A. Voutama, “ANALISIS SENTIMEN TERHADAP RESPON PERUBAHAN NAMA TWITTER MENJADI ‘ X ’ MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” vol. 8, no. 5, pp. 10987–10994, 2024.
- [25] A. B. Sasmita, B. Rahayudi, and L. Muflikhah, “Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1208–1214, 2022.
- [26] R. F. R. Pohan, D. E. Ratnawati, and I. Arwani, “Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 10, pp. 4924–4931, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [27] K. Hadi and E. Utami, “Analysis of K-NN with the Integration of Bag of Words , TF-IDF , and N-Grams for Hate Speech Classification on Twitter,” *JUITA J. Inform.*, vol. 12, no. 2, pp. 289–298, 2024.
- [28] A. D. W. M. Sidik, I. H. Kusumah, A. Suryana, Edwinanto, M. Artiyasa, and A. P. Junfithrana, “Gambaran Umum Metode Klasifikasi Data Mining,” *Fidel. J. Tek. Elektro*, vol. 2, no. 2, pp. 34–38, 2020.
- [29] L. Nurhidayati, Y. Umaidah, and U. Enri, “Analisis Sentimen Isu Childfree Di Media Sosial Twitter Menggunakan Algoritma Support Vector Machine,” *J. Ilm. Wahana Pendidik.*, vol. 10, no. 4, pp. 422–430, 2024.