Analisis Sentimen Hasil Pertandingan Timnas Indonesia di Kualifikasi Piala Dunia Pada Platform Instagram Menggunakan Algoritma Support Vector Machine (SVM).

Muchlas Andrey Pahlevi¹, Dwi Reza Ariyadi²

Universitas Mulawarman Samarinda Jl. Kuaro, Gn. Kelua, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75119

Article Info

Article history:

Received month dd, yyyy Revised month dd, yyyy Accepted month dd, yyyy

Kata Kunci:

Analisis Sentimen Timnas Indonesia Kualifikasi Piala Dunia Support Vector Machine (SVM) Media Sosial

ABSTRAK

Perkembangan sepak bola di Indonesia, khususnya Timnas Indonesia, telah menunjukkan kemajuan yang signifikan dalam beberapa tahun terakhir, terutama dalam kompetisi internasional seperti kualifikasi Piala Dunia. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap hasil pertandingan Timnas Indonesia melalui platform Instagram, menggunakan algoritma Support Vector Machine (SVM). Data dikumpulkan dari kolom komentar pada postingan akun Instagram resmi Timnas Indonesia yang berkaitan dengan pertandingan. Setelah melalui proses praproses teks, data tersebut ditransformasikan menggunakan metode TF-IDF.

Model SVM dengan kernel linear diterapkan untuk mengklasifikasikan sentimen menjadi tiga kategori: positif, netral, dan negatif. Hasil evaluasi menunjukkan akurasi keseluruhan mencapai 79%, dengan precision untuk sentimen negatif sebesar 77%, netral 79%, dan positif 80%. Recall untuk kategori netral mencapai 93%, mengindikasikan kemampuan model dalam mengidentifikasi sentimen netral dengan baik. F1-score menunjukkan keseimbangan antara precision dan recall, tertinggi pada kategori netral (0.86).

Temuan penelitian ini memberikan wawasan tentang persepsi masyarakat terhadap performa Timnas Indonesia, diharapkan dapat membantu pengelola tim dan pemangku kepentingan dalam merumuskan strategi komunikasi yang lebih efektif di media sosial.

This is an open access article under the CC BY-SA license.



Corresponding Author:

Suttichai Premrudeeprechacharn Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University 239 Huay Kaew Road, Muang District, Chiang Mai 50200, Thailand Email: suttichai@mail.com

Journal homepage: http://ijece.iaescore.com

1. PENDAHULUAN

Sepak bola adalah olahraga yang sangat digemari semua kalangan masyarakat Indonesia. Kini, sepak bola telah bertransformasi menjadi industri yang sangat menjanjikan. Selain menjadi industri, sepak bola juga menjadi produk dengan nilai yang terus berubah, menjadikannya komersial bagi banyak pihak. Berbagai aspek kehidupan, baik secara langsung maupun tidak langsung, mempengaruhi sepak bola, termasuk politik, sosial, ekonomi, hiburan, teknologi, dan lainnya [1]. Indonesia, yang belum pernah lolos ke Piala Dunia, kini mulai diperhitungkan dalam olahraga ini melalui ajang *kualifikasi Piala Dunia*. Ajang ini menjadi panggung unjuk kebolehan bagi negara-negara di seluruh dunia. Masyarakat Indonesia pun aktif berbagi pendapat mereka di *platform* media sosial, seperti Instagram.

Sebagai salah satu *platform* media sosial yang paling populer di Indonesia, Instagram menjadi tempat bagi masyarakat untuk mengungkapkan pandangan mereka tentang berbagai hal, termasuk hasil pertandingan sepak bola timnas. Opini-opini tersebut mencakup beragam sentimen, mulai dari dukungan positif hingga kritik negatif. Dalam konteks ini, penting untuk menganalisis sentimen yang berkembang di kalangan pengguna Instagram guna memahami persepsi mereka terhadap performa timnas Indonesia.

Berdasarkan penelitian sebelumnya, metode analisis sentimen telah banyak digunakan untuk mengevaluasi opini publik di media sosial. Salah satu penelitian yang relevan oleh Muhammad Fadli Asshiddiqi dan Kemas Muslim Lhaksmana [2], yang membandingkan metode *Decision Tree* dan *Support Vector Machine (SVM)* dalam analisis sentimen di Instagram terkait kinerja PSSI. Hasilnya menunjukkan bahwa metode SVM memberikan akurasi yang lebih tinggi, yaitu 94.36%, dibandingkan *Decision Tree*. Oleh karena itu, dalam penelitian ini, metode SVM akan digunakan untuk menganalisis sentimen publik terhadap hasil pertandingan timnas Indonesia di *kualifikasi Piala Dunia*.

2. LANDASAN TEORI

2.1. Media Sosial dan Persepsi Publik

Media sosial merupakan wadah untuk saling berinteraksi dan berkomunikasi, memungkinkan masyarakat untuk berbagi pandangan mereka tentang berbagai topik secara terbuka. Salah satu media sosial yang ramai digunakan masyarakat indonesia adalah Instagram. Sebagai negara demokrasi opini publik pada media sosial berguna sebagai alat untuk mengkritisi kinerja timnas Indonesia[5]. efektivitas media sosial sebagai alat kritik dinilai sangat berpengaruh bahkan sampai di tingkat kebijakan pemerintah[6].

2.2. Analisis Sentimen

Analisis sentimen adalah teknik dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang digunakan untuk mengidentifikasi dan mengkategorikan opini dalam teks menjadi positif, negatif, atau netral (Indurkhya & Damerau, 2010)[3]. Dalam konteks media sosial seperti Instagram, analisis sentimen sangat bermanfaat untuk memahami reaksi publik terhadap suatu peristiwa, termasuk hasil pertandingan sepak bola. Opini pengguna dapat mencerminkan kepuasan, kekecewaan, atau harapan mereka terhadap performa Timnas Indonesia di ajang Kualifikasi Piala Dunia.

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang sering digunakan untuk klasifikasi data, termasuk analisis sentimen. Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang sering digunakan untuk klasifikasi data, termasuk analisis sentimen. Popularitas SVM tidak lepas dari kesederhanaan dan fleksibilitasnya dalam menangani berbagai masalah klasifikasi. SVM memberikan kinerja prediktif yang seimbang, bahkan ketika jumlah sampel terbatas, yang membuatnya cocok untuk berbagai aplikasi (Pisner & Schnyer, 2020)[4]. Faktor fleksibilitas lah yang menjadikan SVM banyak digunakan dalam analisis sentimen pada data yang kompleks, seperti komentar di media sosial, untuk mengklasifikasikan opini publik terhadap peristiwa tertentu, termasuk kinerja tim nasional dalam pertandingan internasional.

2.4. Pemrosesan Text

Pemrosesan teks adalah tahapan penting dalam analisis sentimen untuk mengubah data mentah menjadi bentuk yang siap diolah oleh algoritma pembelajaran mesin. Langkah-langkah ini meliputi proses *preprocessing* yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dan mengurangi noise dalam teks (Indurkhya & Damerau, 2010)[3]. Beberapa tahapan utama dalam pemrosesan teks meliputi *tokenizing*, *stopword removal*, *stemming*, dan *normalization* (Rismi Nurlaely et al., 2023)[4].

Setelah melalui tahap *preprocessing*, data teks yang telah dibersihkan kemudian diubah menjadi representasi numerik menggunakan metode seperti *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk memberi bobot pada kata-kata dalam dokumen berdasarkan seberapa sering kata tersebut muncul dalam dokumen dan keseluruhan korpus. Teknik ini membantu algoritma seperti Support Vector Machine (SVM) untuk fokus pada kata-kata yang memiliki nilai informasi yang signifikan (Pisner & Schnyer, 2020)[5]. Tahap pemrosesan teks ini sangat penting dalam memastikan akurasi dan efisiensi analisis sentimen dalam data yang bersifat tidak terstruktur, seperti komentar di platform media sosial Instagram.

3. METODE

Sebelum melakukan analisis sentimen, penelitian ini melalui beberapa langkah penting untuk memastikan data siap diolah dan memberikan hasil yang akurat. Tahapan-tahapan yang diterapkan dalam penelitian ini adalah sebagai berikut:



gambar 3.1 tahapan penelitian

3.1. Identifikasi Masalah dan Tujuan Penelitian

Hal yang mendasari penelitian ini adalah permasalahan banyaknya opini masyarakat di media sosial, khususnya Instagram, yang menyampaikan reaksi terkait hasil pertandingan Timnas Indonesia di ajang kualifikasi Piala Dunia. Opini yang tersebar luas tersebut sulit untuk diukur dan diklasifikasikan secara manual. Tantangan utama terletak pada pengelompokan sentimen publik yang terdiri dari sentimen positif, negatif, atau netral terhadap performa Timnas. Oleh karena itu, diperlukan metode analisis yang tepat untuk mengolah data yang besar dan tidak terstruktur. Algoritma Support Vector Machine (SVM), yang telah terbukti efektif dalam klasifikasi data teks, digunakan dalam penelitian ini untuk mengkategorikan sentimen dari komentar-komentar pengguna Instagram. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen masyarakat, serta memberikan wawasan bagi manajemen Timnas Indonesia dan PSSI mengenai persepsi publik, yang dapat digunakan sebagai dasar untuk pengambilan keputusan yang lebih baik.

3.2. Pengumpulan Dataset

Pengumpulan dataset dilakukan dengan memilih postingan pada akun Instagram resmi Timnas Indonesia yang berkaitan dengan kualifikasi Piala Dunia sebagai sumber data. Postingan yang telah dipilih kemudian diambil menggunakan ekstensi Chrome, yaitu IG Comment Export Tool. Dari proses ini, diperoleh 7.257 data mentah berisi kolom id,username,text comment, profile Url,date yang belum berlabel.

3.3. Data Preprocessing

Sebelum melakukan modeling, terlebih dahulu melakukan data preprocessing. Data yang diperoleh dari instagram memiliki banyak noise dan data yang tidak terstruktur dengan baik maka perlu dilakukan proses preprocessing. Preprocessing merupakan tahap awal untuk penyeleksian atau membuang kata yang tidak diperlukan pada teks komentar instagram[7]. Berikut adalah beberapa tahapan preprocessing.



3.3.1. Cleaning Text

Pada proses cleaning text digunakan untuk membersihkan kata dari yang tidak diperlukan seperti simbol, emoji, spasi, kolom null dan lain-lain.

3.3.2. Case Folding

Mengkonversi keseluruhan teks pada dokumen menjadi suatu bentuk standar. Proses ini mengkonversi uppercase menjadi lowercase[8].

3.3.2. Tokenization

Pada proses tokenization digunakan untuk memisahkan kata atau huruf dari tanda baca dan simbol[8].

3.3.4. Stopword

Pada proses Stopword digunakan untuk menghilangkan kata yang dianggap tidak perlu dalam pengolahan data review [8].

3.3.5. Labeling

Pada proses Labeling setiap teks atau kalimat diberi label positif, negatif, atau netral.

3.3. Pembagian Data Latih dan Data Uji

Data pelatihan dan pengujian dapat dipecah/bagi ke dalam tiga skema pembagian data, yaitu 60/40, 70/30, dan 80/20 [12]. Data mentah penelitian yang berjumlah 7.257 menghasilkan data bersih setelah proses cleaning sebanyak 6.731, dibagi menggunakan skema 80/20.

3.3. Pemodelan

Pemodelan dilakukan setelah pelabelan dengan memberi label positif, negatif atau netral. Pemodelan dilakukan menggunakan term weighting atau pembobotan setiap dokumen. Metode yang digunakan dalam pembobotan ini menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) yaitu metode untuk memberikan bobot setiap kata untuk menentukan seberapa jauh keterhubungan kata terhadap dokumen. Pembobotan ini merupakan ukuran statistik untuk mengevaluasi kepentingan sebuah kata bagi dokumen [9]. Melakukan klasifikasi menggunakan Support Vector Machine (SVM). SVM adalah metode pembelajaran yang menganalisis data dan mengenali pola yang digunakan untuk klasifikasi [10].

3.3. Evaluasi

Setelah pembuatan model, perlu dilakukan evaluasi terhadap nilai akurasi, presisi,recall dan fl-score. Semakin tinggi nilai akurasi menggunakan algoritma Support Vector Machine(SVM) berarti

algoritma SVM terbukti mampu menjadi pengklasifikasi text yang baik untuk pengujian sentimen untuk kategori positif,negatif, dan netral terhadap hasil pertandingan Timnas Indonesia di Kualifikasi Piala Dunia.

4. HASIL DAN PEMBAHASAN

Dalam upaya mencapai penyelesaian masalah disebutkan di bagian pendahuluan. Penelitian menggunakan metode SVM ini memproses dataset komentar masyarakat yang potongan datanya dapat dilihat pada tabel 4.1

userName	text	date
januari_initiative	Pak Erick Tohirapa tindakan bagi wasit perta	2024/09/15 22:09:32
okingtarni	Lewat dulu kawan😆	2024/09/16 00:56:42
ggh.yper	@pssi tolong deh wasitnya di blacklist kontrov	2024/09/16 01:21:29
vikranhasri	Gacorrr waaak 🙌 🔥	2024/09/11 10:35:04
jejegyuu	kak jepun ngeri bgt	2024/09/11 10:35:04
uchilll.	Semangat terus Garuda 🔥 🔥 ❤️ ❤️	2024/09/11 10:35:07
agung_cloud	1 poin yg bearti 👌	2024/09/11 10:35:07

tabel 4.1 data mentah

4.1. Data Preprocessing

Pengolahan data mentah dilakukan melalui beberapa tahapan, yaitu: **pertama**, cleaning text yang berfungsi untuk membersihkan kata-kata yang tidak diperlukan, seperti simbol, emoji, spasi berlebih, serta kolom kosong (null); **kedua**, *case folding* yang mengonversi seluruh teks menjadi bentuk standar, yaitu mengubah huruf kapital (uppercase) menjadi huruf kecil (lowercase); **ketiga**, tokenization, yaitu proses yang digunakan untuk memisahkan kata atau huruf dari tanda baca dan simbol; **keempat**, stopword removal, yang digunakan untuk menghilangkan kata-kata yang dianggap tidak relevan dalam pengolahan data ulasan.

text	text_clean	translated_text	subjektivitas	polaritas	sentimen
Lewat dulu kawan ≅	kawan	friends	0.000000	0.000000	netral
kak jepun ngeri bgt	kak jepun ngeri bgt	japan is so scary	1.000000	-0.500000	negatif
1 poin yg bearti	poin bearti	meaningful points	0.500000	0.500000	positif

tabel 4.2 hasil tahapan preprocessing data

Pada Tabel 4.1, data mentah yang digunakan belum memiliki label (unsupervised). Agar dapat melakukan pemodelan, data tersebut perlu diberi label sentimen (netral, positif, atau negatif) sehingga menjadi data supervised, dengan menentukan nilai subjektivitas dan polaritasnya [11].

4.2. Pemodelan

Model pengklasifikasian dilakukan dengan cara pembagian data menjadi dua bagian, yaitu training data dan test ini untuk menentukan jenis banding terbaik, pada penelitian ini menggunakan perbandingan 20% training dan 80% testing.

TF-IDF (Term Frequency-Inverse Document Frequency), teknik representasi teks yang sering digunakan dalam Natural Language Processing (NLP) untuk mengubah teks mentah menjadi bentuk numerik yang bisa diproses oleh model pembelajaran mesin. berikut adalah rumus untuk TF-IDF

$$TF - IDF(w, d)) = TF(w, d) * (\log (DFN(w)))$$

Keterangan:

TF - IDF(w, d)): bobot setiap kata dalam dokumen

w: kata d dokumen

TF(w, d): frekuensi munculnya 'kata' dalam 'dokumen'

IDF(w): invers DF dari 'kata'N: nilai total dokumen

DF(w): jumlah dokumen yang mengandung 'kata'

Word Cloud, teknik visualisasi yang digunakan untuk melihat distribusi frekuensi kata-kata dalam teks, kata yang sering muncul dalam dokumen akan tampil dengan ukuran lebih besar.

Trombined $\bigcup_{i=1}^{n} Ti$, $f(w) = \sum_{i=1}^{n} fi(w)$, $S(w) = \alpha \cdot f(w)$, $P(w) \propto S(w)$,

 $WordCloud = WordCloud(Tcombined, width = 800, height = 400, background_color = 'white')$

Keterangan:

Tcombined : gabungan dari seluruh teks Ti dari dokumen bersih
Ti : teks hasil pembersihan dari dokumen ke-i dalam data

f(w): frekuensi dari kata w yang dihitung dari seluruh teks Tcombined

fi(w): frekuensi kata w dalam dokumen Ti

S(w): ukuran visual dari kata w dalam word cloud

α : konstanta skala yang mengatur ukuran maksimum kata

P(w): menunjukkan prioritas posisi dari kata w,dimana kata dengan ukuran lebih besar di pusat

width : lebar 800 height : tinggi 400



gambar 4.1 WordCloud

4.3. Evaluasi

Pada tahap evaluasi, model klasifikasi *Support Vector Machine (SVM)* diukur menggunakan matrik evaluasi seperti precision, recall, f1-score, dan accuracy. Berdasarkan hasil evaluasi, model memiliki performa yang cukup baik dengan akurasi sebesar 0.79 dari total 1347 data uji. Dapat dilihat pada tabel 4.3

Kategori	Precision	Recall	F1-Score	Support
Negatif	0.77	0.52	0.62	181
Netral	0.79	0.93	0.86	808
Positif	0.80	0.61	0.70	358
Accuracy			0.79	1347
Macro Avg	0.79	0.69	0.72	1347
Weighted Avg	0.79	0.79	0.78	1347

tabel 4.3

Untuk kelas *negatif*, precision adalah 0.77, recall 0.52, dan f1-score 0.62, menunjukkan bahwa model cukup baik dalam mengidentifikasi sentimen negatif, namun recall yang rendah menunjukkan bahwa model kurang mampu mendeteksi semua instance negatif dengan baik.

Pada kelas *netral*, precision sebesar 0.79 dan recall 0.93, menghasilkan f1-score 0.86. Ini mengindikasikan bahwa model sangat baik dalam mengidentifikasi dan mendeteksi sentimen netral, dengan recall yang tinggi. Untuk kelas *positif*, precision 0.80, recall 0.61, dan f1-score 0.70 menunjukkan bahwa meskipun model akurat dalam memprediksi instance positif, ada beberapa instance positif yang tidak teridentifikasi (recall yang rendah).

Dari segi rata-rata, *macro average* precision adalah 0.79, recall 0.69, dan f1-score 0.72, yang berarti performa model sedikit tidak seimbang di semua kelas. Namun, *weighted average* untuk precision, recall, dan f1-score adalah 0.79, 0.79, dan 0.78, yang menunjukkan bahwa model secara keseluruhan memberikan performa yang stabil dan cukup baik dalam menganalisis sentimen dari data komentar Instagram terkait pertandingan Timnas Indonesia.

5. KESIMPULAN

Penelitian dengan dataset mentah sebanyak 7.257 menghasilkan data bersih setelah proses cleaning sebanyak 6.731, dengan jumlah data untuk masing-masing kategori: positif 1.848, negatif 889, dan netral 3.994. Pembagian dataset dilakukan dengan perbandingan 80% untuk pelatihan dan 20% untuk pengujian. Model Support Vector Machine (SVM) yang diterapkan menunjukkan akurasi keseluruhan sebesar 79%. Precision untuk sentimen negatif mencapai 77%, netral 79%, dan positif 80%. Recall untuk kategori netral mencapai 93%, yang menunjukkan bahwa model sangat baik dalam mengidentifikasi sentimen netral. F1-score tertinggi terdapat pada kategori netral (0.86), yang menunjukkan keseimbangan antara precision dan recall. Nilai akurasi 79% ini membuktikan bahwa model SVM yang digunakan cukup efektif dalam menganalisis sentimen masyarakat terhadap hasil pertandingan Timnas Indonesia di Kualifikasi Piala Dunia pada platform Instagram, meskipun terdapat ruang untuk perbaikan dalam meningkatkan akurasi pada kategori negatif dan positif.

6. ACKNOWLEDGEMENTS

Penulis ingin menyampaikan terima kasih yang tulus kepada semua pihak yang telah berkontribusi dalam penyelesaian penelitian ini. Ucapan terima kasih khusus disampaikan kepada Ibu Ir. Novianti Puspitasari, S.Kom., M.Eng. atas bimbingan dan materi yang diberikan selama di kelas, sehingga penulis dapat menyelesaikan penelitian ini. Terima kasih juga ditujukan kepada pedagang nasi campur seharga 11.000, karena berkat beliau, penulis memiliki energi untuk mengerjakan penelitian ini.

7. REFERENCES

- [1] Athalarik, Fadli Muhammad, and Udi Rusadi. "Sepak Bola Indonesia dalam Perspektif Komodifikasi Budaya Populer." Jurnal Pendidikan Tambusai 7.3 (2023): 25476-25487.
- [2] G. Nguyen *et al.*, "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.
- [3] Indurkhya, N., & Damerau, F.J. (Eds.). (2010). Handbook of Natural Language Processing (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781420085938
- [4] Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.
- [5] Pangestu, D., Malik, M., & Pribadi, M. R. (2024). Analisis Sentimen Hasil Pertandingan Sepakbola Timnas Indonesia di Piala

- Asia U-23 pada Platform Youtube menggunakan Algoritma Suport Vector Machine (SVM): Sentiment Analysis of Indonesian National Team Football Match Results in the U-23 Asian Cup on the YouTube Platform using the Support Vector Machine (SVM) Algorithm. Applied Information Technology and Computer Science, 3(1), 38-48.
- [6] Latifah, K. N., & Najicha, F. U. (2022). Implikasi media sosial terhadap formulasi kebijakan publik. Jurnal Kewarganegaraan, 6(1), 494-501.
- [7] Darwis D Pratiwi E S and Pasaribu a F O, 2020 Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia Edutic Sci. J. Informatics Educ. 7, 1 p. 1–11.
- [8] Ernawati S Wati R Nuris N Marita L S and Yulia E R, 2020 Comparison of Naïve Bayes Algorithm with Genetic Algorithm and Particle Swarm Optimization as Feature Selection for Sentiment Analysis Review of Digital Learning Application J. Phys. Conf. Ser. 1641, 1.
- [9] Savitri N L P C Rahman R A Venyutzky R and Rakhmawati N A, 2021 Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning J. Tek. Inform. dan Sist. Inf. 7, 1 p. 47–58.
- [10] Basari A S H Hussin B Ananta I G P and Zeniarja J, 2013 Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization Procedia Eng. 53 p. 453–462
- [11] Afdal, M., & Elita, L. R. (2022). Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor. Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, 8(1), 78-87.
- [12] Damasela, R., Tomasouw, B. P., & Leleury, Z. A. (2022). Penerapan Metode Support Vector Machine (SVM) Untuk Mendeteksi Penyalahgunaan Narkoba. PARAMETER: Jurnal Matematika, Statistika Dan Terapannya, 1(2), 111-122.

BIOGRAPHIES OF AUTHORS



Muchlas Andrey Pahlevi 2209106082.



Dwi Reza Ariyadi 2209106090