

Analisis Sentimen Terhadap Pemain Naturalisasi dan Lokal Tim Nasional Sepakbola Indonesia Menggunakan Support Vector Machine

Fadlan Arrazak*, M Afdal, Rice Novita, Megawati Megawati

Fakultas Sains dan Teknologi, Prodi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹*12150312250@students.uin-suska.ac.id, ²m.afdal@uin-suska.ac.id, ³rice.novita@uin-suska.ac.id,

⁴megawati@uin-suska.ac.id

Email Penulis Korespondensi: 12150312250@students.uin-suska.ac.id

Submitted: 30/05/2025; Accepted: 25/06/2025; Published: 02/09/2025

Abstrak—Masuknya pemain naturalisasi ke tim nasional sepak bola Indonesia memicu berbagai respons publik, khususnya di media sosial seperti Twitter. Penelitian ini bertujuan membandingkan opini publik terhadap pemain naturalisasi dan pemain lokal melalui analisis sentimen. Sebanyak 2.342 tweet diklasifikasikan ke dalam tiga kategori sentimen: positif, netral, dan negatif. Pemain naturalisasi menerima lebih banyak sentimen positif 809 dibandingkan sentimen negatif 333 dan netral 231. Sementara itu, pemain lokal mendapat 465 sentimen positif, 317 negatif, dan 187 netral, menunjukkan kecenderungan publik yang lebih menyukai pemain naturalisasi. Analisis lanjutan dilakukan menggunakan algoritma Support Vector Machine (SVM) dengan metode SMOTE untuk penyeimbangan data, berdasarkan lima aspek yaitu performa, pengalaman, kondisi fisik, adaptasi, dan komunikasi. Hasil klasifikasi menunjukkan pemain naturalisasi unggul dalam aspek kondisi fisik dengan nilai akurasi 96%, performa 90%, dan adaptasi 90%. Sebaliknya, pemain lokal hanya lebih unggul dalam aspek komunikasi dengan akurasi 92%. Dari segi presisi dan recall, pemain naturalisasi kembali mencatat hasil terbaik pada kondisi fisik sebanyak 97% dan recall 96%, sementara pemain lokal tetap unggul pada aspek komunikasi presisi dan recall 92%. Temuan ini memberikan gambaran penting bagi pembuat kebijakan dan institusi sepakbola dalam menyusun strategi yang lebih efektif.

Kata Kunci: Analisis Sentimen; Pemain Lokal; Pemain Naturalisasi; Support Vector Machine; Tim Nasional Indonesia; Text Mining; X

Abstract—The inclusion of naturalized players in Indonesia's national football team has sparked diverse public reactions, particularly on social media platforms like Twitter. This study aims to compare public opinion toward naturalized and local players through sentiment analysis. A total of 2,342 tweets were categorized into three sentiment classes: positive, neutral, and negative. Naturalized players received a higher number of positive sentiments, totaling 809, compared to 333 negative and 231 neutral sentiments. In contrast, local players gained 465 positive sentiments, 317 negative, and 187 neutral, indicating a generally more favorable perception of naturalized players among the public. Further analysis was conducted using the Support Vector Machine (SVM) classification algorithm along with the SMOTE technique for data balancing, focusing on five key aspects: performance, experience, physical condition, adaptability, and communication. The classification results showed that naturalized players outperformed in physical condition with an accuracy of 96 percent, followed by performance and adaptability, each at 90 percent. On the other hand, local players showed superiority only in communication with an accuracy of 92 percent. In terms of precision and recall, naturalized players again led in physical condition, achieving 97 percent precision and 96 percent recall, while local players excelled in communication with both precision and recall at 92 percent. These findings offer valuable insights for policymakers and football organizations in formulating more effective naturalization strategies.

Keywords: Indonesian National Team; Local Players; Naturalized Players; Sentiment Analysis; Support Vector Machine; Text Mining; X

1. PENDAHULUAN

Dalam beberapa tahun terakhir, tim nasional sepak bola Indonesia mengalami perubahan besar seperti adanya kebijakan naturalisasi pemain asing dan pengembangan pemain lokal menjadi dua fokus utama dalam transformasi tim nasional sepak bola Indonesia. Pemain naturalisasi adalah pesepak bola asing yang memperoleh kewarganegaraan Indonesia melalui proses hukum, sehingga dapat membela Timnas setelah memenuhi syarat dari FIFA, proses ini bisa lebih cepat jika pemain memiliki garis keturunan Indonesia [1]. Sementara itu, pemain lokal merujuk pada individu yang dibina melalui sistem pengembangan sepak bola nasional dan telah memiliki status kewarganegaraan Indonesia sejak lahir, tanpa melalui proses naturalisasi maupun berasal dari latar belakang diaspora [2]. Umumnya, mereka meniti karier dari level usia dini melalui berbagai jalur pembinaan seperti akademi klub lokal, kompetisi antar pelajar, hingga liga profesional dalam negeri. Sejak Shin Tae-yong, pelatih asal Korea Selatan, mulai melatih tim pada akhir 2019, perhatian publik terhadap kedua aspek ini semakin meningkat [3].

Kebijakan naturalisasi pemain dianggap sebagai upaya untuk meningkatkan kualitas dan performa Timnas Indonesia dengan menghadirkan pemain berpengalaman dan berkualitas dari luar negeri [4]. Namun, kebijakan ini juga menghadapi kritik dari beberapa kalangan yang merasa bahwa naturalisasi menghambat perkembangan pemain lokal dan merusak identitas nasional dalam olahraga. Perdebatan antara pro dan kontra ini menciptakan beragam opini yang tersebar luas di media sosial twitter [5]. Banyaknya komentar di Twitter dapat dimanfaatkan melalui analisis sentimen, yaitu metode untuk mengklasifikasikan opini dalam bentuk teks. Tweet dikumpulkan dan diproses menggunakan teknik pengolahan teks untuk mengetahui sentimen yang terkandung di dalamnya.

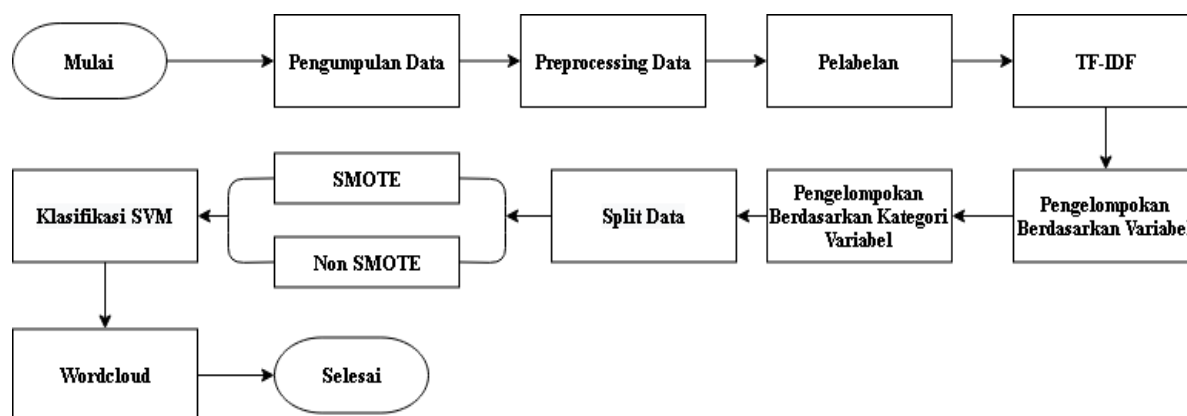
Penelitian sebelumnya yang dilakukan Ramdani et.al. [6] menggunakan algoritma Support Vector Machine (SVM) untuk menganalisis sentimen publik mengenai pemain diaspora Tim Nasional Indonesia di media sosial Instagram, dan mendapatkan akurasi mencapai 82%. Dimana ini merupakan algoritma yang baik dalam

mengklasifikasikan sentiment analisis. SVM menjadi metode populer dalam berbagai studi data mining karena memberikan performa yang unggul. Selain itu, dengan menambahkan fungsi kernel ke SVM, akurasi hasil dapat meningkat. Penelitian yang dilakukan Rusdi Asrianto dan Melda Herwinda (2023) [7] tentang analisis sentimen terhadap kenaikan harga kebutuhan pokok yang diambil dari platform media sosial YouTube menggunakan algoritma support vector machine dan penerapan SMOTE memperoleh tingkat akurasi 86,33%. Penelitian oleh Styawati et al. (2021) [8] yang membahas tentang Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine, memperoleh tingkat akurasi di kernel linear sebesar 98,67%, sedangkan dengan kernel Radial Basis Function (RBF), diperoleh akurasi sebesar 98,34%.

Dalam penelitian ini, algoritma Support Vector Machine (SVM) digunakan untuk mengkaji sentimen masyarakat terhadap kebijakan pemain naturalisasi dan pemain lokal pada timnas sepak bola Indonesia. Studi ini juga bertujuan untuk membandingkan kecenderungan opini masyarakat berdasarkan jenis pemain, yaitu naturalisasi dan lokal, guna melihat perbedaan persepsi yang ada. Data dikumpulkan melalui tweet yang relevan dengan menggunakan kata kunci terkait pemain naturalisasi dan lokal, serta variabel pembanding lainnya [9]. Temuan dari penelitian ini diharapkan mampu memberikan informasi yang berguna bagi PSSI dan pihak terkait dalam mengevaluasi dampak kebijakan tersebut, serta memberikan kontribusi pada pengembangan metode analisis sentimen berbasis machine learning yang lebih luas.

2. METODOLOGI PENELITIAN

Gambar 1 berikut merupakan tahapan dari penelitian yang dilakukan.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Pengumpulan dataset diproses dengan mengambil data dari platform Twitter/X. Data dikumpulkan menggunakan kata kunci terkait seperti "naturalisasi timnas," "naturalisasi sepakbola Indonesia", "pemain lokal sepakbola Indonesia", "performa pemain lokal" dll untuk mengidentifikasi tweet yang relevan dengan topik tersebut.

2.2 Preprocessing Data

Preprocessing data adalah langkah awal dalam data mining yang bertujuan untuk mengubah data mentah, yang juga dikenal sebagai raw data, yang diperoleh dari berbagai sumber menjadi informasi yang bersih dan siap digunakan untuk tahap selanjutnya [10]. Adapun langkah preprocessing yang dilakukan dalam penelitian ini adalah

a. **Cleaning Data**

Cleaning data adalah langkah krusial dalam preprocessing data yang bertujuan untuk melakukan proses pembersihan, normalisasi, dan penyiapan data supaya dapat digunakan dalam proses analisis. Tahapan ini bertujuan untuk memperbaiki mutu data sekaligus mengoptimalkan performa model [11].

b. **Case Folding**

Case folding bertujuan untuk mengubah seluruh huruf menjadi bentuk kecil. Pada tahap ini, setiap karakter diperiksa secara terpisah, dan apabila ditemukan huruf besar, maka akan diubah menjadi huruf kecil [12].

c. **Tokenizing**

Tokenisasi merupakan proses memecah data menjadi potongan kata atau karakter sesuai dengan kebutuhan sistem.

d. **Stopword**

Tujuan dari penghapusan stopwords adalah untuk membuang karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna penting. Contoh dari kata-kata tersebut antara lain "dan", "atau", "tetapi", dan sebagainya.

e. **Stemming**

Stemming adalah metode yang digunakan untuk menghapus akhiran dan awalan dari suatu kata, sehingga menghasilkan bentuk dasar kata tersebut. Contohnya, kata "berenang" diubah menjadi "renang".

2.3 Pelabelan

Pelabelan adalah langkah dalam menetapkan kategori atau tanda pada tiap data dalam dataset. Langkah ini bertujuan memberi kejelasan identitas pada data, sehingga bisa dimanfaatkan dalam pelatihan model machine learning untuk tugas seperti prediksi atau klasifikasi [13].

2.4 TF-IDF

Tujuan dari pembobotan TF-IDF adalah untuk menetapkan bobot pada setiap kata dalam dokumen berdasarkan dua metrik utama, yaitu TF dan IDF. TF mengukur seberapa sering sebuah kata muncul dalam satu dokumen, sedangkan IDF mengukur seberapa sering kata tersebut muncul di seluruh kumpulan dokumen. [14].

$$W_{ij} = tf_{ij} \times \log \left(\frac{N}{df_i} \right) \quad (1)$$

Persamaan (1) mendefinisikan W_{ij} sebagai bobot perhitungan kata i yang berasal dari dokumen j , di mana tf_{ij} merepresentasikan frekuensi kemunculan kata i dalam dokumen j , df_i menunjukkan jumlah dokumen yang memuat kata i , dan N adalah total jumlah dokumen dalam korpus.

2.5 Pengelompokan Berdasarkan Variabel

Data dalam penelitian ini dikelompokkan berdasarkan variabel jenis pemain, yaitu pemain naturalisasi dan pemain lokal Timnas Indonesia. Pengelompokan memiliki tujuan untuk membandingkan sentimen masyarakat pada kedua variabel tersebut, setelah data diproses melalui algoritma SVM. Klasifikasi dilakukan dengan mengidentifikasi konteks yang merujuk pada sepakbola Indonesia [15].

2.6 Pengelompokan Berdasarkan Kategori Variabel

Evaluasi dilakukan berdasarkan variabel kategorik, yaitu jenis pemain yang terdiri dari dua kategori yaitu pemain naturalisasi dan pemain lokal. Setelah data sudah memiliki label, hasil pelabelan sentimen dibandingkan antar kategori untuk mengetahui perbedaan kecenderungan opini publik. Evaluasi ini bertujuan untuk melihat apakah terdapat pola sentimen yang signifikan antara kedua jenis pemain berdasarkan hasil analisis data [9].

2.7 Split Data

Penelitian ini menerapkan metode Train-Test Split untuk membagi data menjadi dua komponen utama, yakni data pelatihan dan data pengujian. Teknik ini bertujuan untuk menguji kinerja model secara objektif pada data yang tidak digunakan dalam proses pelatihan [16].

2.8 Penerapan SMOTE

Pada studi ini SMOTE atau Random Oversampling digunakan guna menganani ketimpangan kelas dengan menggandakan sampel dari kelas minoritas, sehingga distribusi kelas menjadi lebih seimbang dan model dapat mempelajari pola kedua kelas secara lebih efektif [17].

2.9 Support Vector Machine (SVM)

Pada tahap ini, algoritma Support Vector Machine diterapkan untuk klasifikasi sentimen. Algoritma ini dipilih karena kemampuannya dalam mengelola data teks yang besar, SVM cepat diadopsi karena kemampuannya beroperasi dengan sejumlah kecil tingkatan hiper, jaminan teoretis, dan hasil yang baik dalam aplikasi praktis. Meskipun awalnya digunakan untuk mengklasifikasikan data numerik, SVM juga mampu menangani permasalahan yang berkaitan dengan data teks [18]. Dalam algoritma SVM, pemilihan kernel memegang peranan krusial karena mengatur ruang representasi data tempat fungsi klasifikasi akan dibentuk. Melalui teknik *kernel trick*, kita dapat memilih jenis kernel yang sesuai seperti *linear*, *rbf* untuk menyesuaikan model dengan karakteristik data yang berbeda-beda. Berikut rumus dari persamaan kernel SVM [19]:

a. Kernel *linear* :

$$K(x, y) = x \cdot y \quad (2)$$

Kernel *linear* menggunakan operasi *dot product* antara dua vektor dan digunakan ketika data dapat dipisahkan secara *linear* tanpa perlu transformasi ruang fitur.

b. Kernel *rbf* :

$$K(x, y) = e^{-|x \cdot y|^2 / 2\sigma^2} \quad (3)$$

Kernel *rbf* mengukur kemiripan antara dua titik berdasarkan jarak *Euclidean* dan sangat efektif untuk data yang tidak dapat dipisahkan secara *linear*.

2.10 WordCloud

Wordcloud menyajikan frekuensi kemunculan kata dalam kumpulan teks secara visual, di mana kata-kata yang muncul lebih sering ditampilkan dengan ukuran huruf yang lebih besar. Dalam analisis opini publik di Twitter,

wordcloud dimanfaatkan untuk mengidentifikasi topik atau kata kunci yang paling banyak digunakan oleh pengguna, sehingga memberikan gambaran mengenai fokus atau sentimen yang paling menonjol dalam analisis tersebut [20].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data yaitu langkah yang dikerjakan agar memperoleh data yang diperlukan untuk mendukung penelitian. Data di riset ini diperoleh dari Media Sosial Twitter/X dengan Proses crawling data dilakukan melalui bahasa pemrograman Python. Total data yang di peroleh sebanyak 2342 data, dengan rentang waktu dari tanggal 01 Mei 2024 sampai dengan 01 Mei 2025.

3.2 Tahap Preprocessing

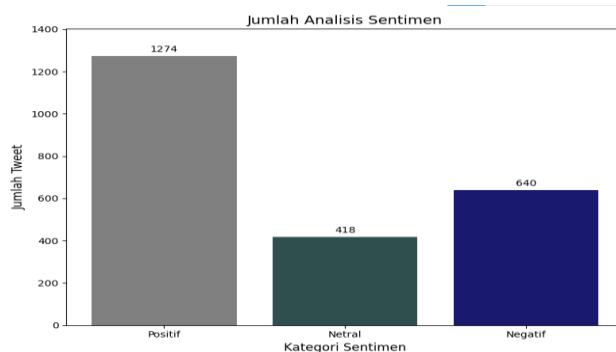
Bagian ini menyoroti pembersihan teks dari karakter yang tidak relevan, normalisasi bahasa, serta penghapusan kata umum (stopword) yang tidak memiliki makna penting dalam analisis sentimen. Hasil dari tahapan praproses ditampilkan pada Tabel 1.

Tabel 1. Hasil *Preprocessing* Data

Tahapan	Hasil
Data Awal	@TheIndoLad @rice_ur tapi local pride inilah yang jauh membawa sepakbola indonesia lebih bermarwah. klub local pride dengan membina pemain lokal sebelum diekspor utk abroad kualitas pemain meningkat + penopang buat timnasnya....
<i>Cleaning</i>	tapi local pride inilah yang jauh membawa sepakbola indonesia lebih bermarwah klub JLeague local pride dengan membina pemain lokal sebelum diekspor utk abroad kualitas pemain meningkat penopang buat timnasnya
<i>Case Folding</i>	tapi local pride inilah yang jauh membawa sepakbola indonesia lebih bermarwah klub jleague local pride dengan membina pemain lokal sebelum diekspor utk abroad kualitas pemain meningkat penopang buat timnasnya
<i>Tokenizing</i>	['tapi', 'local', 'pride', 'inilah', 'yang', 'jauh', 'membawa', 'sepakbola', 'indonesia', 'lebih', 'bermarwah', 'klub', 'jleague', 'local', 'pride', 'dengan', 'membina', 'pemain', 'lokal', 'sebelum', 'diekspor', 'utk', 'abroad', 'kualitas', 'pemain', 'meningkat', 'penopang', 'buat', 'timnasnya']
<i>Stopword Removal</i>	['local', 'pride', 'membawa', 'sepakbola', 'indonesia', 'bermarwah', 'klub', 'jleague', 'local', 'pride', 'membina', 'pemain', 'lokal', 'diekspor', 'abroad', 'kualitas', 'pemain', 'meningkat', 'penopang', 'timnasnya']
<i>Stemming</i>	local pride bawa sepakbola indonesia bermarwah klub jleague local pride bina main lokal ekspor abroad kualitas main tingkat topang timnas

3.3 Pelabelan

Pelabelan data dikerjakan dengan manual, dengan bantuan ahli untuk memastikan akurasi dalam menentukan sentimen (positif, negatif, atau netral) dari setiap tweet. Pakar yang dilibatkan memiliki pemahaman kontekstual yang baik terhadap isi komentar, sehingga dapat memberikan penilaian yang objektif dan tepat.. Hasil pelabelan ini digunakan sebagai data latih dalam proses klasifikasi menggunakan algoritma SVM. Hasil pelabelan ditampilkan pada Gambar 2 dan Tabel 2.



Gambar 2. Hasil Pelabelan

Gambar 2 menunjukkan hasil analisis sentimen dari sejumlah tweet. Mayoritas tweet bersentimen positif sebanyak 1274, diikuti oleh negatif sebanyak 640, dan sisanya netral sejumlah 418. Ini menunjukkan bahwa respons pengguna secara umum cenderung positif.

Tabel 2. Hasil Pelabelan

No	Teks	Hasil Pelabelan
1	local pride bawa sepakbola indonesia bermarwah klub jleague local pride bina main lokal ekspor abroad kualitas main tingkat topang timnasnya	Positif
2	komunikasi dasar kayak panggil bola atau kasih abaaba aja sering miss antara mereka	Negatif
3	kemaren bukannya indo semua ya dan gimana hasilnya dengan yang naturalisasi	Netral
...
2342	pemain lokal sesekali kasih arahan walau hanya sebatas isyarat kepala	Positif

Tabel 2 tersebut menampilkan contoh hasil pelabelan sentimen dari teks tweet. Setiap entri diklasifikasikan ke dalam kategori Positif, Netral, atau Negatif berdasarkan isi teks. Proses ini membantu memahami persepsi publik terhadap topik tertentu secara sistematis.

3.4 TF-IDF

Proses ini bertujuan untuk Menghitung bobot untuk kata tertentu berdasarkan seberapa sering kata tersebut muncul dalam sebuah dokumen. Metode ini menggunakan perbandingan antara frekuensi kata dalam dokumen (TF) dan seberapa umum kata itu muncul di seluruh kumpulan dokumen (IDF). Hasil TF-IDF dapat dilihat pada Tabel 3.

Tabel 3. Hasil TF-IDF

No	bola	komunikasi	indonesia	lokal	naturalisasi	pemain	pride
1	0.00	0.00	0.15	0.21	0.00	0.26	0.38
2	0.25	0.25	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.22	0.23	0.38	0.30	0.00
...
2342	0.00	0.00	0.00	0.24	0.00	0.28	0.00

Tabel 3 menunjukkan nilai TF-IDF untuk beberapa kata dalam dokumen. Nilai tinggi menandakan kata tersebut penting dalam dokumen tertentu. Misalnya, pride bernilai 0.38 di dokumen 1, menunjukkan kata itu relevan. Sebaliknya, bola bernilai 0.00, artinya tidak muncul atau tidak penting dalam konteks dokumen tersebut.

3.5 Pengelompokan Berdasarkan Variabel

Pengelompokan dilakukan berdasarkan variabel jenis pemain, yaitu pemain naturalisasi dan pemain lokal Timnas Indonesia. Pengelompokan memiliki tujuan untuk membandingkan sentimen masyarakat pada kedua variabel tersebut, setelah data diproses melalui algoritma SVM. Rincian selengkapnya terdapat pada tabel 4.

Tabel 4. Hasil Pengelompokan Variabel

Variabel	Positif	Netral	Negative
Naturalisasi	809	231	333
Lokal	465	187	317

Berdasarkan hasil analisis sentimen pada Tabel 4, pemain naturalisasi memperoleh sentimen positif sebesar 809, netral 231, dan negatif 333. Sementara itu, pemain lokal mendapatkan 465 sentimen positif, 187 netral, dan 317 negatif. Hal ini menunjukkan bahwa pemain naturalisasi cenderung mendapat tanggapan yang lebih positif dari publik dibandingkan pemain lokal, meskipun keduanya sama-sama menerima sentimen negatif dengan kuantitas yang cukup banyak.

3.5 Pengelompokan Berdasarkan Kategori Variabel

Pengelompokan dilakukan dengan menggunakan variabel kategori, yaitu jenis pemain, yang terbagi menjadi dua kelompok: pemain naturalisasi dan pemain lokal. Setelah sentimen data diberi label, hasil pelabelan tersebut akan dibandingkan antara kedua kategori ini untuk melihat perbedaan kecenderungan opini publik. Hasil dari pengelompokan berdasarkan kategori variabel dapat dilihat pada tabel 5.

Tabel 5. Hasil Pengelompokan Kategori Variabel

Kategori Variabel	Naturalisasi	Lokal
Performa Bermain	415(Positif:236,Negatif:95,Netral:84)	318(Positif:118,Negatif:126,Netral:74)
Pengalaman	441(Positif:273,Negatif:94,Netral:74)	194(Positif:101,Negatif:55,Netral:38)
Kondisi Fisik	144(Positif:108,Negatif:25,Netral:11)	156(Positif:88,Negatif:37,Netral: 31)
Adaptasi	213(Positif:105,Negatif:29,Netral:26)	117(Positif:50,Negatif:47,Netral: 20)
Komunikasi	160(Positif:87,Negatif:90,Netral: 36)	184(Positif:108,Negatif:52,Netral:24)

Tabel 5 menampilkan Lima kategori variabel utama digunakan dalam pengelompokan berdasarkan kategori variabel, yakni performa bermain, pengalaman, kondisi fisik, adaptasi, dan komunikasi, karena dianggap paling berpengaruh terhadap kontribusi dan penerimaan pemain di tim nasional. Masing-masing mewakili aspek teknis, kematangan, kesiapan fisik, penyesuaian strategi, dan efektivitas kerja sama tim.

3.6 Split Data

Studi ini menerapkan algoritma Support Vector Machine (SVM) untuk proses klasifikasi. Algoritma ini dikenal mampu menangani data berlabel secara efektif dan bekerja baik pada data berdimensi tinggi. Dalam penerapannya, digunakan metode Train-Test Split dengan proporsi 80% data untuk pelatihan dan 20% untuk pengujian. Pendekatan ini bertujuan untuk menguji performa model secara objektif dengan membandingkan hasil prediksi terhadap data yang belum pernah dilatih sebelumnya.

3.7 Penerapan SMOTE

Untuk mengatasi ketimpangan jumlah data antar kelas, digunakan pendekatan Random Oversampling dengan memperbanyak data pada kelas yang jumlahnya lebih sedikit. Dengan metode ini, distribusi data antar kelas menjadi lebih merata, sehingga dapat meningkatkan akurasi prediksi. Hasil sebelum random over sampling dan setelah random over sampling dapat dilihat pada Tabel 6 dan tabel 7.

Tabel 6. Hasil Non SMOTE

Kategori Variabel	Variabel	Positif	Negatif	Netral
Performa Bermain	Naturalisasi	236	95	84
	Lokal	118	126	74
Pengalaman	Naturalisasi	273	94	74
	Lokal	101	55	38
Kondisi Fisik	Naturalisasi	108	25	11
	Lokal	88	37	31
Adaptasi	Naturalisasi	105	29	26
	Lokal	50	47	20
Komunikasi	Naturalisasi	87	90	36
	Lokal	108	52	24

Pada Tabel 6 tersebut menunjukkan distribusi data sebelum dilakukan SMOTE atau Random Over Sampling berdasarkan lima kategori variabel, yaitu Performa Bermain, Pengalaman, Kondisi Fisik, Adaptasi, dan Komunikasi. Setiap kategori dibedakan berdasarkan status pemain (Naturalisasi dan Lokal) serta klasifikasi sentimen (Positif, Negatif, Netral). Terlihat bahwa pemain naturalisasi cenderung memiliki jumlah data positif yang lebih tinggi dibandingkan pemain lokal di sebagian besar kategori.

Tabel 7. Hasil dengan SMOTE

Kategori Variabel	Variabel	Positif	Negatif	Netral
Performa Bermain	Naturalisasi	236	236	236
	Lokal	126	126	126
Pengalaman	Naturalisasi	273	273	273
	Lokal	101	101	101
Kondisi Fisik	Naturalisasi	108	108	108
	Lokal	88	88	88
Adaptasi	Naturalisasi	105	105	105
	Lokal	50	50	50
Komunikasi	Naturalisasi	90	90	90
	Lokal	108	108	108

Tabel 7 menunjukkan hasil setelah dilakukan proses SMOTE (Synthetic Minority Over-sampling Technique) atau Random Over Sampling pada data sentimen yang dikelompokkan berdasarkan kategori variabel. Pada tahap ini, jumlah data untuk setiap kelas sentimen Positif, Negatif, dan Netral telah disamakan, baik untuk tipe pemain Naturalisasi maupun Lokal. Tujuan utama dari teknik oversampling ini adalah untuk mengatasi ketidakseimbangan jumlah data antar kelas sentimen yang dapat memengaruhi performa model klasifikasi. Dengan data yang seimbang, proses analisis dan pelatihan model menjadi lebih adil dan hasilnya pun cenderung lebih akurat serta representatif.

3.8 Klasifikasi Algoritma Support Vector Machine

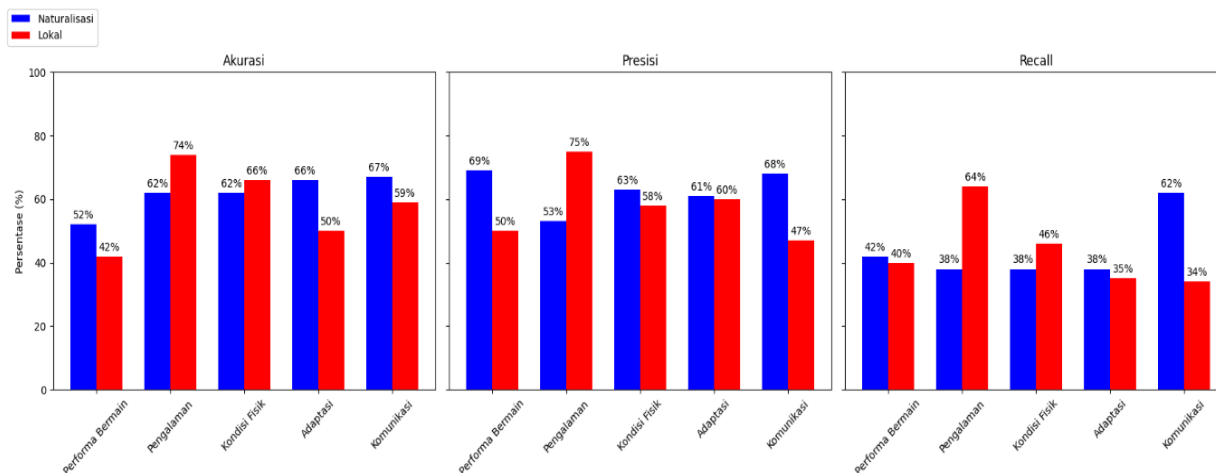
Setelah data melalui proses prapengolahan dan pelabelan, tahap berikutnya adalah melakukan proses klasifikasi dengan memanfaatkan algoritma Support Vector Machine (SVM). Algoritma ini dipilih karena dikenal efektif dalam mengolah data berlabel dan memiliki performa yang baik dalam membedakan kelas secara jelas, terutama pada data

dengan dimensi tinggi. Hasil dari klasifikasi algoritma SVM untuk pemain Naturalisasi dan Lokal sebelum dan setelah Random Over Sampling dapat dilihat pada Gambar 3, Gambar 4, Tabel 8 dan Tabel 9.

Tabel 8. Hasil Klasifikasi Algoritma SVM Naturalisasi dan Lokal Non SMOTE

Variabel	Kategori Variabel	Akurasi (%)	Presisi (%)	Recall (%)
Naturalisasi	Perform Bermain	52	69	42
	Pengalaman	62	53	40
	Kondisi Fisik	66	63	38
	Adaptasi	66	61	38
	Komunikasi	67	68	62
Lokal	Perform Bermain	42	50	40
	Pengalaman	74	75	64
	Kondisi Fisik	62	58	46
	Adaptasi	59	60	35
	Komunikasi	59	47	34

Tabel 8 memperlihatkan kinerja algoritma Support Vector Machine (SVM) dalam mengelompokkan sentimen tanpa dilakukan penyeimbangan data sebelumnya (tanpa SMOTE). Proses klasifikasi dilakukan berdasarkan sejumlah variabel yang dikaitkan dengan tipe pemain, yakni Naturalisasi dan Lokal. Evaluasi dilakukan dengan menggunakan tiga ukuran utama, yaitu akurasi, presisi, dan recall, untuk menilai sejauh mana model mampu mengenali pola sentimen secara tepat. Hasil yang ditampilkan menunjukkan bahwa model memberikan hasil yang lebih baik saat diterapkan pada data pemain lokal, terutama pada variabel yang berkaitan dengan pengalaman dan kemampuan beradaptasi, yang menunjukkan bahwa data dari kelompok ini lebih mudah dipelajari oleh model.



Gambar 3. Hasil Klasifikasi Algoritma SVM Naturalisasi dan Lokal Non SMOTE

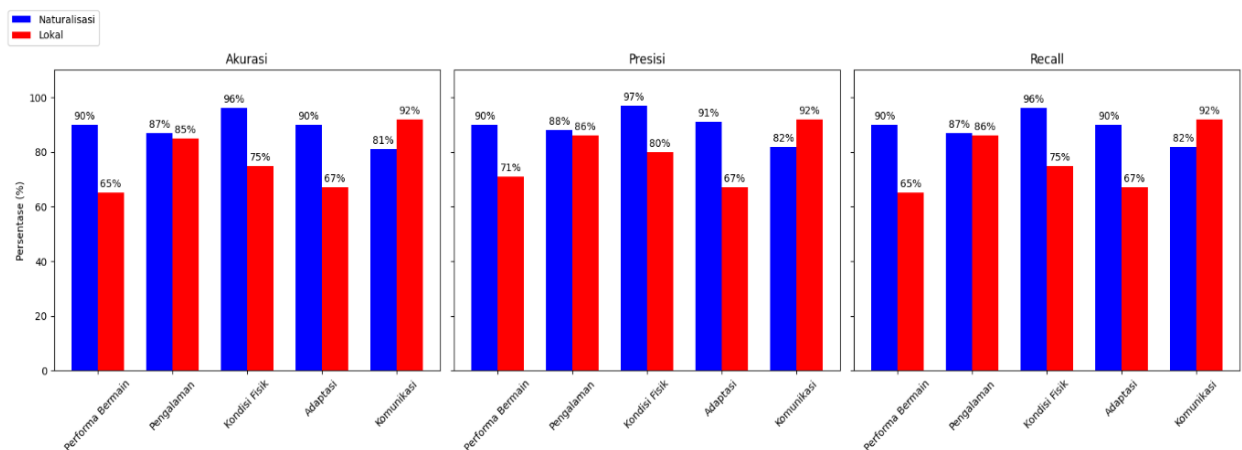
Gambar 3 tersebut menunjukkan hasil klasifikasi menggunakan algoritma Support Vector Machine (SVM) sebelum mengalami proses Random Over Sampling, terdapat lima kategori variabel, yaitu Performa Bermain, Pengalaman, Kondisi Fisik, Adaptasi, dan Komunikasi, dengan perbandingan dua kelompok data: Naturalisasi (biru) dan Lokal (merah), berdasarkan tiga metrik evaluasi: Akurasi, Presisi, dan Recall. Secara umum, data Naturalisasi menunjukkan performa yang lebih tinggi pada sebagian besar kategori. Pada metrik Akurasi, Naturalisasi lebih unggul dalam empat dari lima kategori, terutama di Adaptasi (66%) dan Komunikasi (67%), sedangkan Lokal hanya unggul di Pengalaman (74%). Pada metrik Presisi, pemain Lokal unggul dalam Pengalaman (75%), namun Naturalisasi tetap lebih baik dalam aspek lainnya seperti Performa Bermain (69%) dan Komunikasi (68%). Untuk metrik Recall, Naturalisasi menunjukkan dominasi pada kategori seperti Komunikasi (62%) dan Performa Bermain (42%), meskipun selisihnya tipis dibandingkan Lokal. Secara keseluruhan, hasil ini mengindikasikan bahwa pemain Naturalisasi cenderung memiliki keunggulan dalam aspek teknis seperti performa dan fisik, sedangkan pemain Lokal menunjukkan kekuatan pada aspek pengalaman.

Tabel 9. Hasil Klasifikasi Algoritma SVM Naturalisasi dan Lokal dengan SMOTE

Variabel	Kategori Variabel	Akurasi (%)	Presisi (%)	Recall (%)
Naturalisasi	Perform Bermain	90	90	90
	Pengalaman	87	88	87
	Kondisi Fisik	96	97	96
	Adaptasi	75	80	75
	Komunikasi	90	91	90

Variabel	Kategori Variabel	Akurasi (%)	Presisi (%)	Recall (%)
Lokal	Perform Bermain	65	71	65
	Pengalaman	85	86	86
	Kondisi Fisik	90	80	90
	Adaptasi	67	67	67
	Komunikasi	92	92	92

Tabel 9 menampilkan hasil evaluasi model klasifikasi SVM setelah dilakukan proses penyeimbangan data menggunakan metode SMOTE. Penilaian dilakukan terhadap dua kelompok pemain, yaitu Naturalisasi dan Lokal, berdasarkan lima kategori variabel yang berbeda. Setiap variabel dianalisis dengan tiga metrik utama: akurasi, presisi, dan recall. Secara umum, penerapan SMOTE memberikan peningkatan signifikan terhadap performa model di seluruh kategori. Untuk pemain Naturalisasi, kategori Kondisi Fisik dan Komunikasi mencatat skor tertinggi, mencapai nilai 90% ke atas pada semua metrik. Sementara itu, pemain Lokal juga menunjukkan hasil yang kuat, terutama pada aspek Komunikasi dan Kondisi Fisik, yang mendapatkan nilai konsisten tinggi pada ketiga metrik evaluasi. Hal ini menunjukkan bahwa setelah dilakukan oversampling, model dapat belajar secara lebih seimbang dari masing-masing kelas sentimen dan memberikan hasil prediksi yang lebih stabil dan akurat.



Gambar 4. Hasil Klasifikasi Algoritma SVM Naturalisasi dan Lokal dengan SMOTE

Gambar 4 tersebut menunjukkan hasil klasifikasi menggunakan algoritma Support Vector Machine (SVM) setelah melakukan proses Random Over Sampling, terdapat lima kategori variabel, yaitu Performa Bermain, Pengalaman, Kondisi Fisik, Adaptasi, dan Komunikasi, dengan perbandingan dua kelompok data: Naturalisasi (biru) dan Lokal (merah), berdasarkan tiga metrik evaluasi: Akurasi, Presisi, dan Recall. Secara umum, data Naturalisasi menunjukkan performa yang lebih tinggi pada hampir semua kategori. Pada metrik Akurasi, Naturalisasi unggul terutama di Kondisi Fisik (96%) dibandingkan Lokal (75%), sedangkan satu-satunya keunggulan Lokal terdapat pada Komunikasi (92%, sedangkan Naturalisasi hanya 81%). Hal serupa juga tampak pada Presisi, di mana Naturalisasi kembali lebih unggul di Kondisi Fisik (97%) dan Adaptasi (91%), sementara Lokal hanya lebih tinggi pada Komunikasi (92% dibanding 82%). Pada Recall, Naturalisasi juga dominan di sebagian besar kategori, seperti Kondisi Fisik (96%), Adaptasi (90%), dan Performa Bermain (90%), sedangkan Lokal tetap unggul hanya di Komunikasi (92% berbanding 82%). Keseluruhan hasil mengindikasikan bahwa data Naturalisasi memberikan performa klasifikasi yang lebih baik dalam aspek fisik dan permainan, sedangkan data Lokal lebih kuat dalam aspek komunikasi.

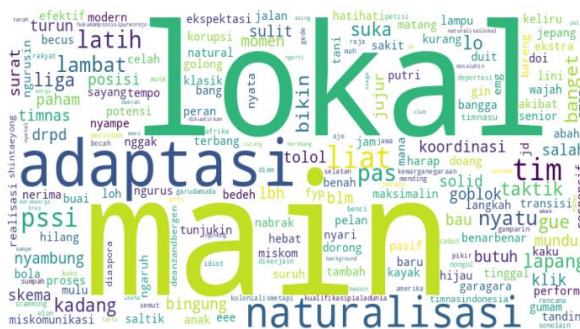
3.9 Visualisasi WordCloud

Visualisasi WordCloud dilakukan untuk menampilkan frekuensi kemunculan kata pada sentiment positif, negative, dan netral. Hasil Wordcloud ditampilkan di Gambar 5, Gambar 6, dan Gambar 7.



Gambar 5. Wordcloud Sentimen Positif

Gambar 5 menunjukkan Wordcloud positif yang merepresentasikan dukungan publik terhadap pemain naturalisasi dan lokal di timnas Indonesia. Kata-kata yang paling sering muncul dan berukuran besar antara lain "Indonesia", "sepakbola", "lokal", "program", "baik", "kualitas", dan "bagus". Ini menunjukkan bahwa topik-topik tersebut adalah fokus utama dari teks yang dianalisis.



Gambar 6. Wordcloud Sentimen Negatif

Gambar 6 menunjukkan Wordcloud negatif. Kata-kata yang paling sering muncul dalam suatu kumpulan teks. Kata-kata yang lebih besar seperti "lokal" dan "main" mengindikasikan frekuensi kemunculan yang tinggi, sementara kata-kata lain seperti "adaptasi", "naturalisasi", "miskomunikasi", dan "lambat" juga menonjol.



Gambar 7. Wordcloud Sentimen Netral

Gambar 7 menunjukkan Wordcloud netral yang mencerminkan kritik publik terhadap program naturalisasi dan pemain lokal timnas Indonesia. Dalam *wordcloud* ini, kata-kata yang paling menonjol (berukuran besar) adalah "lokal", "main", "adaptasi", dan "naturalisasi". Kata-kata lain yang juga terlihat cukup besar antara lain "liga", "tim", "fisik", dan "sepakbola". Secara umum, Wordcloud ini merefleksikan kekhawatiran atas kualitas dan arah perkembangan sepak bola nasional.

4. KESIMPULAN

Berdasarkan hasil penelitian terhadap data tweet mengenai pemain naturalisasi dan lokal di tim nasional sepak bola Indonesia, diperoleh bahwa pemain naturalisasi mendapatkan sentimen positif sebesar 809, netral 231, dan negatif 333, sedangkan pemain lokal memperoleh 465 sentimen positif, 187 netral, dan 317 negatif. Hal ini menunjukkan bahwa sentimen publik terhadap pemain naturalisasi cenderung lebih positif dibandingkan pemain lokal. Pengujian menggunakan algoritma Support Vector Machine (SVM) menunjukkan kelebihan pada kategori kondisi fisik pemain naturalisasi dengan akurasi 96%, presisi 97%, dan recall 96%. Sementara itu, kategori komunikasi pada pemain lokal memperoleh akurasi, presisi, dan recall sebesar 92%. Namun, terdapat kekurangan pada kategori performa bermain pemain lokal yang hanya mencapai akurasi dan recall 65%. Penelitian ini unggul dalam pemanfaatan machine learning, khususnya algoritma SVM, serta teknik random oversampling dan SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi ketidakseimbangan data, yang secara signifikan membantu meningkatkan kualitas analisis model klasifikasi. Selain itu, pelabelan data secara manual oleh pakar turut memperkuat validitas hasil. Keterbatasan penelitian ini terletak pada sumber data yang hanya berasal dari satu jejaring sosial, serta masih adanya kategori dengan akurasi yang rendah. Untuk pengembangan selanjutnya, disarankan agar penelitian diperluas dengan menambah sumber data dari platform lain serta menguji algoritma klasifikasi lainnya guna meningkatkan akurasi pada kategori yang masih rendah.

REFERENCES

[1] M. Graciela and dan Hafiz Irsyad, "Klasifikasi Opini Masyarakat Terhadap Naturalisasi Pemain Sepak Bola Menggunakan KNN dan SMOTE," Applied Information Technology and Computer Science (AICOMS), Vol 3, No 1, Juni 2024, doi: <https://doi.org/10.58466/aicoms.v3i1.1547>.



- [2] N. M. Sania, N. Baitillah, M. H. Indriani, F. Fernanda, and T. Aditya, “Survei Kepuasan Opini Publik terhadap Kebijakan Naturalisasi Pemain PSSI: dalam Upaya Meningkatkan Prestasi Timnas Indonesia,” *Indonesian Journal of Public Administration Review*, vol. 2, no. 2, p. 19, Jan. 2025, doi: 10.47134/par.v2i2.3564.
- [3] D. Rizky, P. Jaya, and S. Lestari, “Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors,” *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK)*, Vol 5, No 3, 2024, doi: <https://doi.org/10.35870/jimik.v5i3.1024>.
- [4] M. P. Silitonga and D. Kristina, “Naturalisasi Pemain Asing Dalam Hukum Konstitusi Indonesia,” *Honeste Vivere*, vol. 35, no. 1, pp. 129–137, Jan. 2025, doi: 10.55809/hv.v35i1.429.
- [5] H. Sinaga, T. Br Sembiring, and I. Artikel, “Perlakuan Penghinaan Pemain Naturalisasi Maupun Lokal Sepak Bola Indonesia Di Media Sosial Dan Hukum Yang Berlaku,” *Journal of International Multidisciplinary Research*, vol. 2, no. 1, 2024
- [6] S. Ramdani and T. Bustomi, “Analisis Sentimen Terhadap Pemain Diaspora Timnas Indonesia Pada Media Sosial Instagram Menggunakan Metode Support Vector Machine (SVM),” *Sebatik*, 2024, doi: 10.46984/sebatik.v28i2.0000.
- [7] R. Asrianto and M. Herwinanda, “Analisis Sentimen Kenaikan Harga Kebutuhan Pokok Dimedia Sosial Youtube Menggunakan Algoritma Support Vector Machine,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 431–440, Dec. 2022, doi: 10.37859/coscitech.v3i3.4368.
- [8] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, 2021, doi: <https://doi.org/10.30591/jpit.v6i3.2870>.
- [9] M. Cakal Cendekia, G. Yulisatria, Y. H. Bhakti, M. Suhariyanti, and M. F. Dongoran, “Strategi Pelatihan yang Efektif untuk Pengembangan Pemain Muda Sepakbola: Sebuah Tinjauan Literatur,” *Jurnal Ilmiah STOK Bina Guna Medan (JISBG)*, vol. 13, no 1, 2025, doi: <https://doi.org/10.55081/jsbg.v13i1.3668>.
- [10] D. Musfiroh *et al.*, “Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, 2021, doi: <https://doi.org/10.57152/malcom.v1i1.20>.
- [11] R. T. Aldisa and P. Maulana, “Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 106–109, Jun. 2022, doi: 10.47065/bits.v4i1.1581.
- [12] E. A. Lisangan *et al.*, “Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia,” *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol 2, no 1, 2022, doi: <https://doi.org/10.24002/konstelasi.v2i1.5609>.
- [13] A. Rachmadana Ismail, R. Bagus, and F. Hakim, “Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Mengetahui Trend Wisata Pantai Di DI Yogyakarta Berdasarkan Data Twitter,” *Emerging Statistics and Data Science Journal (ESDS)*, vol 11, no 1, 2023, doi: <https://doi.org/10.20885/esds.vol11.iss.1.art5>.
- [14] D. Farah Zhafira, and B. Rahayudi, “Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada YouTube,” *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, vol 2, no 1, 2021, doi: <https://doi.org/10.25126/justsi.v2i1.24>.
- [15] O. : Rifqy, M. Turjaman, and I. Budi, “Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompot Digital (Studi Kasus: Aplikasi LinkAja pada Twitter),” *Jurnal Darma Agung*, Vol 30, No 2, 2022, doi: <http://dx.doi.org/10.46930/ojsuda.v30i2.1672>.
- [16] R. A. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, and L. Efrizoni, “Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 211–218, Oct. 2023, doi: 10.57152/malcom.v3i2.901.
- [17] O. N. Cahyani and F. Budiman, “Performa Logistic Regression dan Naive Bayes dalam Klasifikasi Berita Hoax di Indonesia,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 60–68, 2025, doi: 10.29408/edumatic.v9i1.28987.
- [18] D. Atika, A. Ari Aldino, “Term Frequency-Inverse Document Frequency Support Vector Machine untuk Analisis Sentimen Opini Masyarakat terhadap Tekanan Mental pada Media Sosial Twitter,” *JURNAL Teknologi dan Sistem Informasi (JTISI)*, 2022. doi: <https://doi.org/10.33365/jtsi.v3i4.2054>.
- [19] P. M. Susanti, M. Afdal, I. Permana, and A. Marsal, “Klasifikasi Sentimen Pengguna X Terhadap Pemboikotan Produk Pro Israel Menggunakan Algoritma Machine Learning,” *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6533.
- [20] V. Fitriyana *et al.*, “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine,” *Jurnal Buana Informatika Jurnal Buana Informatika*, Vol 14, No 1, 2023. doi: <https://doi.org/10.24002/jbi.v14i01.6909>.