

Evaluasi Kinerja Algoritma *Support Vector Machine* (SVM) Dalam Analisis Sentimen Publik Terhadap Naturalisasi Timnas Indonesia di Twitter

Muhammad Ghafur Rahman Lubis¹ Dicky Sambora Sitompul² Teuku Muhammad Giovanni³ Fanny Ramadhani⁴ Sri Dewi⁵

Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Kota Medan, Provinsi Sumatera Utara, Indonesia^{1,2,3,4,5}

Email: rahmanghafur@gmail.com¹ dickysambora.xiimia1@gmail.com²
teukumgio@gmail.com³ fannyr@unimed.ac.id⁴ sridewi@unimed.ac.id⁵

Abstrak

Studi ini bertujuan untuk menganalisis efektivitas algoritma Support Vector Machine (SVM) dalam menilai sentimen masyarakat terkait kebijakan naturalisasi pemain tim nasional sepak bola Indonesia yang dibicarakan di Twitter. Mengingat meningkatnya perhatian terhadap isu naturalisasi, pemahaman tentang reaksi publik menjadi sangat penting. Data diperoleh melalui teknik crawling dari Twitter, menghasilkan kumpulan data yang mencakup beragam opini dan tanggapan publik. Setelah melakukan praproses data, termasuk penghilangan kata yang tidak penting dan stemming, model SVM diterapkan untuk mengkategorikan sentimen menjadi positif, negatif, atau netral. Kinerja model dievaluasi dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil dari penelitian ini menunjukkan bahwa algoritma SVM mampu mencapai tingkat akurasi yang memuaskan dalam analisis sentimen publik, serta memberikan wawasan yang berguna mengenai pandangan masyarakat terhadap naturalisasi Timnas Indonesia. Temuan ini diharapkan dapat memberikan bantuan kepada pembuat kebijakan dan pemangku kepentingan lainnya dalam memahami dinamika opini publik seputar isu-isu penting dalam dunia olahraga nasional.

Kata Kunci: Analisis, Sentimen, Twitter, Naturalisasi Timnas Indonesia, SVM

Abstract

This study aims to analyze the effectiveness of the Support Vector Machine (SVM) algorithm in assessing public sentiment regarding the naturalization policy of players in the Indonesian national football team discussed on Twitter. Given the increasing attention to the issue of naturalization, understanding public reactions becomes crucial. Data was obtained through crawling techniques from Twitter, resulting in a dataset that encompasses a variety of opinions and public responses. After preprocessing the data, including the removal of stop words and stemming, the SVM model was applied to classify sentiment as positive, negative, or neutral. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. The results of this study indicate that the SVM algorithm can achieve a satisfactory level of accuracy in analyzing public sentiment, providing valuable insights into public perceptions regarding the naturalization of the Indonesian national team. These findings are expected to assist policymakers and other stakeholders in understanding the dynamics of public opinion surrounding important issues in national sports.

Keywords: Analysis, Sentiment, Twitter, Naturalization of the Indonesian National Team, SVM



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Kemajuan teknologi yang sangat pesat, membuat hampir semua kegiatan yang kita lakukan sangat dipengaruhi dari teknologi yang tersedia. Hal ini dapat dilihat dari seberapa ketergantungannya kita terhadap teknologi yang tersedia ini. Dengan itu, kemajuan ini harus dapat dipergunakan atau dimanfaatkan semaksimal mungkin (Salman, et al, 2023). Indihome merupakan operator penyedia ISP yang paling banyak digunakan oleh masyarakat Indonesia, dimana sebanyak 8,7% memilih Indihome sebagai operator penyedia ISP yang mereka

gunakan. Persentase tersebut lebih banyak dari pada persentase operator penyedia ISP lainnya. Melalui Twitter, pengguna dapat saling berdiskusi mengenai kritik, saran ataupun kepuasan mereka terhadap layanan Indihome. Perbedaan Twitter dari media sosial yang lain adalah Twitter memungkinkan tiap pengguna dapat melacak *tweet* pengguna lain tanpa persetujuan dari pengguna tersebut. Pertumbuhan media sosial yang sangat cepat dalam beberapa tahun terakhir telah menjadikan internet sebagai platform yang layak untuk memberikan sebuah pengetahuan. Penggunaan big-data dari media sosial dapat digunakan untuk prediksi kejadian di dunia nyata (Sekioka, et al, 2023). Dalam beberapa tahun terakhir, kemajuan dalam teknologi informasi dan komunikasi telah menyebabkan pertumbuhan pesat platform media sosial, dengan Twitter menjadi salah satu yang paling banyak digunakan di Indonesia. Indonesia menjadi salah satu pengguna Twitter terbanyak dengan 59% masyarakatnya menggunakan Twitter dan menjadi platform media sosial yang cukup berpengaruh di Indonesia. Hal ini disebabkan tidak lain karna kemudahan untuk menyampaikan opini pengguna dalam membicarakan suatu topik masalah. (Utomo, et al, 2023).

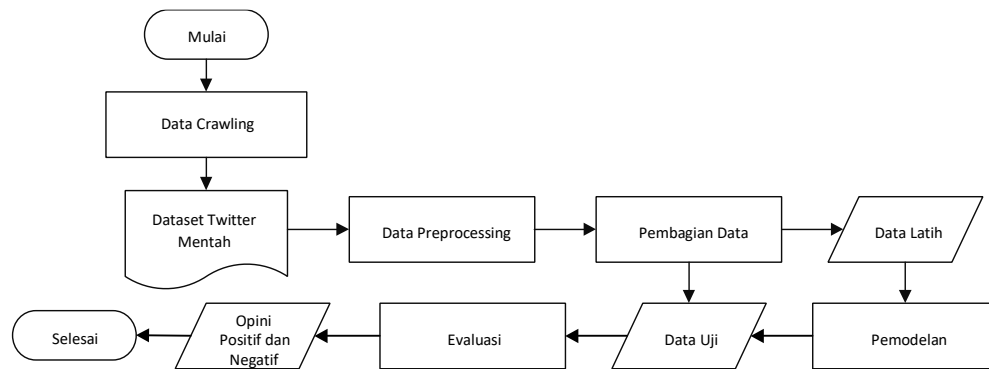
Dengan jutaan pengguna aktif, Twitter berfungsi sebagai saluran bagi individu untuk menyampaikan pendapat, membagikan informasi, dan terlibat dalam diskusi tentang berbagai isu, termasuk olahraga. Salah satu topik yang menjadi sorotan adalah proses naturalisasi pemain dalam tim nasional sepak bola Indonesia. Kebijakan ini, yang melibatkan pemberian kewarganegaraan kepada pemain asing, seringkali menimbulkan reaksi beragam di kalangan penggemar dan masyarakat secara umum. Analisis sentimen merupakan suatu istilah yang biasa digunakan untuk memberikan gambaran topik secara faktual atau non faktual dengan opini positif ataupun negatif (Wongkar dan Angdresey, 2019). Menganalisis sentimen publik mengenai naturalisasi timnas Indonesia sangat penting untuk memahami seberapa besar penerimaan masyarakat terhadap kebijakan tersebut. Oleh karena itu, analisis sentimen menjadi alat yang sangat berharga untuk menggali reaksi masyarakat terhadap isu ini.

Analisis ini menggunakan suatu pendekatan dalam menganalisis teks untuk mengetahui subjektivitas pendapat/opini postingan tweet (Agustina and Ihsan, 2023). Karena masyarakat cenderung banyak memberi opini berupa tanggapan atau kritik terhadap suatu situasi melalui media sosial (Tri Saputra et al., 2021). Analisis sentimen memungkinkan untuk membaca dan memahami emosi orang dan memberikan kesimpulan berdasarkan hasil deteksi bahasa postingan tweet. Dalam konteks ini, algoritma pembelajaran mesin seperti Support Vector Machine (SVM) dapat digunakan untuk mengklasifikasikan sentimen dari data yang diperoleh dari Twitter. SVM dikenal efektif dalam menangani dataset yang besar dan kompleks serta unggul dalam klasifikasi. Namun, kinerja SVM dalam analisis sentimen di Twitter, khususnya mengenai topik naturalisasi tim nasional, belum banyak diteliti. Oleh karena itu, penelitian ini bertujuan untuk menilai sejauh mana algoritma SVM dapat digunakan dalam menganalisis sentimen publik terhadap naturalisasi timnas Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang pandangan masyarakat serta membantu pengambil kebijakan dalam merumuskan langkah-langkah yang tepat terkait kebijakan naturalisasi.

METODE PENELITIAN

Diagram Alur Proses Penelitian

Dalam penelitian ini, digunakan metode klasifikasi pembelajaran mesin yang dikenal sebagai Support Vector Machine (SVM). Meskipun terdapat banyak metode klasifikasi lainnya, penelitian ini akan difokuskan hanya pada SVM. Dalam melakukan penelitian ini diperlukan alur proses penelitian, agar dalam proses penelitian berjalan sesuai dengan rencana dan mencapai hasil yang optimal. Gambar 1 menunjukkan diagram alur proses penelitian dalam penelitian ini.



Gambar 1. Diagram Alur Proses Penelitian

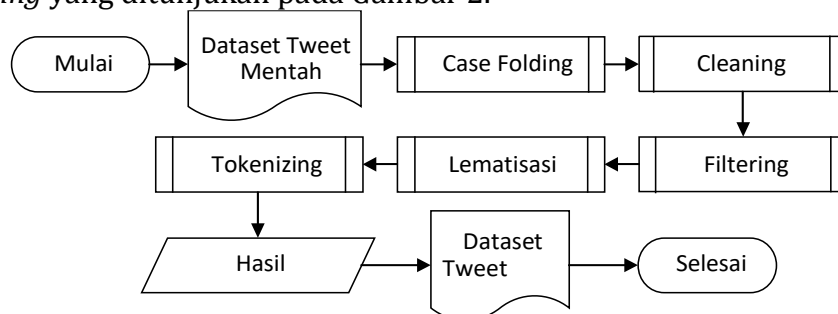
Pada Gambar 1. Terdapat beberapa tahap dalam penelitian kali ini diantaranya tahap pengumpulan data, *text preprocessing*, vektorisasi, pembagian data, pemodelan metode klasifikasi, dan terakhir tahap evaluasi.

Pengumpulan Data

Penelitian ini memanfaatkan dataset yang berisi informasi dari tweet yang diambil melalui crawling data di platform Twitter, dengan bantuan API Twitter. API ini didapatkan dengan membuat akun developer, yang menyediakan API Key serta Access Token unik. Dataset yang digunakan berjumlah 100 tweet, dengan kata kunci "Naturalisasi Timnas Indonesia" yang digunakan untuk mengumpulkan data. Pelabelan data dilakukan secara manual dan divalidasi. Tiga label yang digunakan dalam klasifikasi adalah negatif, positif, dan netral. Penelitian ini menggunakan berbagai pustaka Python seperti pandas, numpy, nltk, matplotlib.

Text Preprocessing

Dataset yang dikumpulkan masih berupa data tidak terstruktur, sehingga diperlukan preprocessing agar data lebih terstruktur. Beberapa tahapan preprocessing yang dilakukan dalam penelitian ini meliputi *case folding*, *cleaning*, *filtering*, lemmatisasi, dan *tokenizing*. Setelah data bersih maka dilakukan proses pelabelan secara manual. Berikut ini adalah diagram alur *text preprocessing* yang ditunjukkan pada Gambar 2.



Gambar 2. Diagram Alur Proses *Text Preprocessing*

Vektorisasi

Pada penelitian ini algoritma yang diterapkan adalah pembobotan kata atau TF-IDF (*Term Frequency-Inverse Document Frequency*) yang berguna dalam pembobotan sebagai upaya untuk memberi nilai pada kata (*term*) berdasarkan suatu dokumen/*tweet*. Term Frequency – Inverse Document Frequency (TF-IDF) adalah model yang digunakan untuk menemukan hubungan antara term dengan data, dengan cara memberikan bobot pada setiap kata atau term. Perhitungan nilai TF dilakukan secara manual menggunakan persamaan (1):

$$TF(k, d) = \frac{F(k, d)}{\sum_j F(j, d)} \quad (1)$$

Keterangan :

$F(k, d)$ = frekuensi kata k dalam dokumen d

$\sum_j F(j, d)$ = total jumlah kata dalam dokumen d

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah teknik pembelajaran mesin yang digunakan untuk tugas-tugas prediksi, baik dalam klasifikasi maupun regresi. Tujuan utama dari SVM adalah menemukan *hyperplane* yang memisahkan data ke dalam kelas yang berbeda dengan margin terbesar. Dalam kasus regresi, SVM mencoba menemukan garis atau kurva yang paling sesuai untuk memprediksi nilai target. Persamaan dasar untuk fungsi keputusan SVM adalah sebagai berikut:

$$f(x) = \sum_{i=1}^M a_i y_i K(x, x_i) + b \quad (2)$$

Keterangan:

a_i = koefisien Lagrange untuk data latih ke-i, diperoleh dari proses pelatihan

y_i = label atau kelas dari data latih ke-i

M = jumlah data latih

$K(x, x_i)$ = fungsi kernel yang mengukur kesamaan antara data uji x dan data latih x_i

x = data uji

x_i = data latih ke-i

b = bias, digunakan untuk menggeser garis keputusan

Pembagian Data

Data dibagi menjadi dua kelompok utama, yaitu data latih dan data uji. Terdapat tiga skenario pemisahan data yang dilakukan, yaitu:

1. Pada percobaan pertama, data latih terdiri dari 70% dan data uji 30%.
2. Pada percobaan kedua, data latih terdiri dari 80% dan data uji 20%.
3. Pada percobaan ketiga, data latih terdiri dari 90% dan data uji 10%.

Data Latih

Data latih digunakan untuk melatih model klasifikasi SVM dalam penelitian ini. Data tersebut membantu metode SVM dalam mempelajari cara mengklasifikasikan komentar ke dalam kategori negatif, positif, atau netral.

Data Uji

Setelah model klasifikasi selesai dilatih, langkah berikutnya adalah mengevaluasi kinerjanya menggunakan data uji. Tujuan dari pengujian ini adalah untuk melihat bagaimana metode SVM mampu mengklasifikasikan data baru dengan benar, mengelompokkan komentar menjadi negatif, netral, atau positif sesuai hasil analisis.

Klasifikasi Algoritma

Penelitian ini menggunakan algoritma klasifikasi Support Vector Machine (SVM). SVM, yang berakar dari teori pembelajaran statistik, dianggap mampu memberikan hasil yang lebih unggul dibandingkan dengan metode lainnya. Oleh sebab itu, SVM dipilih untuk mengelompokkan label yang telah diperoleh melalui proses pembobotan dan pelabelan. Pengelompokan dilakukan ke dalam tiga kategori: positif, negatif dan netral.

Evaluasi

Pengujian model diperlukan untuk menilai performa metode Support Vector Machine (SVM). Dalam penelitian ini, evaluasi dilakukan berdasarkan nilai akurasi. Akurasi merupakan metrik evaluasi yang mengukur sejauh mana model klasifikasi atau prediksi dapat menghasilkan hasil yang akurat sesuai dengan data yang tersedia. Nilai akurasi menggambarkan persentase keberhasilan model dalam memprediksi kelas atau label yang tepat, serta menunjukkan seberapa efektif model dalam mengklasifikasikan data secara keseluruhan. Dengan kata lain, akurasi mencerminkan seberapa baik model dapat memprediksi kelas yang benar (3).

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total jumlah data}} \quad (3)$$

HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kumpulan tweet di platform media sosial X untuk naturalisasi timnas senior Indonesia. Tiga kata kunci yang sedang tren akan digunakan sebagai acuan untuk membandingkan respon masyarakat terhadap isu-isu yang berkembang di media sosial. Penelitian ini ditujukan untuk melihat seberapa besar warga negara luar yang ingin membela kancah sepak bola tanah air tingkat nasional. Data yang digunakan dalam penelitian ini dikumpulkan dari media sosial Twitter dengan metode scrapping tweet berbahasa Indonesia menggunakan Google Collabs, dengan kata kunci "Naturalisasi Timnas Indonesia". Data diambil pada bulan Oktober 2024, mencakup periode dari 3 Desember 2020 hingga 3 Juni 2024. Preprocessing yang dilakukan pada penelitian ini adalah untuk mempersiapkan data agar sesuai dengan kebutuhan analisis. Adapun tahapan preprocessing adalah meliputi normalisasi kata, pembersihan teks, case folding, tokenisasi, penghapusan stopword, dan stemming. Berikut ini akan dilakukan tahapan preprocessing:

Normalisasi Kata

Tabel 1. Sampel Hasil dari Normalisasi Kata

Tweet	Normalisasi
@Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain #TimnasIndonesia #TurnamenToulon2024 #Prancis https://t.co/vMpbAZWyCR	lima calon pemain naturalisasi sudah bergabung dengan timnas indonesia u20 di como italia

Cleaning Kata

Tabel 2. Sampel Hasil dari Cleaning Kata

Tweet	Cleaning
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain #TimnasIndonesia #TurnamenToulon2024 #Prancis https://t.co/vMpbAZWyCR	Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U20 di Como Italia

Case Folding

Tabel 3. Sampel Hasil dari Case Folding

Tweet	Case Folding
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain	lima calon pemain naturalisasi sudah bergabung dengan timnas indonesia

#TimnasIndonesia #TurnamenToulon2024 #Prancis
<https://t.co/vMpbAZWyCR>

u-20 di como italia #thephraseid
 #naturalisasipemain
 #timnasindonesia
 #turnamentoulon2024 #prancis
<https://t.co/vmpbazwyrcr>

Tokenizing

Tabel 4. Sampel Hasil dari Tokenizing

Tweet	Tokenizing
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain #TimnasIndonesia #TurnamenToulon2024 #Prancis https://t.co/vMpbAZWyCR	['lima','calon','pemain','naturalisasi','sudah','bergabung','dengan','Timnas','Indonesia','U-20','di','Como','Italia']

Stopword

Tabel 5. Sampel Hasil dari Stopword

Tweet	Stopword
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain #TimnasIndonesia #TurnamenToulon2024 #Prancis https://t.co/vMpbAZWyCR	Lima calon pemain naturalisasi bergabung Timnas Indonesia U-20 Como Italia

Stemming

Tabel 6. Sampel Hasil dari Stemming

Tweet	Stemming
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U-20 di Como Italia #ThePhraseID #NaturalisasiPemain #TimnasIndonesia #TurnamenToulon2024 #Prancis https://t.co/vMpbAZWyCR	lima calon pemain naturalisasi bergabung timna indonesia u20 como italia

Sentimen

Tabel 7. Sampel Sentimen

Translate	Sentimen
Lima calon pemain naturalisasi sudah bergabung dengan Timnas Indonesia U20 di Como Italia	Positif
Sepak bola indonesia masih primitif dan akan tetap primitif bubarin aja sepakbola indonesia sekalian federasinya timnas ambil aja 11 orang naturalisasi gausah pake pemain lokal sama sekali	Negatif
Emil Audero Mulyadi merespons kabar mengenai naturalisasi dirinya untuk memakai jersi timnas Indonesia di Inter Serie A.	Netral

Pada baris pertama, tweet tersebut mencerminkan sentimen positif karena menyatakan bahwa calon pemain naturalisasi telah bergabung dengan Timnas Indonesia U20, menunjukkan harapan dan potensi kontribusi untuk tim. Pada baris kedua, tweet tersebut mencerminkan sentimen negatif terhadap sepak bola Indonesia, menyatakan bahwa kualitasnya primitif dan menyarankan untuk mengganti semua pemain lokal dengan pemain naturalisasi. Sedangkan baris ketiga mencerminkan sentimen netral karena Emil Audero Mulyadi memberikan respons netral terkait kabar naturalisasinya untuk memperkuat timnas Indonesia di Inter Serie A.

Evaluasi

Setelah melakukan tahap evaluasi dengan data 80% training dan 20% data testing, menghasilkan nilai akurasi sebesar 0,7% dengan jumlah sentimen netral 29, dan positif 54.

```

Akurasi Model SVM:
0.7108

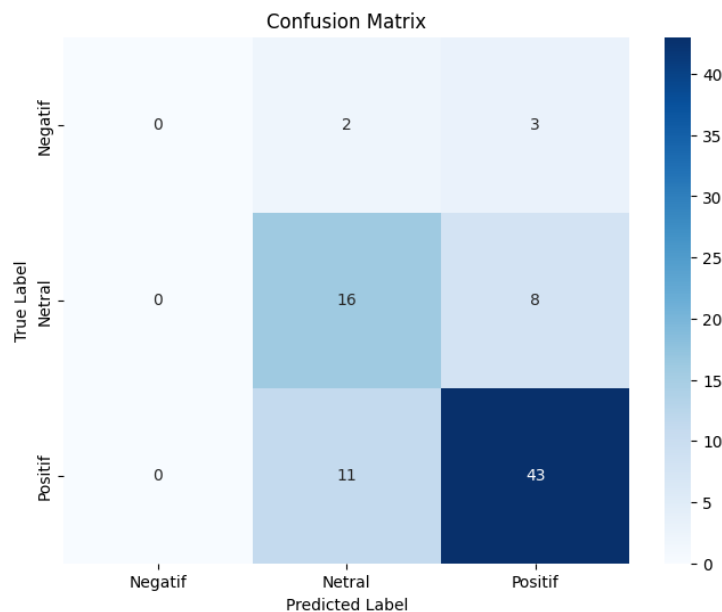
Classification Report:
              precision    recall  f1-score   support

   Negatif         1.00      0.00      0.00         5
    Netral         0.55      0.67      0.60        24
    Positif         0.80      0.80      0.80        54

   accuracy          0.71
  macro avg          0.78
 weighted avg          0.74

Jumlah Sentimen Prediksi:
Netral: 29
Positif: 54
    
```

Gambar 1. Nilai Akurasi



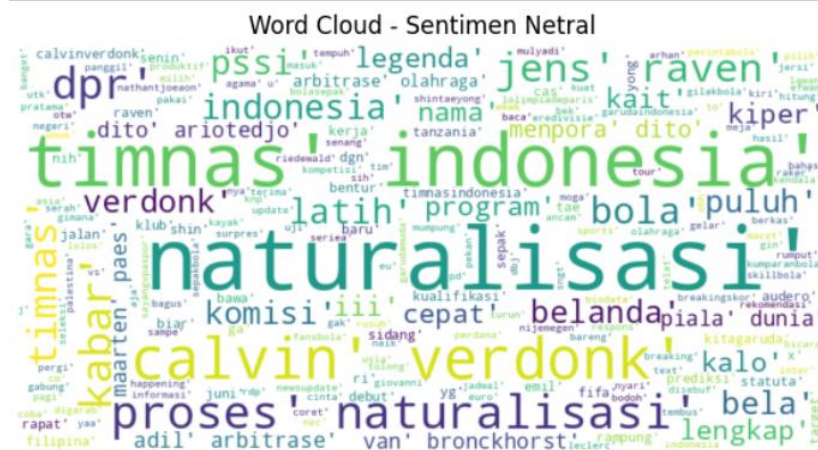
Gambar 2. Heatmap Confusion Matrix

Heatmap confusion matrix yang divisualisasikan pada Gambar 2 bertujuan untuk mempermudah dalam menentukan perbandingan antara sentimen aktual dan sentimen prediksi. Visualisasi ini memberikan gambaran yang lebih jelas mengenai performa model dalam mengklasifikasikan sentimen. Berdasarkan hasil dari tahapan-tahapan yang telah dilakukan, penelitian ini disajikan dalam bentuk visualisasi. Visualisasi hasil penelitian ditampilkan dalam bentuk word cloud dan dashboard. Word cloud untuk keseluruhan sentimen ditunjukkan pada:

- Gambar 3 untuk sentimen positif
- Gambar 4 untuk sentimen netral



Gambar 3. World Cloud Sentimen Positif



Gambar 4. World Cloud Sentimen Netral

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, algoritma Support Vector Machine (SVM) terbukti efektif dalam menganalisis sentimen publik terkait kebijakan naturalisasi pemain tim nasional Indonesia di media sosial Twitter. Melalui pengolahan data dengan teknik crawling, preprocessing, serta pembobotan menggunakan metode TF-IDF, model SVM berhasil mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral dengan tingkat akurasi yang memuaskan. Penelitian ini memberikan wawasan penting bagi para pembuat kebijakan dan pemangku kepentingan dalam memahami reaksi dan opini publik terhadap kebijakan naturalisasi, yang dapat dijadikan acuan dalam pengambilan keputusan di masa mendatang. Kesimpulannya, penerapan SVM dalam analisis sentimen media sosial memiliki potensi besar untuk digunakan pada isu-isu lain yang mempengaruhi opini publik secara luas.

DAFTAR PUSTAKA

- Abdusyukur, Fatwa. "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter." *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, vol. 12, no. 1, 18 May 2023, pp. 73-82, ojs.unikom.ac.id/index.php/komputa/article/view/9418, <https://doi.org/10.34010/komputa.v12i1.9418>. Accessed 19 May 2024.
- Lathifah, Laila, et al. "Sistem Crawling Data Instrumen Akreditasi Berbasis Selenium Dan Pandas." *Transient: Jurnal Ilmiah Teknik Elektro*, vol. 10, no. 1, 17 Mar. 2021, pp. 84-91, <https://doi.org/10.14710/transient.v10i1.84-91>. Accessed 12 Apr. 2023.

- Nanda, Robbi, et al. "Klasifikasi Berita Menggunakan Metode Support Vector Machine." *Jurnal Nasional Komputasi Dan Teknologi Informasi/Jurnal Nasional Komputasi Dan Teknologi Informasi*, vol. 5, no. 2, 28 Apr. 2022, pp. 269–278, <https://doi.org/10.32672/jnkti.v5i2.4193>. Accessed 25 May 2024.
- R Damasela, et al. "Penerapan Metode Support Vector Machine (Svm) Untuk Mendeteksi Penyalahgunaan Narkoba." *Parameter*, vol. 1, no. 2, 13 Oct. 2022, pp. 111–122, <https://doi.org/10.30598/parameter.v1i2pp111-122>. Accessed 1 Oct. 2024.
- Tri Yuli Pahtoni, and Handaru Jati. "Analisis Sentimen Data Twitter Terkait Chatgpt Menggunakan Orange Data Mining." *Jurnal Teknologi Informasi Dan Ilmu Komputer*, vol. 11, no. 2, 25 Apr. 2024, pp. 329–336, <https://doi.org/10.25126/jtiik.20241127276>. Accessed 1 Oct. 2024.
- Utomo, Muchammad Chandra Cahyo, et al. "Analisis Sentimen Media Sosial Twitter Pada Kasus Pemberlakuan Pembatasan Kegiatan Masyarakat Dengan Menggunakan Metode Naïve Bayes Classifier." *Journal of Mathematics & Information Technology*, vol. 1, no. 1, 2023, pp. 41–48