

Analisis Perbandingan Metode *Support Vector Machine* dan *Naïve Bayes* pada Persepsi Publik terhadap Naturalisasi Pemain Timnas Sepak Bola Indonesia: Studi Kasus Sosial Media X

Yoga Sulistiyo Widodo, Akhsin Nurlayli

Universitas Negeri Yogyakarta

Email: yogasulistiyo.2021@student.uny.ac.id

ABSTRAK

Pemerintah menerapkan kebijakan naturalisasi untuk meningkatkan prestasi Timnas Sepak Bola Indonesia, namun hal ini menimbulkan sentimen yang beragam di masyarakat. Penelitian ini bertujuan untuk membandingkan performa algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* serta menganalisis persepsi publik terhadap kebijakan tersebut melalui data media sosial X. Dengan menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM), penelitian ini mengolah 10.454 data yang diambil dari Januari 2020 hingga Januari 2025. Hasil evaluasi menunjukkan SVM lebih unggul dengan akurasi 0,80 dan F1-score 0,72, dibandingkan *Naïve Bayes* dengan akurasi 0,76 dan F1-score 0,66. Analisis sentimen mengungkapkan bahwa persepsi publik secara keseluruhan cenderung negatif jika dibandingkan langsung dengan sentimen positif, meskipun sentimen netral sedikit lebih unggul. Implikasi dari penelitian ini adalah memberikan wawasan berbasis data bagi PSSI untuk mengevaluasi dan merumuskan strategi komunikasi kebijakan naturalisasi yang lebih efektif dan sejalan dengan aspirasi publik.

Kata kunci: Analisis Sentimen, Naturalisasi, Persepsi Publik, Support Vector Machine, *Naïve Bayes*

ABSTRACT

The government implemented a naturalization policy to improve the achievements of the Indonesian national football team, but this caused mixed sentiments in the community. This research aims to compare the performance of Support Vector Machine (SVM) and Naïve Bayes algorithms and analyze public perception of the policy through X social media data. Using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, this research analyzed 10,454 data points collected from January 2020 to January 2025. The evaluation results showed that SVM was superior, with an accuracy of 0.80 and an F1-score of 0.72, compared to Naïve Bayes, which had an accuracy of 0.76 and an F1-score of 0.66. Sentiment analysis revealed that overall public perception tended to be negative when directly compared to positive sentiment, although neutral sentiment was slightly superior. This research implies that it provides data-based insights for PSSI to evaluate and formulate a more effective naturalization policy communication strategy that is in line with public aspirations.

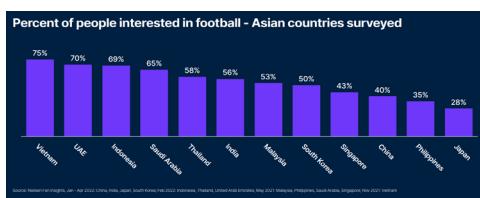
Keywords: sentiment analysis, naturalization, public perception, Support Vector Machine, *Naïve Bayes*

PENDAHULUAN

Sepak bola merupakan olahraga yang sangat populer di Indonesia. Berdasarkan (World Football Report, 2022), 69% populasi Indonesia merupakan penggemar sepak bola,

menjadikannya peringkat ketiga di Asia setelah Vietnam dan Uni Emirat Arab. Popularitas ini didukung oleh kemudahan akses, peralatan yang sederhana, serta peranannya sebagai budaya masyarakat (Mulyana et al., 2024).

Antusiasme publik terhadap sepak bola nasional juga tinggi. BRI Liga 1 musim ini mencatatkan 811.000 penonton langsung di stadion, sementara 28,7% masyarakat mengaku rutin menonton pertandingan Liga 1 (Rainer, 2025). Di sisi lain, media sosial dan platform digital seperti OTT menjadi sarana baru untuk menikmati tayangan sepak bola, dengan 72% penggemar mengakses pertandingan melalui media sosial (World Football Report, 2022). Media sosial juga menjadi ruang diskusi dan kritik publik terhadap klub, pemain, serta kebijakan sepak bola nasional.



Gambar 1. Peringkat Penggemar Sepak Bola

Salah satu isu yang ramai dibicarakan adalah kebijakan naturalisasi pemain oleh Tim Nasional Indonesia. Naturalisasi dalam konteks Timnas bertujuan meningkatkan kualitas tim melalui pemain keturunan, dengan total 19 pemain dinaturalisasi jumlah tertinggi dibanding era sebelumnya (Adi Yaksa, 2024). Kepemimpinan Erick Thohir dan kehadiran pelatih Shin Tae Yong turut mendorong kemajuan sepak bola nasional, tercermin dari peningkatan peringkat FIFA Indonesia dari 173 ke 127 (CNBC Indonesia, 2025).

Namun, kebijakan naturalisasi menuai pro dan kontra. Survei Indikator

Politik Indonesia (2024) menunjukkan bahwa meskipun 71,5% masyarakat mendukung kebijakan ini, sebagian menolaknya karena alasan nasionalisme dan identitas budaya. Isu ini juga mendapat sorotan media dan menjadi perdebatan publik, baik di ruang nyata maupun media sosial seperti platform X (Adharani et al., 2025).

Media sosial X sendiri sangat populer di Indonesia, dengan 25,2 juta pengguna aktif (Kemp, 2025). Banyak pengguna menyampaikan opini mereka secara langsung terkait naturalisasi. Karena volume data yang besar, diperlukan pendekatan otomatis untuk menganalisis sentimen publik secara efektif.

Untuk itu, penelitian ini menggunakan metode analisis sentimen berbasis *machine learning* dengan pendekatan metodologi CRISP-DM. Metodologi ini dipilih karena bersifat fleksibel, terstruktur, dan telah menjadi standar dalam proyek *data mining* serta cocok untuk data sosial yang besar dan tidak terstruktur (Cazacu et al., 2020; Dåderman & Rosander, 2018). Tahapan CRISP-DM meliputi: pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment* (Kurniawan, 2022).

Penelitian ini menerapkan dua algoritma populer, *Support Vector Machine* (SVM) dan *Naïve Bayes*, yang telah terbukti efektif dalam berbagai studi analisis sentimen, dengan akurasi di atas 75% (Hafidz et al., 2020; Luthfanida, 2022). Meski sudah digunakan dalam beberapa studi terkait naturalisasi

(Franko et al., 2024; Lubis et al., 2024), perbandingan langsung antara SVM dan *Naïve Bayes* pada topik ini masih jarang dilakukan.

Penelitian ini bertujuan untuk membandingkan performa kedua algoritma dalam mengklasifikasikan sentimen masyarakat terhadap naturalisasi pemain berdasarkan data dari media sosial X. Hasil dari analisis ini diharapkan dapat memberi masukan kebijakan bagi PSSI serta menjembatani kesenjangan antara kebijakan dan aspirasi publik dalam perkembangan sepak bola nasional.

METODE

Penelitian ini menggunakan pendekatan kerangka kerja berbasis CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri atas enam tahap: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* (Ayele, 2021).

Data dikumpulkan melalui crawling menggunakan tools Tweet Harvest berbasis Node.js, dengan autentikasi melalui bearer token yang terhubung ke akun media sosial X (dahulu Twitter). Dataset terdiri atas 13.276 tweet berbahasa Indonesia yang mengandung topik naturalisasi pemain Timnas Indonesia. Proses *preprocessing* data mencakup tahapan *cleaning* (penghapusan simbol dan URL), *case folding* (mengubah huruf menjadi *lowercase*), *stopword removal*, *stemming*, dan pelabelan data menjadi tiga kategori sentimen: positif, negatif, dan netral. Data kemudian direpresentasikan dalam bentuk numerik menggunakan metode

Term Frequency–Inverse Document Frequency (TF-IDF).

Tahap pemodelan dilakukan dengan dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM). Data dibagi menjadi dua subset: 80% untuk data latih dan 20% untuk data uji. Untuk proses pelatihan dan pengujian model, digunakan pembobotan TF-IDF terhadap token-token dalam tweet. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, berdasarkan hasil *confusion matrix*. Tahap akhir berupa *deployment*, di mana hasil klasifikasi disajikan dalam bentuk visualisasi serta disusun menjadi laporan yang mendukung analisis sentimen publik terhadap kebijakan naturalisasi pemain Timnas Indonesia.

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini disajikan secara objektif pada pembahasan analisis sentimen terkait perbandingan metode SVM dan *Naïve Bayes* dalam Analisis sentimen naturalisasi pemain Timnas Indonesia di media sosial X.

1. *Business Understanding*

Tahap ini bertujuan untuk memahami arah dan tujuan penelitian, yaitu menganalisis sentimen publik terhadap kebijakan naturalisasi pemain Timnas Indonesia melalui media sosial X, mengevaluasi dan membandingkan performa algoritma *Naïve Bayes* dan SVM, serta memberikan wawasan bagi PSSI dalam merumuskan kebijakan yang selaras dengan opini publik.

2. *Data Understanding*

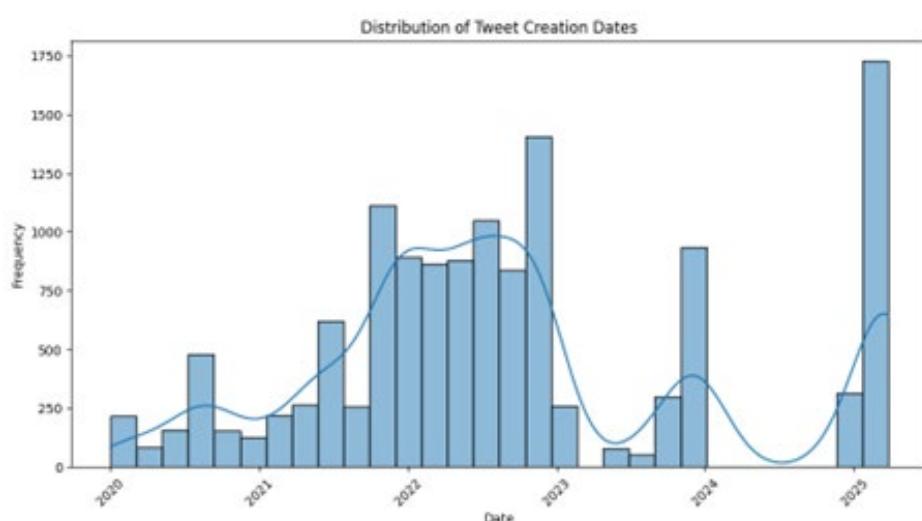
Data dalam penelitian ini dikumpulkan dari media sosial X (Twitter) menggunakan tool Tweet Harvest berbasis Node.js. Proses

crawling dilakukan terhadap tweet berbahasa Indonesia yang mengandung kata kunci: "Naturalisasi Timnas Indonesia", "Naturalisasi Timnas Sepak Bola Indonesia", dan "Naturalisasi Timnas Shin Tae Yong", dengan rentang waktu dari Januari 2021 hingga Januari 2025. Akses API dilakukan menggunakan token autentikasi (*auth_token*) yang dikonfigurasi pada Jupyter Notebook untuk menyesuaikan parameter pencarian seperti jumlah tweet, kata kunci, dan nama file output.

Data berhasil dikumpulkan dalam enam batch dengan total 13.276 tweet, dan disimpan dalam format CSV (Satria, 2024). Kolom utama yang digunakan untuk analisis adalah *full_text* yang berisi isi tweet. Data awal terdiri atas 15 kolom, dengan berbagai tipe data seperti object, int64,

dan float64. Analisis statistik deskriptif awal menunjukkan bahwa sebagian besar tweet tidak mendapatkan likes atau retweet, namun terdapat beberapa tweet dengan engagement sangat tinggi (maksimal 869 retweet).

Distribusi data berdasarkan waktu dianalisis menggunakan pustaka Seaborn, dengan konversi kolom *created_at* menjadi format datetime. Hasil visualisasi menunjukkan puncak aktivitas tweet terjadi pada tahun 2025, yang bertepatan dengan agenda kualifikasi Piala Dunia. Sebaliknya, terdapat penurunan drastis pada 2024, kemungkinan akibat dominasi isu politik menjelang pemilihan umum. Visualisasi dari distribusi tersebut dapat dilihat pada Gambar 2.



Gambar 2. Distribusi Data Tweet

3. Data Preparation

Data hasil crawling dari media sosial X tidak dapat langsung digunakan karena adanya variasi gaya bahasa, penulisan, dan elemen non-standar seperti emoji, tagar, atau tautan. Oleh

karena itu, dilakukan proses text preprocessing untuk memastikan konsistensi dan meningkatkan kemampuan generalisasi model klasifikasi. Tahapan preprocessing meliputi:

- a. Data *Cleaning*, Proses ini merupakan tahapan yang bertujuan untuk membersihkan data dari elemen-elemen yang tidak dibutuhkan atau tidak relevan seperti angka, tanda baca, karakter khusus, tautan, hingga duplikasi data (F. Karaca, 2024). Hasil dari proses ini dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Hasil Data *Cleaning*

Sebelum	Sesudah
@Indotransfer	padahal
Padahal pngin	pngin banget
banget lihat	lihat konate
Konate	dinaturalisasi
dinaturalisasi	dan bisa
dan bisa main	main buat
buat Timnas	Timnas
Indonesia	indonesia

- b. *Normalization*, Tahapan ini merupakan proses memperbaiki berbagai macam penulisan kata, kesalahan penulisan, penggunaan bahasa yang salah dan tidak baku. Selain itu, kata-kata yang disingkat juga diubah agar menjadi kata yang benar dan dapat diproses dengan mudah menjadi sebuah informasi yang baik dan jelas (Makmur et al., 2024). Seperti pada Tabel 2 di bawah ini.

Tabel 2. Hasil *Normalization*

Sebelum	Sesudah
padahal pngin	padahal
banget lihat	pngin banget
konate	lihat konate
dinaturalisasi	dinaturalisasi
dan bisa main	dan bisa main
buat Timnas	buat Timnas
Indonesia	indonesia

- c. *Labeling*, setelah melalui tahap *cleaning* dan *normalization*, data

diberi label sentimen menggunakan pretrained transformer model IndoBERT untuk klasifikasi awal. Hasil pelabelan otomatis ini kemudian direvisi secara manual guna memastikan akurasi dan relevansi konteks. Proses verifikasi penting dilakukan karena tidak semua tweet yang diperoleh, meskipun menggunakan kata kunci spesifik, relevan terhadap topik naturalisasi pemain Timnas (Pratiwi & Kamayani, 2024).

Tabel 3. Kriteria Pelabelan Sentimen

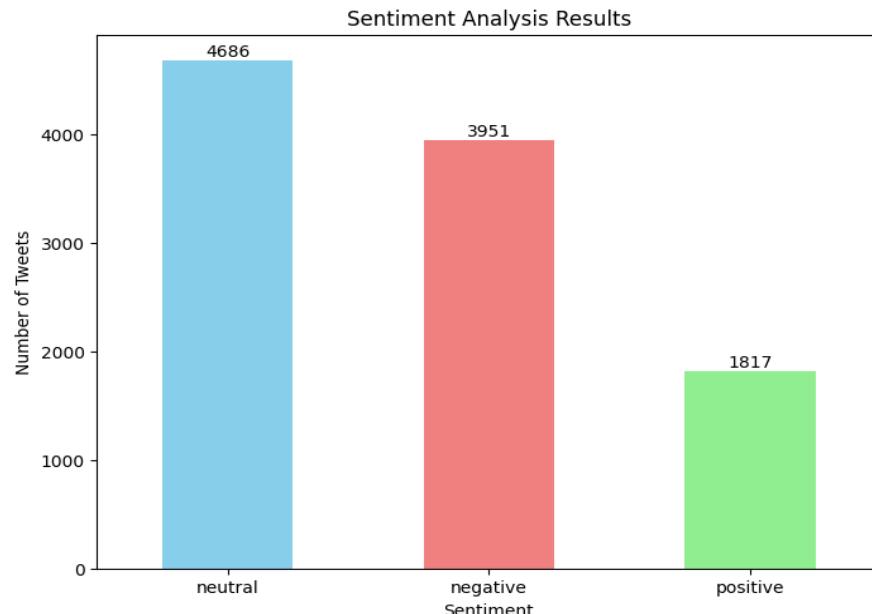
Label	Kriteria
Positif	<ol style="list-style-type: none"> 1. Mendukung kebijakan naturalisasi Timnas 2. Memuji performa pemain naturalisasi 3. Bangga terhadap Timnas Indonesia
Negatif	<ol style="list-style-type: none"> 1. Menolak kebijakan naturalisasi Timnas 2. Mengomentari dan mencari kesalahan terhadap performa pemain naturalisasi 3. Tidak Bangga dengan Timnas Indonesia
Netral	<p><i>Tweet</i> yang tidak termasuk ke dalam kriteria <i>tweet</i> positif maupun negatif yaitu <i>tweet</i> yang ambigu atau tidak memiliki seperti headline berita.</p>

Oleh karena itu, pelabelan dilakukan secara dua tahap: sentimen (positif, negatif, netral) dan konteks (relevan/tidak relevan) untuk menghasilkan dataset yang valid dan terfokus.

Tabel 4. Kriteria Pelabelan Sentimen

Sesuai	Tidak Sesuai
Berhubungan dengan naturalisasi Timnas Indonesia	Tidak berhubungan dengan naturalisasi Timnas Indonesia

Hasil dari proses *labeling* berdasarkan pendekatan IndoBERT dan manual maka diperoleh distribusi dari label seperti pada Gambar 3.



Gambar 3. Distribusi Hasil Label Tweet

- d. *Tokenization*, yaitu memecah kalimat menjadi unit-unit yang lebih kecil atau biasa disebut token yang berfungsi untuk memudahkan proses analisis pada tahap *stemming* (Schmidt et al., 2024). Seperti pada Tabel 5.

Tabel 5. Hasil Tokenization

Sebelum	Sesudah
Mantan pemain naturalisasi Timnas indonesia jadi pemain terbaik liga	[padahal,pengin,b anget,lihat,konate, di,naturalisasi,dan bisa,main,buat,Ti mnas,Indonesia]

- e. *Stopword*, menghilangkan kata-kata yang tidak mendeskripsikan

sesuatu atau tidak memiliki bobot atau arti (Agustine Fitrrana et al., 2024). Kode untuk melakukan *stopword* menggunakan pustaka NLTK

Tabel 6. Hasil Stopword

Sebelum	Sesudah
padahal pengin banget lihat konate	pengin banget lihat konate
di naturalisasi dan bisa main buat	naturalisasi main Timnas
Timnas indonesia	indonesia

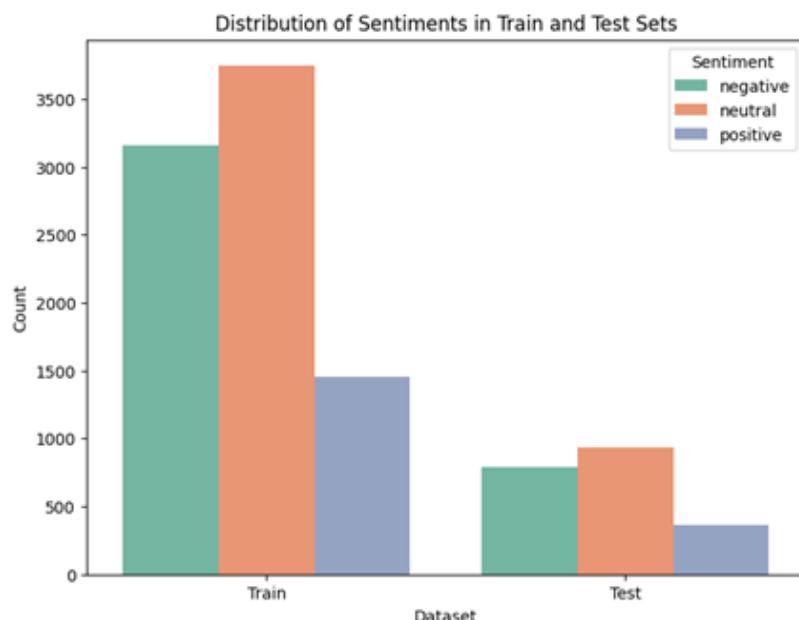
- f. *Stemming*, yaitu mengubah kata yang berawalan *infix* maupun berakhiran *suffix* menjadi sebuah kata aslinya atau kata dasarnya, di

mana kata dasar tersebut lebih memiliki makna seperti pada Tabel 7.

Tabel 7. Hasil Stemming

Sebelum	Sesudah
Jokowi tidak persoalkan rencana naturalisasi pemain Timnas indonesia	jokowi tidak soal rencana naturalisasi pemain Timnas indonesia

- g. Data *Splitting*, Sebelum melatih model, kita perlu melakukan pembagian set data menjadi 2 kategori untuk pelatihan dan pengujian. Dalam penelitian ini perbandingan yang digunakan adalah perbandingan 80:20 atau sebanyak 80% data dijadikan sebagai data latih dan 20% data dijadikan data uji. Hasilnya seperti pada Gambar 4.



Gambar 4. Komposisi Data Splitting

4. Modelling

Selanjutnya setelah tahap data preparation, akan dilakukan tahap ekstraksi fitur menggunakan TF-IDF untuk transformasi data teks menjadi angka (fitur). Dalam penelitian ini, TF-IDF tidak diberikan parameter.

Tabel 8. Hasil *TF-IDF*

Kata	Nilai
orang	0,4095
aneh	0,5974
Tidak	0,2315
setuju	0,5161
naturalisasi	0,1148

Selanjutnya model diberikan parameter terbaik menggunakan metode *GridSearchCV* untuk menemukan parameter terbaik untuk model SVM maupun *Naïve Bayes*. Seperti *Hyperparameter Naïve Bayes* pada Tabel 9.

Tabel 9. *Hyperparameter Naïve Bayes*

Parameter	Nilai
alpha	0,21544346900318845
fit_prior	True

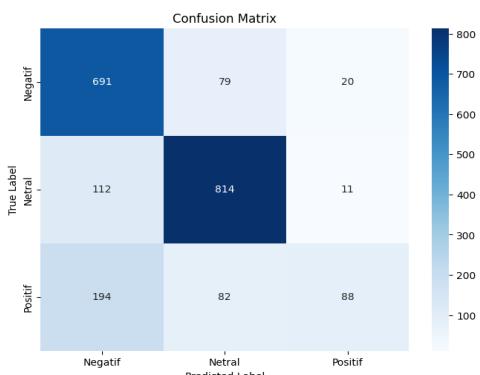
Selanjutnya untuk Parameter terbaik SVM dapat dilihat pada Tabel 9 di bawah ini.

Tabel 10. Hyperparameter SVM

Parameter	Nilai
C	1
Gamma	1

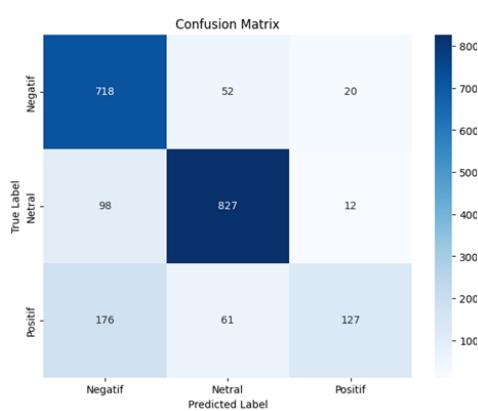
5. Evaluation

Evaluasi atas implementasi model *machine learning* menggunakan algoritma *Naïve Bayes* dan *SVM* menghasilkan algoritma terbaik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1 score*. Evaluasi dilakukan terhadap data uji sebesar 2091 data menggunakan visualisasi *confusion matrix* seperti pada *Naïve Bayes* pada Gambar 5 di bawah ini.



Gambar 5. Confusion matrix *Naïve Bayes*

Sedangkan, untuk *SVM* dapat dilihat pada Gambar 6 di bawah ini.



Gambar 6. Confusion matrix *SVM*

Hasil perbandingan performa antara algoritma *Naïve Bayes* dan *Support Vector Machine* (*SVM*) menunjukkan bahwa *SVM* memiliki kinerja yang lebih unggul dalam melakukan analisis sentimen terhadap kebijakan naturalisasi pemain Timnas Indonesia. Setelah melalui proses optimasi menggunakan *GridSearchCV*, model *SVM* mencapai nilai akurasi 0,80 dan *F1-score* 0,72. Performa ini melampaui model *Naïve Bayes* yang mencatatkan akurasi 0,76 dan *F1-score* 0,66 saja.

6. Deployment

Pada penelitian ini, proses *deployment* dilakukan dengan cara menyusun laporan akhir berupa dokumentasi hasil analisis sentimen yang telah dilakukan menggunakan dua metode, yaitu *Support Vector Machine* dan *Naïve Bayes*. Hasil analisis ini disajikan dalam bentuk visualisasi seperti *wordcloud*, komposisi sentimen, serta *confusion matrix* untuk masing-masing model.

- Hasil Analisis Sentimen *Naïve Bayes*, berdasarkan data yang diambil dari X (Januari 2020 - Januari 2025) mengenai naturalisasi Tim Nasional Indonesia, *Naïve Bayes* memprediksi 4542 sentimen negatif, 4903 netral, dan 1009 positif.
- Hasil Analisis Sentimen *SVM*, berdasarkan data yang diambil dari X (Januari 2020 - Januari 2025) mengenai naturalisasi Tim Nasional Indonesia, *SVM* memprediksi 4279 sentimen negatif, 4718 netral, dan 1457 positif.

KESIMPULAN

Penelitian ini membahas analisis perbandingan metode *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengidentifikasi persepsi publik terhadap kebijakan naturalisasi pemain Timnas Sepak Bola Indonesia, berdasarkan data dari media sosial X pada periode Januari 2020 hingga Januari 2025.

Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa yang lebih unggul dibandingkan dengan *Naïve Bayes* dalam melakukan analisis sentimen. Setelah dilakukan proses optimasi menggunakan *GridSearchCV*, model SVM mencapai akurasi sebesar 0,80 dan F1-score sebesar 0,72, mengungguli *Naïve Bayes* yang hanya mencatatkan akurasi 0,76 dan F1-score 0,66.

Dari segi distribusi sentimen, model *Naïve Bayes* memprediksi 4.542 sentimen negatif, 4.903 sentimen netral, dan 1.009 sentimen positif. Sementara itu, model SVM memprediksi 4.279 sentimen negatif, 4.718 sentimen netral, dan 1.457 sentimen positif. Hasil ini mengindikasikan bahwa persepsi publik terhadap kebijakan naturalisasi cenderung terpolarisasi, dengan dominasi sentimen netral dan negatif.

Secara keseluruhan, temuan ini memberikan gambaran bahwa kebijakan naturalisasi pemain Timnas memicu respons publik yang cukup signifikan. Dominasi sentimen netral dan negatif menjadi perhatian penting yang dapat dijadikan landasan bagi pemangku kepentingan, seperti PSSI dan pemerintah, dalam mengevaluasi efektivitas kebijakan, merumuskan strategi komunikasi publik yang tepat, serta menentukan arah kebijakan yang lebih inklusif dan berkelanjutan dalam pengembangan sepak bola nasional.

DAFTAR RUJUKAN

- Adharani, S. K., Kacung, S., & Vitianingsih, A. V. (2025). *Sentiment Analysis on Indonesian National Football Team Naturalization using KNN and SVM*.
- Adi Yaksa, M. (2024, September 17). *Membandingkan Jumlah Pemain Naturalisasi Timnas Indonesia di Era Shin Tae-yong dan Sebelumnya*. bola.com. <https://www.bola.com/indonesia/read/5703718/membandingkan-jumlah-pemain-naturalisasi-timnas-indonesia-di-era-shin-tae-yong-dan-sebelumnya>
- Agustine Fitrrana, Lady, Linawati, S., Herlinawati, N., Sa'adah, R., & Seimahuria, S. (2024). ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP BRAND INDOSAT MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 4291–4297. <https://doi.org/10.36040/jati.v8i3.9866>
- Ayele, W. Y. (2021). *Adapting CRISP-DM for Idea Mining: A Data Mining Process for Generating Ideas Using a Textual Dataset*. <https://doi.org/10.48550/ARXI.V.2105.00574>
- Cazacu, M., Titan, E., & The Bucharest University of Economic Studies. (2020). Adapting CRISP-DM for Social Sciences. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 11(2sup1), 99–106.

- <https://doi.org/10.18662/brain/11.2Sup1/97>
- CNBC Indonesia, T. I. (2025, January 6). *Era STY Timnas Indonesia Naik 46 Peringkat, Garuda Kembali Ganas!* [NewsLetter]. CNBC Indonesia. <https://www.cnbcindonesia.com/research/20250106123430-128-601046/era-sty-timnas-indonesia-naik-46-peringkat-garuda-kembali-ganas>
- Däderman, A., & Rosander, S. (2018). *Evaluating Frameworks for Implementing Machine Learning in Signal Processing*.
- F. Karaca, M. (2024). Effects of Preprocessing on Text Classification in Balanced and Imbalanced Datasets. *KSI Transactions on Internet and Information Systems*, 18(3). <https://doi.org/10.3837/tiis.2024.03.004>
- Franko, B., Wilyanto, N., & Irsyad, H. (2024). *Analisis Sentimen Terhadap Naturalisasi Pemain pada Youtube Menggunakan Decision Tree dan Naïve Bayes.* 03.
- Hafidz, N., Anggraeni, S., & Gata, W. (2020). *Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes*.
- Kemp, S. (2025, March 12). *X Users, Stats, Data, Trends, and More*. DataReportal – Global Digital Insights. <https://datareportal.com/essential-x-stats>
- Kurniawan, D. (2022). *OPTIMIZATION SENTIMENT ANALYSIS USING CRISP-DM AND NAÏVE BAYES METHODS IMPLEMENTED ON SOCIAL MEDIA*.
- Lubis, M. G. R., Sitompul, D. S., Giovanni, T. M., Ramadhani, F., & Dewi, S. (2024). Evaluasi Kinerja Algoritma Support Vector Machine (SVM) Dalam Analisis Sentimen Publik Terhadap Naturalisasi Timnas Indonesia di Twitter. *Journal of Accounting Law Communication and Technology*, 2(1), 81–89. <https://doi.org/10.57235/jalakotek.v2i1.4180>
- Luthfanida, L. (2022). ANALISIS SENTIMEN DATA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) TENTANG PRESIDEN JOKOWI 3 PERIODE. *Djtechno: Jurnal Teknologi Informasi*, 3(1), 5–11. <https://doi.org/10.46576/djtecnovo.v3i1.2143>
- Makmur, H., Wulandari, W., Surianto, D. F., & Fajar B, M. (2024). Analisis Sentimen Penghapusan Skripsi sebagai Tugas Akhir Mahasiswa Menggunakan Metode Multi-Layer Perceptron. *Komputika : Jurnal Sistem Komputer*, 13(2). <https://doi.org/10.34010/komputika.v13i2.12402>
- Mulyana, A., Shafira, D., Aditya, F. A., Anggia, I. R., Haq, M. A., Daud, M. R. B., Putri, N. N. K., & Ardhiyansah, S. (2024). Mengapa Sepak Bola Menjadi Populer di Kalangan Siswa Sekolah Dasar? *Indo-MathEdu Intellectuals Journal*, 5(3), 3284–3289. <https://doi.org/10.54373/imeij.v5i3.1261>
- Pratiwi, A. A., & Kamayani, M. (2024). Perbandingan Pelabelan

- Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes. *Jurnal Eksplora Informatika*, 14(1), 96–107. <https://doi.org/10.30864/eksplora.v14i1.1093>
- Rainer, P. (2025, February 13). *Seberapa Sering Masyarakat Tonton Pertandingan Liga Utama Indonesia? Seberapa Sering Masyarakat Tonton Pertandingan Liga Utama Indonesia?* <https://goodstats.id/article/menilik-data-tingkat-ketersaksian-liga-utama-di-indonesia-X9hyu>
- Satria, H. (2024). *Helmisatria/tweet-harvest* [TypeScript]. <https://github.com/helmisatria/tweet-harvest> (Original work published 2023)
- Schmidt, C. W., Reddy, V., Zhang, H., Alameddine, A., Uzan, O., Pinter, Y., & Tanner, C. (2024). *Tokenization Is More Than Compression* (No. arXiv:2402.18376). arXiv. <https://doi.org/10.48550/arXiv.2402.18376>
- World Football Report.* (2022). Nielsen Company. <https://denisdoeland.com/wp-content/uploads/2022/09/Nielsen-World-Football-Report-2022.pdf>