

ANALISIS SENTIMEN PENGGEMAR SEPAK BOLA TERHADAP PROSES NATURALISASI PEMAIN TIMNAS INDONESIA PADA MEDIA SOSIAL X MENGGUNAKAN SUPPORT VECTOR MACHINE

Muh. Amal Anugra.S^{*1}, Natalis Ransi², Gusti Arviana Rahman³

Email: ¹amalanugra286@gmail.com, ²natalis.ransi@gmail.com, ³arviana.rahman@uho.ac.id

Abstrak

Proses naturalisasi pemain untuk tim nasional Indonesia telah menjadi topik yang hangat diperbincangkan di kalangan penggemar sepak bola. Media sosial, khususnya platform X, menjadi salah satu tempat utama bagi para penggemar untuk mengungkapkan pendapat dan sentimen mereka mengenai hal ini. Penelitian ini bertujuan untuk menganalisis sentimen penggemar sepak bola terhadap proses naturalisasi pemain timnas Indonesia dengan menggunakan metode Support Vector Machine (SVM). Analisis sentimen dapat digunakan untuk menganalisis pendapat suporter dan mengelompokkannya menjadi tiga kategori yaitu positif, netral, dan negatif. Penelitian ini menggunakan metode *Support Vector Machine* untuk melakukan klasifikasi yang dilakukan dengan garis pembatas (*hyperplane*) yang memisahkan kelas-kelas yang ada pada kumpulan data *tweet* suporter di media sosial X (*twitter*) yang memiliki opini tentang naturalisasi. Hasil penelitian menunjukkan bahwa sentimen penggemar sepak bola Indonesia terhadap proses naturalisasi pemain timnas sebagian besar cenderung positif, dengan proporsi sentimen positif mencapai 38%, sentimen negatif sebesar 37%, dan netral 25%. Analisis ini juga mengungkapkan beberapa faktor yang mempengaruhi sentimen penggemar, termasuk performa pemain naturalisasi di lapangan, kebijakan federasi sepak bola, serta pengaruh media dan opini publik.

Kata kunci: *analisis sentimen, naturalisasi pemain, timnas Indonesia, support vector machine, X (twitter).*

Abstract

The naturalization process for players for the Indonesian national team has become a hot topic of discussion among football fans. Social media, especially platform X, has become one of the main places for fans to express their opinions and sentiments regarding this matter. This research aims to analyze football fans' sentiments towards the naturalization process of Indonesian national team players using the Support Vector Machine (SVM) method. Sentiment analysis can be used to analyze supporters' opinions and group them into three categories, namely positive, neutral and negative. This research uses the Support Vector Machine method to carry out classification which is carried out with a dividing line (hyperplane) that separates the classes in the data collection of supporters' tweets on social media X (twitter) who have opinions about naturalization. The research results show that the sentiment of Indonesian football fans towards the naturalization process for national team players mostly tends to be positive, with the proportion of positive sentiment reaching 38%, negative sentiment of 37%, and neutral sentiment of 25%. This analysis also reveals several factors that influence fan sentiment, including the performance of naturalized players on the field, football federation policies, and the influence of the media and public opinion.

Keywords: *sentiment analysis, naturalization of players, Indonesian national team, support vector machine, x (twitter).*

1. PENDAHULUAN

Data mining adalah proses mengekstraksi informasi yang bermanfaat dari kumpulan data yang besar dan kompleks. Ini melibatkan penggunaan teknik statistik, algoritma *machine learning*, dan metode komputasi lainnya untuk mengidentifikasi pola, tren, dan hubungan dalam data yang bisa digunakan untuk membuat keputusan yang lebih baik. Analisis sentimen, atau *opinion mining*, adalah cabang dari data mining yang berguna untuk menganalisis pendapat orang terhadap suatu entitas seperti layanan, organisasi, produk, individu,

masalah, peristiwa, dan lain-lain. Analisis ini memiliki fokus pada pendapat yang mengekspresikan sentimen positif, netral, dan negatif. Analisis sentimen dapat diterapkan menggunakan beberapa algoritma atau metode *machine learning*.

Sepak bola merupakan olahraga yang sangat populer di Indonesia. Negara ini memiliki salah satu jumlah penggemar sepak bola terbesar di dunia, yang mencerminkan *antusiasme* tinggi terhadap perkembangan tim nasional dan klub lokal. Popularitas sepak bola di Indonesia terlihat dari berbagai aspek, mulai dari banyaknya penonton di

stadion, tingginya rating siaran pertandingan di televisi, hingga aktivitas yang ramai di media sosial. Selain itu, pertandingan-pertandingan internasional yang melibatkan tim nasional Indonesia selalu dinanti-nantikan oleh para penggemar. Isu yang sering dibicarakan suporter Tanah Air adalah naturalisasi pemain asing untuk memperkuat timnas. Naturalisasi pemain asing sering diadopsi oleh banyak negara sebagai strategi untuk memperkuat tim nasional mereka dengan pemain berbakat dari luar negeri. Namun, proses naturalisasi ini sering kali menimbulkan perdebatan di kalangan penggemar, terutama terkait dengan identitas nasional dan kebanggaan bangsa yang mereka miliki.

Naturalisasi adalah proses dimana warga negara asing menjadi warga negara dari negara di mana mereka tidak dilahirkan sebagai warga dari negara tersebut. Ini adalah salah satu langkah paling signifikan menuju integrasi karena memberikan hak dan tanggung jawab yang sama kepada individu sebagai warga negara yang lahir di negara tersebut. Pemain naturalisasi adalah seorang atlet yang lahir di negara lain tetapi kemudian memperoleh kewarganegaraan dari negara yang akan diwakilinya dalam kompetisi olahraga internasional. Proses ini biasanya melibatkan perpindahan kewarganegaraan secara resmi, di mana pemain tersebut memenuhi persyaratan hukum dan administratif untuk menjadi warga negara baru. Di dunia sepak bola, pemain naturalisasi sering kali merupakan pemain yang telah menunjukkan kemampuan dan prestasi tinggi di negara asalnya, dan kemudian ditarik untuk memperkuat tim nasional negara baru. Proses naturalisasi ini sering digunakan untuk meningkatkan kualitas tim nasional, dengan mengintegrasikan pemain-pemain berbakat yang dapat membawa pengalaman dan keahlian tambahan. Di Indonesia naturalisasi pemain asing sudah berlangsung sejak lama, banyak pemain asing yang sebelumnya pernah memperkuat tim nasionalnya atau baru memulai karir internasionalnya telah resmi menjadi warga negara Indonesia dan bergabung dengan tim nasional Indonesia.

Support Vector Machine (SVM) adalah salah satu metode machine learning yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang termasuk dalam kategori *supervised learning*. Dalam prosesnya, metode SVM memiliki tujuan yaitu untuk menemukan *hyperplane* paling optimal yang berfungsi untuk memisahkan dua buah kelas. Tingkat akurasi pada model yang dihasilkan oleh proses klasifikasi dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan.

Dalam melakukan penelitian ini, digunakan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi yang dilakukan dengan garis pembatas (*hyperplane*) yang memisahkan tiga kelas yaitu kelas opini positif, opini netral, dan opini negatif. Dalam penelitian ini, SVM dipilih karena memiliki beberapa kelebihan yang cocok dengan

penelitian ini. Salah satu kelebihan metode SVM adalah bisa menghasilkan model klasifikasi yang baik walaupun hanya dilatih dengan data yang sedikit.

Selain itu, ada alasan lain juga yang menyatakan secara implisit bahwa SVM merupakan metode yang cukup *fleksibel* dan akurat untuk penelitian ini. *Fleksibel* karena proses klasifikasi dengan SVM dapat dilakukan dengan memilih salah satu di antara 4 kernel populer yang tersedia sesuai dengan himpunan data yang dimiliki yaitu *linear*, *polynomial*, RBF, dan *sigmoid* dan dapat memanfaatkan *kernel trick* sampai mendapatkan *hyperplane* yang optimal.

Penelitian ini diharapkan mampu mengidentifikasi dan menganalisis secara komprehensif emosi positif, negatif, dan netral suporter Indonesia terhadap kasus naturalisasi pemain timnas Indonesia. Penelitian Lainnya dilakukan oleh. Analisis sentimen berdasarkan metode SVM, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam memahami pandangan fans Indonesia terhadap kasus naturalisasi pemain timnas Indonesia. Beberapa peneliti sebelumnya telah menerapkan analisis sentimen dalam berbagai konteks, termasuk industri sepak bola. Misalnya, penelitian yang dilakukan oleh Liu et al. (2019) menggunakan analisis sentimen untuk mengukur persepsi penggemar terhadap manajemen klub, sementara penelitian lain yang dilakukan oleh Chen et al. (2020) menggunakan teknik serupa untuk menganalisis reaksi penggemar terhadap performa pemain. Namun belum banyak penelitian yang fokus khusus menganalisis emosi suporter terhadap proses naturalisasi pemain timnas Indonesia.

Berdasarkan uraian latar belakang yang telah dijelaskan peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Sentimen Penggemar Sepak Bola Terhadap Proses Naturalisasi Pemain Timnas Indonesia Pada Media Sosial X Menggunakan *Support Vector Machine*” penelitian ini dilakukan sebagai salah satu bentuk informasi agar melihat tingkat sentiman masyarakat di media sosial X (*Twitter*) terhadap kasus naturalisasi pemain timnas.

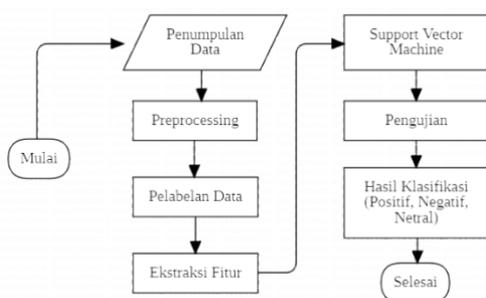
2. METODE

2.1 Waktu dan Tempat Penelitian

Penelitian ini akan dilaksanakan pada bulan April 2024 - Juli 2024 menggunakan media sosial. Penelitian ini bertempat di Laboratorium Penelitian Mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Halu Oleo.

2.2 Prosedur Penelitian

Prosedur penelitian yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Prosedur Penelitian

Berikut penjelasan langka-langkah prosedur penelitian yang berdasarkan pada Gambar 1.

1. Pengumpulan data, Data *tweet* diambil dari media sosial X (*twitter*) dengan teknik *crawling* menggunakan *tools Tweet Harvest* pada ekstensi *Google Notebook Colab* menggunakan kata kunci “*naturalisasi*” dari periode bulan Januari 2024 - Februari 2024. Data *tweet* disimpan dalam bentuk *file csv*.
2. Preprocessing, adalah tahap untuk mengubah kata data mentah menjadi data yang siap digunakan supaya memudahkan dalam proses analisis sentimen. Tahap ini perlu dilakukan karena beberapa kalimat *tweet* yang didapatkan tidak sepenuhnya menggunakan kata baku dan tidak menggunakan bahasa yang baik dan benar. Terdiri dari *case folding*, *cleaning*, *tokenizing*, *filtering*, dan *stemming*.
3. Pelabelan, setelah melalui tahap *preprocessing*, proses selanjutnya adalah memberi label positif, netral, negatif pada data *tweet*. Sebelum dilakukan pelabelan data, dataset hasil *preprocessing* tersebut diterjemahkan terlebih dahulu ke dalam bahasa Inggris menggunakan Google Spreadsheet. Dataset harus diterjemahkan terlebih dahulu karena TextBlob hanya bisa mendeteksi bahasa Inggris. Pada penelitian ini digunakan library Python yaitu TextBlob untuk memberi label positif, netral, dan negatif sesuai dengan polaritas data *tweet*.
4. Ekstraksi fitur, Setelah melalui tahap *preprocessing*, proses selanjutnya adalah memberi label positif, netral, negatif pada data *tweet*. Sebelum dilakukan pelabelan data, dataset hasil *preprocessing* tersebut diterjemahkan terlebih dahulu ke dalam bahasa Inggris menggunakan Google Spreadsheet. Dataset harus diterjemahkan terlebih dahulu karena TextBlob hanya bisa mendeteksi bahasa Inggris. Pada penelitian ini digunakan library Python yaitu TextBlob untuk memberi label

positif, netral, dan negatif sesuai dengan polaritas data *tweet*.

5. Support Vector Machine, selanjutnya pelatihan dan pengujian data dengan metode support vector machine. Dengan metode ini, garis pembatas atau *hyperplane* digunakan untuk memisahkan kelas sentimen yaitu *tweet* positif, netral dan negatif.

2.3 Pengumpulan data

Pada penyusunan penelitian ini, peneliti mengumpulkan data dan informasi yang dapat mendukung proses dalam penelitian yang berkaitan dengan proses pengumpulan data sebagai berikut:

1. Studi literatur
Studi literatur dilakukan dengan cara peneliti mengumpulkan dan mempelajari buku, artikel, dan jurnal ilmiah penelitian sebelumnya yang berkaitan dengan penulisan skripsi seperti teori analisis sentimen, metode *Support Vector Machine*, pengaplikasian bahasa pemrograman Python dalam analisis sentimen. Peneliti juga menggunakan jurnal, e-book, dan website resmi sebagai referensi pada penulisan skripsi ini sehingga membantu peneliti dalam menyelesaikan penelitian ini.
2. Crawling data
Peneliti melakukan crawling data dengan mengambil data dari media sosial X (*Twitter*) tentang *Naturalisasi*. Penarikan data diambil menggunakan *tools Tweet Harvest* menggunakan ekstensi *Google Notebook Colab* bahasa pemrograman Python yang dilakukan di tanggal Januari 2024 - Februari 2024. Pengambilan data ini mendapatkan total data sebanyak 1531 *tweets* dan akan disimpan ke dalam format *Comma Separated Values (CSV)*.

2.4 Pengujian

Setelah dilakukan implementasi maka dilakukan tahap pengujian Tahap pengujian dilakukan untuk tujuan mengetahui bagaimana kinerja metode. Dalam pengujian metode akan menggunakan pengujian *Confussion Matrix*

2.5 Instrumen penelitian

Instrumen yang digunakan pada penelitian ini terbagi menjadi dua, yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*). Instrumen penelitian dapat dilihat pada Tabel 1.

Tabel 1 Perangkat Lunak dan Perangkat Keras

Perangkat Keras	Perangkat Lunak
1. VivoBook_ASUSLaptop X415EA-A416EA, dengan spesifikasi: a. 11th Gen Intel(R) Core(TM) i3-1115G4 @ 3.00GHz b. Penyimpanan 4,00 GB c. Core(TM) i3-1115G4 @ 3.00GHz	1. Sistem operasi : Sistem operasi 64-bit, prosesor berbasis x64 2. Tools desain: Diagrams.net. 3. Bahasa pemrograman : Python 3.9.7 4. IDE : Jupyter Notebook dan Visual Studio Code

3. HASIL DAN PEMBAHASAN

3.1. Implementasi

Tahap implementasi merupakan tahap yang menguraikan tentang pembangunan dan penerapan sistem sesuai dengan rancangan yang sebelumnya telah dibuat.

1. Hasil Preprocessing

Hasil akhir dari proses preprocessing sebagai berikut:

Tabel 2 Hasil Preprocessing

Data 1	Data 2	Data 3
polemik	jerit	main
naturalisasi	hati	naturalisasi
timnas	latih	maju
indonesia	lokal	sepak
bambang	sikap	bola
pamungkas	marak	indonesia
bela	main	
shin	naturalisasi	
taeyong	balik	
	menang	
	untun	
	timnas	
	vietnam	

2. Labeling

Proses pelabelan tweet pada dataset dilakukan secara manual dengan cara memberikan label pada database twitter oleh dosen Bahasa Indonesia yang bernama Harmin, S.Pd., M. Pd dan Kusuma Wardani, S.Pd., M.Pd melihat referensi terhadap penelitian terkait sebelumnya dengan komposisi 80%: 20% dari

1531 tweet. Yang sebelumnya telah dilakukan random di excel sebanyak 10 kali random dengan perbandingan yang berbeda-beda. Untuk setiap tweet diberi label.

Penentuan kelas positif, netral dan negatif didasari oleh nilai polaritas. Teks *tweet* dengan nilai polaritas mengarah ke nilai 1 menunjukkan sentimen kelas positif, nilai polaritas mengarah ke nilai -1 menunjukkan kelas sentimen negatif dan nilai polaritas bernilai 0 masuk kedalam kelas netral. Hasil pelabelan dengan TextBlob dapat dilihat pada Gambar 2.

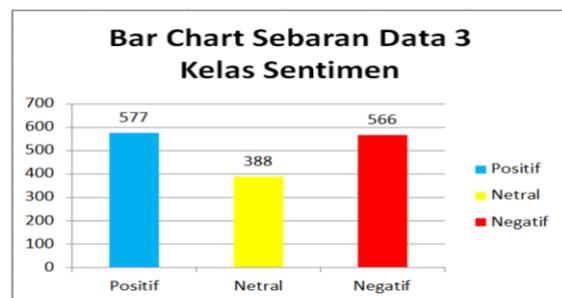
Gambar 2 Hasil Pelabelan dengan TextBlob

Didapatkan hasil akhir dari pelabelan dengan menggunakan library textblob sebanyak 1531 data *tweet* adalah 577 *tweet* yang masuk dalam kelas positif, 388 *tweet* kelas netral, dan 566 *tweet* kelas negatif yang ditunjukkan di Tabel 3.

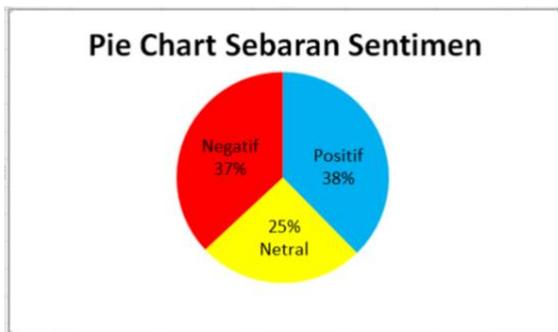
Tabel 3 Hasil Pelabelan dengan TextBlob

Positif	Netral	Negatif	Jumlah
577	388	566	1531

Jumlah sebaran data hasil labeling data dengan TextBlob dapat dilihat pada Gambar 4 dan 5.



Gambar 4 Bar Chart Labeling



Gambar 5 Pie Chart Labeling

3. Ekstraksi fitur

Hal pertama yang dilakukan dalam tahap ini adalah membagi data menjadi dua bagian yaitu data latih dan data uji. Data uji atau *test size* sebanyak 20% dari total keseluruhan data, dan sisanya yaitu 80% data menjadi data latih. Menghasilkan 1224 data latih dan 307 data uji.

Tahap selanjutnya adalah mengubah isi dataset ke dalam representasi *vector* sekaligus menerapkan *N-Gram*. Pada penelitian ini, digunakan *library scikitlearn* untuk mengambil *CountVectorizer* yang dapat mengubah fitur teks menjadi sebuah representasi *vector*, kemudian parameter *N-Gram* akan menyusun kata yang menjadi *bag of words* berbentuk *unigram*. *Unigram* artinya setiap representasi *vector* akan mewakili 1 kata.

4. Implementasi support vector machine

Tahap ini adalah tahap membuat model *Support Vector Machine classifier*. Pada data latih terdapat kelas sentimen positif, netral, dan negatif. SVM akan mempelajari karakteristik kata-kata yang terdapat pada masing-masing kelas pada data latih tersebut, kemudian SVM akan mencoba memprediksi kelas sentimen pada data uji sebanyak 307 data. Pada tahap pemodelan ini, dicoba tiga kernel SVM yaitu Linear, RBF, dan Polynomial dengan nilai *default* pada *C*, *gamma*, dan *degree* untuk mengetahui kernel apa yang paling baik akurasi dalam mengklasifikasikan 1531 data pada penelitian ini.

Tabel 4 Perbandingan Akurasi Tiga Kernal SVM

No	Kernal SVM	Akurasi
1	Linear	83,061%
2	Radial Basis Function	64,495%
3	Polynomial	44,951%

Pada Tabel 4 dapat disimpulkan bahwa Linear SVM dan memiliki performa paling baik dalam menguji data *tweet* pada penelitian ini. Maka dari itu, kernel tersebut adalah kernel yang akan digunakan di penelitian ini untuk evaluasi performansi dan sekaligus menjadi *classifier* untuk aplikasi web analisis sentimen.

5. Evaluasi performansi

Untuk mengetahui performa dari metode *Linear SVM*, maka dilakukan pengujian terhadap model yang telah dibuat. Hasil klasifikasi akan ditampilkan dalam bentuk *confusion matrix*. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Hasil evaluasi model *confusion matrix* dapat dilihat di Tabel 5.

	Prediksi Kelas		
	Negatif	Netral	Positif
Negatif	77	10	9
Netral	8	66	8
Positif	9	8	112

Tabel 5 *confusion matrix*

Model mengklasifikasikan 77 data negatif yang teridentifikasi dengan benar bersentimen negatif, 10 data negatif teridentifikasi netral, dan 9 data negatif teridentifikasi positif. 66 data netral benar teridentifikasi dengan benar bersentimen netral, 8 data netral teridentifikasi negatif, dan 8 data netral teridentifikasi positif. Kemudian 112 data positif teridentifikasi dengan benar bersentimen positif, 9 data positif teridentifikasi negatif, dan 8 data positif teridentifikasi netral.

Setelah mendapatkan hasil *confusion matrix*, selanjutnya dapat dilakukan perhitungan manual nilai akurasi dari model Linear SVM.

Accuracy

$$= \frac{\text{True Positive} + \text{True Neral} + \text{True Negative}}{\text{Total Data Uji}} \times 100\%$$

$$= \frac{255}{307} \times 100\%$$

$$= 83,061\%$$

Accuracy menggambarkan seberapa besar tingkat akurat model yang telah dibuat dapat mengklasifikasi data dengan benar. *Accuracy* didapatkan dari perhitungan rasio data benar dengan keseluruhan data. Dengan mengetahui besarnya nilai akurasi pada kinerja model *machine learning*, dapat diketahui tingkat kemampuan model dalam mencari ketepatan antara informasi yang diinginkan dengan jawaban yang diberikan oleh model. Tingkat keberhasilan sistem dalam menemukan sebuah informasi dalam penelitian ini sebesar 83,061 %. Selanjutnya, untuk melihat nilai performa klasifikasi

dari setiap kelas dapat diketahui melalui nilai *precision*, *recall*, dan *f1-score*.

Berikut perhitungannya:

1. Perhitungan *Precision*

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

$$Precision\ negative = \frac{77}{77 + 17} \times 100\% = 81\%$$

$$Precision\ Netral = \frac{66}{66 + 18} \times 100\% = 78\%$$

$$Precision\ Positive = \frac{112}{112 + 17} \times 100\% = 86\%$$

2. Perhitungan *Recall*

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

$$Recall\ negative = \frac{77}{77 + 19} \times 100\% = 80\%$$

$$Recall\ Netral = \frac{66}{66 + 16} \times 100\% = 80\%$$

$$Recall\ Positive = \frac{112}{112 + 17} \times 100\% = 86\%$$

3. Perhitungan *F1-Score*

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

$$F1 - Score\ Negative = \frac{2 * (0,81 * 0,80)}{0,81 + 0,80} \times 100\% = 80\%$$

$$F1 - Score\ Netral = \frac{2 * (0,78 * 0,80)}{0,78 + 0,80} \times 100\% = 79\%$$

$$F1 - Score\ Positive = \frac{2 * (0,86 * 0,86)}{0,86 + 0,86} \times 100\% = 86\%$$

Nilai *precision* untuk kelas negatif sebesar 81%, untuk kelas netral sebesar 78%, untuk kelas positif sebesar 86%. Angka ini dapat diartikan bahwa proporsi presisi label yang diprediksi dengan benar dari total prediksi cukup tinggi untuk kelas positif dan negatif. Sedangkan *recall* untuk kelas negatif sebesar 80%, untuk kelas netral sebesar 80%, dan kelas positif sebesar 86%. Hal ini berarti keberhasilan kinerja sistem dalam menemukan kembali informasi yang bernilai positif dalam dokumen lebih baik dibandingkan dengan menemukan informasi kembali yang bernilai netral dan negatif. Sementara *F1-Score* bisa diartikan sebagai *harmonic mean* (rata-rata yang dihitung dengan cara mengubah semua data menjadi pecahan) dari *precision* dan *recall*. *F1-*

Score yang baik mengindikasikan bahwa model klasifikasi memiliki *precision* dan *recall* yang baik.

3.2. Uji Coba dan Pembahasan Sistem

Tahap ini berisi pengujian kinerja sistem beserta pembahasannya.

1. Antarmuka aplikasi

Aplikasi web yang dibuat di penelitian ini memiliki 3 halaman yaitu Beranda, *Confusion Matrix*, dan Sebaran Data seperti yang bisa dilihat di Gambar 8.

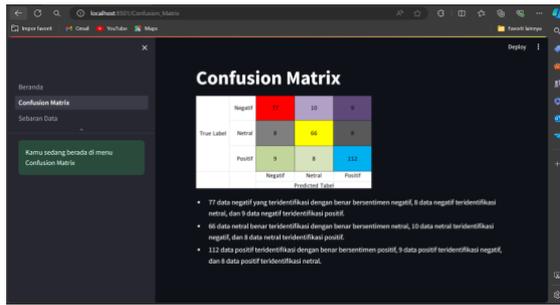


Gambar 8 Halaman Beranda

Pada menu Beranda, pengguna akan disambut dengan sebuah *text field* berfungsi untuk memasukkan teks untuk menguji kemampuan model SVM dalam mengklasifikasikan teks ke dalam salah satu dari tiga kelas sentimen.

Aplikasi web ini merupakan sebuah sistem *sentiment analyzer* yang bisa mendeteksi sentimen dari teks Bahasa Indonesia yang dimasukkan oleh pengguna. Di dalam *back end* aplikasi web tersebut, sudah dimasukkan model *machine learning Linear SVM* beserta pustakan ekstraksi fitur yaitu *CountVectorizer* dan *N-Gram* yang dilatih dan diuji menggunakan data *tweet* yang menghasilkan akurasi sebesar 83,061%.

Ketika pengguna menekan tombol “Klasifikasi sentimen teks”, maka teks Bahasa Indonesia yang sudah dimasukkan tersebut akan langsung diterjemahkan oleh sistem ke dalam bahasa Inggris menggunakan *library googletrans*. Teks tersebut harus diterjemahkan terlebih dahulu sebelum diklasifikasikan oleh model karena model dibuat menggunakan dataset berbahasa Inggris. Lalu, model akan memasukkan teks yang sudah diterjemahkan tersebut ke dalam 3 kelas sentimen yaitu positif, negatif atau netral kemudian hasilnya akan ditampilkan.



Gambar 9 Halaman *Confusion Matrix*

Kemudian pada menu *Confusion Matrix* di Gambar 9, disediakan tampilan *confusion matrix* dari model Linear SVM yang sudah melalui proses pengujian.



Gambar 10 Halaman Sebaran Data

Kemudian halaman terakhir bernama "Sebaran Data" yang bisa di lihat di Gambar 10 adalah halaman yang memuat infografis tentang penyebaran sentimen data *tweet* pada dataset yang berisi 1531 *tweet* selama bulan Januari dan April 2024. Data divisualisasikan menggunakan *Bar Chart* dan *Pie Chart*.

2. Hasil uji coba

Pada tahap ini, sistem menjalani uji coba untuk mengklasifikasikan teks yang dimasukkan pengguna ke dalam *text field*. Pertama, sistem akan mencoba untuk mendeteksi kalimat bersentimen negatif seperti yang ada pada Gambar 11.



Gambar 11 Input Teks dengan Hasil Klasifikasi Negatif

Sistem berhasil mengkalifikasikan teks tersebut ke dalam kelas "negatif" untuk kalimat "kalau yang

ngomong Arya Sinulingga mah sulit dipercaya. Apalagi soal duit. Kalau soal negosiasi naturalisasi dia bisa dipercaya". Kemudian uji coba selanjutnya adalah memasukkan kaliman bernada netral. Kaliman bersentimen netral adalah kalimat yang tidak mengandung unsur pro atau kontra terhadap suatu topik. Uji teks netral ditunjukkan di Gambar 12.



Gambar 12 Input Teks dengan Hasil Klasifikasi Netral

Pada Gambar 12, sistem berhasil mengkalsifikasikan kalimat netral ke dalam kelas netral. Selanjutnya sistem diuji dengan teks bernada positif. Menurut hasil pengujian, model SVM dalam penelitian ini memiliki performa paling tinggi dalam mendeteksi data bersentimen positif. Kali ini sistem diuji dengan kalimat "kalau STY diganti pemain diasporanya masih dipakai gak ya? Sayang banget mereka hanya dipakai di jaman STY aja. Soalnya bisa saja pengganti STY bukan penggemar pemain diaspora atau naturalisasi" yang bisa dilihat di Gambar 13.



Gambar 13 Input Teks dengan Hasil Klasifikasi Positif

Dengan aplikasi web *sentiment analyzer* ini, model SVM Linear dapat mengklasifikasikan ketiga teks ke dalam tiga sentimen yang sesuai. Berikut rinciannya.

- Negatif" untuk kalimat "kalau yang ngomong Arya Sinulingga mah sulit dipercaya. Apalagi soal duit. Kalau soal negosiasi naturalisasi dia bisa dipercaya = **Negatif**
- Ternyata ada juga seleksi pemain keturunan yang masih butuh proses naturalisasi = **Netral**
- kalau STY diganti pemain diasporanya masih dipakai gak ya? Sayang banget mereka hanya

dipakai di jaman STY aja. Soalnya bisa saja pengganti STY bukan penggemar pemain diaspora atau naturalisasi = **Positif**

6. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan antara lain: Penelitian ini memberikan hasil analisis sentimen pada topik “naturalisasi” di media sosial X (*twitter*) menggunakan metode *Support Vector Machine* yang mampu mengklasifikasikan *tweet* ke dalam kelas sentimen positif, netral, dan negatif. Dihasilkan sebuah aplikasi web yang memiliki model *machine learning* Linear SVM di dalam *backend* yang dapat mengklasifikasikan teks yang dimasukkan oleh pengguna ke dalam kelas positif, netral, atau negatif. Hasil akurasi pengujian klasifikasi dengan Linear *Support Vector Machine* adalah 83,061%. Sentimen negatif berjumlah sedikit bisa dikarenakan TextBlob yang mendeteksi kata “naturalisasi” menjadi positif.

Dalam penelitian selanjutnya disarankan menggunakan metode *machine learning* lain terutama metode yang didesain untuk klasifikasi multiclass sehingga bisa digunakan sebagai perbandingan hasil uji model untuk mencari metode klasifikasi terbaik.

7. DAFTAR PUSTAKA

- [1] Abdusyukur, F. (2023). Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter. *Komputa: Jurnal Ilmiah Komputer Dan Informatika*, 12(1), 73–82. <https://doi.org/10.34010/komputa.v12i1.9418>.
- [2] Agung Prabowo, D., & Sudianto. (2023). Analisis Sentimen Sepak Bola Indonesia pada Twitter menggunakan K-Nearest Neighbors dan Random Forest. *JSAI (Journal Scientific and Applied Informatics)*, 6(2), 217–227. <https://doi.org/10.36085/jsai.v6i2.5337>
- [3] Aji, R. N. B. (2013). Nasionalisme Dalam Sepak Bola Indonesia Tahun 1950-1965. *Lembaran Sejarah*, 10(2), 137
- [4] Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), 267–281.
- [5] Bimasakti, M. A. (2023). Problematik Konstitusionalitas Naturalisasi di Indonesia. *Pancasila: Jurnal Keindonesiaan*, 3(1), 100–112. <https://doi.org/10.52738/pjk.v3i1.149>
- [6] Dalifah, N., Suarna, N., & Prihartono, W. (2024). Analisis Data Sentimen Negatif Pada Opini Pengguna Twitter Terhadap Berita Sepak Bola Liga 1 Tahun 2022 Dengan Penerapan Support Vector Mechine. 8(1), 209–214.
- [7] Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *EduTic - Scientific Journal of Informatics Education*, 7(1), 1–11. <https://doi.org/10.21107/edutic.v7i1.8779>.
- [8] Fitriyah, N., Warsito, B., & Maruddani, D. A. I. (2020). Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm). *Jurnal Gaussian*, 9(3), 376–390. <https://doi.org/10.14710/j.gauss.v9i3.28932>
- [9] Hasibuan, N. A., Silalahi, N., Nasution, S. D., Sutiksno, D. U., Nurdianto, H., Buulolo, E., Ambon, P. N., Pendahuluan, I., & Mining, A. D. (2017). Implementasi Data Mining Untuk Pengaturan Layout. 4(4), 6–11.
- [10] Ikhsan Yusuf, M., Ransi, N., Arman, A., Tenriawaru, A., Saidi, L. O., & La Surimi, L. S. (2022). Analisis Sentimen Komentar Netizen Instagram Terhadap Racism Di Sepak Bola Indonesia Dengan Metode Naive Bayes. *Jurnal Matematika Komputasi Dan Statistika*, 2(3), 136–143. <https://doi.org/10.33772/jmks.v2i3.10L>. Liu and H. Miao, "A specification based approach to testing polymorphic attributes," in *Formal Methods and Software Engineering: the 6th Int. Conf. on Formal Engineering Methods, ICFEM 2004, Seattle, USA, Nov. 2004*, J. Davies, W. Schulte, M. Barnett, Eds. Berlin: Springer, 2004. pp. 306-19.
- [11] Indriati, I., Marji, M., & Pakpahan, S. (2019). Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(7), 7259–7267.
- [12] Jon, A. M. (2022). Analisis Sentimen Pada Media Sosial Instagram Klub Persija Jakarta Menggunakan Metode Naive Bayes. 958.
- [13] Kadarsih, K., & Andrianto, S. (2022). Membangun Website SMA PGRI Gunung Raya Ranau Menggunakan PHP dan MYSQL. *JTIM: Jurnal Teknik Informatika Mahakarya*, 03(2), 37–44
- [14] Kholilullah, M., Hayati, U., Informatika, T., Informatika, M., & Cirebon, K. (2024). Analisis Sentimen Pengguna Twitter (X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes. 8(1), 392–398
- [15] Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 200. <https://doi.org/10.30645/j-sakti.v2i2.83>

- [16] Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada Pt. Pertamina (Persero). *Jurnal Teknik Informatika*, 13(2), 1–8.
<https://ejurnal.poltekpos.ac.id/index.php/informatika/article/view/1383>.
- [17] Rifaldi, D., Abdul Fadlil, & Herman. (2023). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.” *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171.
<https://doi.org/10.51454/decode.v3i2.131>.
- [18] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
<https://doi.org/10.31294/ijcit.v5i1.7951>
- [19] Setiawan, H., & Zufria, I. (2023). Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 7(3), 1003–1012.
<https://doi.org/10.30865/mib.v7i3.6144>.