

# Implementasi *Text Mining* Pada Analisis Sentimen Pemain Naturalisasi Timnas Indonesia Dengan Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*

Fajar Tri Widodo<sup>1\*</sup>, Moh. Bhanu Setyawan<sup>1</sup>

<sup>1</sup>Program Studi Teknik Informatika Fakultas Teknik

Universitas Muhammadiyah Ponorogo

Jl. Budi Utomo No. 10, Ronowijayan, Siman, Ponorogo, Jawa Timur  
Indonesia

\*Corresponding Author Email: fajartriwido16@gmail.com

*Abstract:* - Sepak bola adalah olahraga paling populer dan banyak dimainkan diseluruh dunia, termasuk Indonesia. FIFA menjadi sebuah lembaga tertinggi yang mengatur segala urusan sepak bola. FIFA banyak memiliki agenda kompetisi sepak bola seperti piala dunia. PSSI sebagai anggota dari FIFA berkesempatan untuk turut serta dalam turnamen yang diadakan oleh FIFA. Piala dunia menjadi mimpi yang terus didambakan oleh seluruh pecinta sepak bola tanah air. Salah satu strategi federasi sepak bola Indonesia yaitu PSSI untuk meningkatkan prestasi tim nasional adalah melalui program naturalisasi pemain asing. Tujuan dari program tersebut tidak lain adalah untuk memperkuat performa timnas agar bisa bersaing untuk merebutkan asa tampil di ajang piala dunia 2026. Program naturalisasi ini memicu berbagai respons dari masyarakat, terutama melalui media sosial twitter. Opini publik yang disampaikan melalui platform tersebut dapat dianalisis untuk mengetahui sentimen masyarakat terhadap kebijakan naturalisasi. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap cuitan pengguna Twitter mengenai pemain naturalisasi timnas Indonesia dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*. Proses analisis mencakup tahapan pengumpulan data, pra-pemrosesan, dan klasifikasi sentimen menjadi dua kategori sentimen, yaitu positif dan negatif. Hasil penelitian menunjukkan *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi dibandingkan *Support Vector Machine*. Akurasi *Naïve Bayes* sebesar 76,23%, sedangkan SVM memiliki tingkat akurasi sebesar 70,29%. Selain itu, model *Naïve Bayes* juga lebih unggul dalam nilai *precision*, *recall*, dan *f1-score* dibandingkan dengan SVM. *Naïve bayes* memiliki nilai 78,43% untuk *precision*, 75,47% untuk *recall*, dan 76,93% untuk *f1-score*. Sementara itu, SVM memiliki nilai presisi sebesar 76,74%, *recall* sebesar 62,26%, dan *f1-score* sebesar 68,73%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode *Naïve Bayes* lebih efektif dalam mengklasifikasikan sentimen masyarakat terhadap program naturalisasi pemain timnas Indonesia.

*Key-Words:* - Analisis Sentimen, Timnas Indonesia, Naturalisasi, *Naïve Bayes*, *Support Vector Machine (SVM)*, Evaluasi Model

Received: July 18, 2025. Revised: July 27, 2025. Accepted: July 29, 2025. Published: August 10, 2025

## 1 Pendahuluan

Sepa kbola merupakan salah satu cabang olahraga terpopuler dan paling banyak diminanti di seluruh dunia, termasuk Indonesia [1]. Sepak bola dimainkan secara beregu yang berisikan sebelas pemain inti dan memiliki pemain cadangan. Induk sepak bola dunia yang memiliki wewenang untuk mengelola dan menaungi persepak bolaan dunia adalah FIFA (*Federation International de Football Association*). Sementara itu, PSSI (Persatuan Sepak Bola Seluruh Indonesia) menjadi induk sepak bola nasional. Sebagai induk sepak bola tertinggi, FIFA memiliki berbagai macam kompetisi sepak bola yang bisa diikuti oleh semua negara yang tergabung sebagai anggotanya.

PSSI sebagai anggota FIFA bisa berpartisipasi pada ajang yang diadakan oleh FIFA, seperti kualifikasi piala dunia dan juga piala dunia.

Sebelum mengikuti piala dunia, setiap peserta harus mengikuti ajang kualifikasi. Indonesia total telah mengikuti fase kualifikasi sebanyak 15 kali [2]. Namun, dari beberapa kualifikasi yang telah diikuti, Indonesia belum pernah lolos ke fase berikutnya dan menjadi peserta pada ajang piala dunia. Lolos dan berlaga ke piala dunia menjadi impian bagi seluruh pecinta sepak bola Indonesia. Berbagai cara telah dilakukan oleh PSSI untuk bisa mewujudkan mimpi masyarakat Indonesia. Berbagai upaya telah dilakukan untuk meningkatkan kualitas sepak bola nasional seperti *upgrade* liga, TC pemain timnas, serta yang

paling populer dan membuat Indonesia disorot dunia adalah program naturalisasi. Naturalisasi merupakan sebuah proses perpindahan status kewarganegaraan, dimana penduduk asing akan mendapatkan status kewarganegaraan apabila telah memenuhi persyaratan dan prosedur naturalisasi [3]. Adanya program naturalisasi bertujuan untuk memajukan persepakbolaan nasional dan mewujudkan mimpi timnas Indonesia berlaga di piala dunia.

Program naturalisasi menuai berbagai macam reaksi publik dari berbagai kalangan masyarakat. Banyak yang berpendapat setuju dan tidak setuju terhadap naturalisasi. Salah satu tokoh yang berargumen terkait program naturalisasi adalah anggota DPR RI, ibu Anita Jacoba Gah. Anggota dewan tersebut berpendapat bahwa negara Indonesia tidak miskin atlet, sehingga PSSI harus memanfaatkan produk lokal bangsa Indonesia, karena negara ini memiliki banyak atlet yang berkualitas [4]. Tentunya pendapat tersebut dapat menjadi gambaran bagaimana persepsi masyarakat umum terhadap naturalisasi timnas Indonesia. Derasnya pertumbuhan teknologi informasi, banyak bermunculan *platform* komunikasi seperti media sosial.

Dengan media sosial, semua orang dapat menyuarakan segala komentar, baik itu komentar yang mengarah untuk hal positif maupun negatif. Menurut situs [databoks.katadata.co.id](http://databoks.katadata.co.id), pengguna media sosial pada tahun 2024 mencapai 191 juta pengguna, atau sekitar 73,7% dari jumlah populasi. Salah satu *platform* media sosial yang paling populer untuk menuangkan pendapat dan opini dalam bentuk cuitan adalah twitter. Twitter merupakan aplikasi media sosial yang memungkinkan pengguna bisa melakukan kegiatan bertukar informasi seperti mengirim dan membaca pesan berbasis teks, gambar dan juga video [5].

Dalam penggunaan media sosial twitter, tidak jarang *netizen* memberikan pendapat dan kritikan yang dinilai tidak baik. Mereka dengan bebas dapat menuangkan segala ‘uneg-uneg’ berdasarkan kejadian yang sedang ramai dibincangkan, termasuk program naturalisasi pemain timnas sepak bola Indonesia. Beberapa ulasan dari *netizen* ini dapat diketahui, apakah ulasan mereka mengarah ke hal positif, negatif, atau bahkan netral. Banyak tanggapan bermunculan terkait program naturalisasi pemain timnas Indonesia. Dalam menanggapi sebuah opini, kita juga harus membedakan tanggapan positif maupun negatif, bahkan yang bernada netral sekalipun. Untuk mengetahui jenis tanggapan tersebut, bisa dilakukan dengan melakukan analisis sentimen.

Analisis sentimen merupakan cara yang digunakan untuk melakukan analisis teks digital dan untuk mengumpulkan data berupa nada pendapat

emosional orang terhadap suatu kejadian tertentu, baik itu positif, negatif, atau netral [6]. Analisis sentimen termasuk dalam bidang *Natural Language Processing* (NLP). Analisis sentimen bekerja untuk mengenali dan melakukan ekstraksi sebuah opini dalam bentuk teks. Dengan memanfaatkan bantuan dari analisis sentimen, informasi yang tidak terstruktur dapat menjadi sebuah data yang terstruktur.

Dalam menerapkan penelitian terkait sentimen, terdapat beberapa metode yang sering digunakan. Beberapa metode yang sering digunakan antara lain seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Decision Tree*, dan sebagainya. Dalam penelitian ini, akan menggunakan metode yang memiliki tingkat akurasi tinggi, yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*. Metode *Naïve Bayes* dan SVM memiliki akurasi yang berbeda. Metode *Naïve Bayes* tergolong dalam metode yang mudah digunakan, sehingga metode ini banyak dipergunakan dalam penelitian yang berkaitan dengan analisis sentimen. Selain mudah digunakan, metode *Naïve Bayes* dapat diterapkan pada domain yang berbeda. Metode *Support Vector Machine (SVM)* merupakan sebuah peningkatan margin yang terletak dalam dua kelas yang berbeda.

Banyak penelitian terdahulu yang menggunakan kedua metode tersebut. Pada penelitian yang dilakukan oleh Dea Nurmatin Novianti, dkk dengan judul penelitiannya *Comparison of Support Vector Machine dan Naïve Bayes Algorithms for Sentiment Analysis of the Metaverse*, menghasilkan metode *Support Vector Machine* memiliki tingkat akurasi dan keakuratan lebih tinggi yaitu sebesar 90,32%, sedangkan metode *Naïve Bayes* memiliki nilai akurasi dan keakuratan lebih rendah sebesar 84,23% [7]. Penelitian lainnya yang dilakukan oleh Suryani, dkk., dengan judul penelitian *Sentiment Analysis of Towards Electric Cars using Naïve Bayes Classifier dan Support Vector Machine Algorithm*, sebuah penelitian yang membahas tentang sentimen terkait mobil elektrik menghasilkan nilai akurasi metode *Support Vector Machine* sebesar 90%, dan metode *Naïve Bayes* sebesar 88% [8].

Pada penelitian ini, penulis akan melakukan perbandingan tingkat akurasi antara model *Naïve Bayes* dan *Support Vector Machine (SVM)*. Kasus yang diangkat dalam penelitian ini adalah ulasan atau opini terkait program naturalisasi timnas Indonesia pada platform media sosial Twitter (sekarang dikenal dengan nama X). Berdasarkan latar belakang di atas, maka penulis merumuskan sebuah penelitian yang berjudul, “Implementasi Text Mining Pada Analisis Sentimen Pemain Naturalisasi Timnas Indonesia Dengan Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*”. Berdasarkan judul tersebut, diharapkan akan mendapatkan sebuah nilai akurasi

yang tepat dari kedua metode tersebut dalam menentukan sentimen masyarakat yang berkaitan dengan tweet positif dan negatif dari program naturalisasi timnas Indonesia.

## 2 Metode Penelitian

### 2.1 Tahapan Penelitian

Pada tahapan penelitian ini, akan disajikan alur tahapan atau langkah-langkah yang dilakukan penelitian yang akan dijalankan. Tahapan penelitian digunakan sebagai prosedur yang menjelaskan proses berjalannya sebuah penelitian. Gambar 2.1 merupakan alur dari tahapan penelitian yang telah dirancang:



Gambar 2. 1 Alur Penelitian

#### 2.1.1 Studi Literatur

Dalam tahap studi literatur dilakukan pencarian semua referensi yang memiliki keterkaitan materi dengan topik penelitian yang diangkat. Studi literatur didapatkan dari berbagai macam sumber informasi seperti artikel, jurnal, serta penelitian terdahulu dimana pembahasan materinya memiliki topik bahasan yang relevan.

#### 2.1.2 Pengumpulan Data

Pengumpulan data merupakan langkah untuk mendapatkan data yang nantinya akan dilakukan pengolahan. Data didapatkan dari cuitan pengguna media sosial twitter.

#### 2.1.3 Preprocessing Data

Tahap selanjutnya dari rangkaian penelitian ini adalah *preprocrssing* data. Pada tahap ini, akan

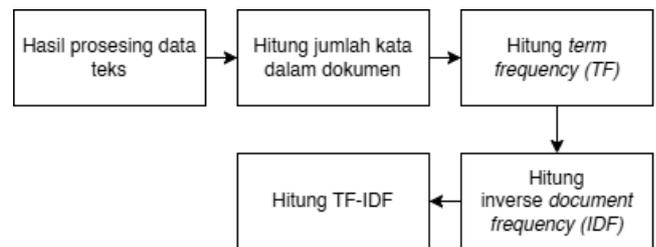
melakukan pemrosesan data untuk mengubah data mentah menjadi data terstruktur. Tahap *preprocrssing* terdiri dari beberapa langkah seperti *case folding*, *cleaning*, *tokenizing*, *stopword removal*, serta *stemming* [9]. Berikut merupakan alur dari tahap *preprocessing* pada penelitian ini:



Gambar 2. 2 Tahapan *Preprocessing* Teks

#### 2.1.4 Ekstraksi Fitur

Tahap selanjutnya adalah ekstraksi fitur. Pada tahap ini bertujuan untuk mengkonversi atau melakukan perubahan pada sebuah data teks kedalam bentuk numerik agar nantinya dapat diproses oleh medel. Teknik yang digunakan untuk melakukan ekstraksi fitur adalah dengan *Term Frequency-Inverse Document Frequency* atau bisa disingkat dengan TFIDF [10]. Teknik ini akan mengubah sebuah teks menjadi vektor berdasarkan pentingnya kata dalam sebuah dokumen. Dalam ekstraksi fitur ini terdapat tahapan atau alur perhitungan yang harus dilalui seperti yang telah disajikan pada gambar 2.3:



Gambar 2. 3 Tahap Perhitungan Ekstraksi Fitur

Dalam menghitung TFIDF, terdapat rumus yang memudahkan dalam proses perhitungan. Berikut ini merupakan rumus untuk ekstraksi fitur dengan TFIDF:

- a. Rumus menghitung nilai *Term Frequency* (TF):

$$tf = (t, d) = \frac{\text{jumlah kata } t}{\text{jumlah total kata dalam dokumen}} \quad (1)$$

- b. Rumus menghitung nilai *Inverse Document Frequency* (IDF):

$$idf = (t) \log_{10} = \frac{\text{jumlah dokument dalam } d}{\text{jumlah dok. dimana kata } t \text{ muncul}} \quad (2)$$

- c. Rumus menghitung TFIDF:

$$tfidf = tf * idf \quad (3)$$

#### 2.1.5 Training Model

Setelah tahap ekstraksi fitur berhasil dilakukan, selanjutnya masuk ke dalam tahap *training model* atau pelatihan model. Tahap ini bertujuan untuk melakukan pengklasifikasian dengan metode *Naïve Bayes* dan *Support Vector Machine*. Model akan diklasifikasikan ke dalam sentimen positif dan negatif.

Pada pelatihan model, data akan dibagi menjadi data latih dan data uji. Jumlah presentase perbandingan antara data latih dan data uji sebesar 90:10, dimana 90% untuk data latih, serta 10% untuk data uji.

### 2.1.6 Evaluasi Model

Tahap evaluasi model bertujuan untuk menilai kinerja dari model yang digunakan. Evaluasi model ini dilakukan dengan metrik evaluasi. Tahap ini penting untuk dilakukan guna mendapatkan informasi terkait keberhasilan model dalam melakukan prediksi sesuai dengan apa yang diharapkan. Dalam melakukan evaluasi model ini, menggunakan *Confusion Matrix*. *Confusion Matrix* memberikan nilai untuk *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Terdapat beberapa ukuran yang bisa dijadikan gambaran objektif terhadap kinerja model. Beberapa ukuran yang digunakan antara lain *accuracy*, *precision*, *recall*, dan *F1-score*.

*Accuracy* merepresentasikan total data sentimen yang benar pada saat dikenali. Untuk perhitungan nilai *accuracy* ini, langkah pertama yang harus dilakukan adalah dengan membagi jumlah data sentimen yang benar dengan total data dan data uji. Berikut ini persamaan rumus yang digunakan untuk menghitung nilai akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

*Precision* atau presisi mengukur seberapa akurat model dalam memberikan prediksi yang benar untuk kelas positif dibandingkan dengan seluruh jumlah prediksi positif yang telah dibuat. Adapun rumus untuk menghitung presisi atau *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

*Recall* akan bekerja untuk mendeteksi seberapa banyak data positif yang berhasil terdeteksi. Berikut ini merupakan rumus dari *recall*:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

*F1-Score* mencerminkan hubungan antara nilai *precision* dengan *recall*. Berikut merupakan rumus dari *F1-score*:

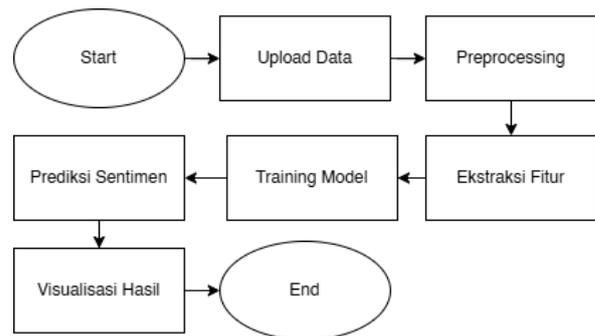
$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

### 2.1.7 Perbandingan Algoritma

Pada tahap ini akan membandingkan kedua metode yang digunakan untuk analisis sentimen. Untuk melakukan perbandingannya, akan dilakukan perbandingan dari hasil evaluasi model untuk setiap nilai matrik seperti *accuracy*, *precision*, *recall*, dan *f1-score*. Tujuan dari perbandingan ini adalah mampu mengetahui metode mana yang memiliki tingkat akurasi lebih tinggi dalam kasus analisis sentimen pemain naturalisasi Timnas Indonesia.

### 2.1.8 Implementasi Sistem

Pada tahap implementasi sistem bertujuan untuk mengembangkan sistem analisis sentimen berbasis web. Tujuan dari implementasi sistem ini adalah dapat melakukan serangkaian proses analisis sentimen secara langsung melalui sebuah tampilan *user interface*. Dalam sistem ini, terdapat beberapa fitur yang akan disediakan seperti fitur *upload file*, melakukan pemrosesan data, *training data*, serta menampilkan hasil sentimen dalam bentuk visualisasi seperti grafik *pie chart*. Berikut ini disajikan alur atau *flowchart* dari sistem yang akan dibangun:



Gambar 2. 4 *Flowchart* Sistem

Gambar 2.4 menyajikan alur jalannya sistem yang akan dirancang. Berikut merupakan penjelasan dari beberapa langkah dari sistem yang akan dirancang:

1. *Upload Data*  
Pada tahap ini, berfungsi agar pengguna mengunggah dataset yang akan diujikan. Dataset yang digunakan adalah sebuah file dengan format csv.
2. *Preprocessing*  
Setelah data berhasil diunggah, selanjutnya masuk ke tahap *preprocessing*. Tahap ini akan menjalankan serangkaian pemrosesan data untuk mengolah dataset yang telah diunggah.
3. *Ekstraksi Fitur*  
Setelah pemrosesan, sistem bisa lanjut ke tahap ekstraksi fitur. Pada tahap ini, nantinya akan digabungkan dengan bagian *training model*.
4. *Training Model*

Pada tahap *training* ini, data yang sudah diekstraksi akan dilatih. Pada *training* ini, akan memproses data latih dan akan menghasilkan perbandingan performa kedua model.

### 5. Prediksi

Pada bagian prediksi sentimen akan menampilkan hasil prediksi data uji kedua model terhadap data latih yang telah dibagi. Kedua model akan menunjukkan hasil prediksi sentimen terhadap data latih sebelumnya.

### 6. Visualisasi Hasil

Pada visualisasi hasil akan menampilkan grafik dalam bentuk *pie chart* perbandingan akurasi dari kedua model.

## 3 Hasil dan Pembahasan

### 3.1 Pengumpulan Data

Dataset yang dikumpulkan berupa *tweet* atau cuitan pengguna twitter yang berkaitan dengan program ataupun pemain naturalisasi Timnas Indonesia. Teknik yang digunakan dalam pengumpulan data adalah dengan *web scrapping*. Pengumpulan data dilakukan dengan memanfaatkan alat Google Colaboratory dengan bahasa pemrograman python. Dalam proses pengumpulan data dengan python, memanfaatkan pustaka *Tweet harvest*. Hasil pengumpulan data akan disimpan sebagai file dalam format *Comma Separated Values* (CSV). Gambar 3.1 berikut merupakan dataset yang telah terkumpul:

nomor	full_text
1	@FaktaSepakbola Nathan sama struick masih dipanggil padahal kemarin udh banyak banget netia
2	@FaktaSepakbola Berikut daftar 27 pemain yang dipanggil Kluivert <a href="https://t.co/MJ5dSt4R2b">https://t.co/MJ5dSt4R2b</a>
3	@FaktaSepakbola Nathan dipanggil karena timnas sangat butuh gelandang pengangkut ultras seb
4	@FaktaSepakbola Sty mematikan karir elkan memang elkan bunuh org apa? Buzzzer bloom @Coa
5	@FaktaSepakbola Yakob ga ada wkwk
6	@FaktaSepakbola Gilak Nathan yg gak pernah main di klub tetep aja dipanggil. Ya mending pang
7	@FaktaSepakbola Setiap panggilan selalu ada yg nanya ini itu. Lebih parah lagi ada yg nanya Elkar
8	@FaktaSepakbola Katanya prioritas menit bermain tp kok ada minim menit masih dipanggil
9	@FaktaSepakbola Septian Bagaskara udh lama gak maen di timnas ya. Lg bagus juga di dewa. God
10	@FaktaSepakbola ATLET PRO LIGA KUAT BANGET SELALU DI PANGGIL
11	@FaktaSepakbola Gada witan yak
12	@FaktaSepakbola Sayuri????
13	@FaktaSepakbola Yakob gak nih? Ferrari? Sure?
14	@FaktaSepakbola Mana SAYURI????????
15	@FaktaSepakbola Ricky kambuya semoga dapat menit bermain di timnas
16	@FaktaSepakbola Kadang bingung yang hate Arhan ini dia nonton bola atau nonton drama zizah
17	@FaktaSepakbola Menurut kamu siapa aja 3 pemain yang bakal dicoret? <a href="https://t.co/wEO4isZW">https://t.co/wEO4isZW</a>
18	@FaktaSepakbola duo sayuri gak ada

Gambar 3. 1 Dataset Cuitan

Setelah data terkumpul, selanjutnya dilakukan pelabelan sentimen secara manual. Gambar 3.2 menampilkan data setelah diberikan label sentimen secara manual:

nomor	full_text	sentimen
1	@detikcom #masukjadulu. ga ada rugi nya... jeda internasional sambil holiday. walo harus pake Positif	Positif
2	@detikcom ... salah satu alasan gw males nonton timnas...	Negatif
3	@detikcom @msaid_didu Coba tanya @erickthohir kenapa?	Negatif
4	@detikcom @msaid_didu itu nyindir kerjaan pemerintah woi gk malu kalian demi bangsa dan ne Negatif	Negatif
5	@detikcom @msaid_didu Siapa dulu ketua @PSSI nya	Positif
6	@detikcom @PSSI ITU FAKTA <a href="https://t.co/bFMUGyPx10">https://t.co/bFMUGyPx10</a>	Positif
7	@detikcom 300 juta dpt pemimpin bangsa sing mbokneancok kabeh fuck	Negatif
8	@detikcom 300 juta penduduk nyari 11 orang kok ke belanda ya??	Negatif
9	@detikcom 300 juta tolong jangan ajari kami cari pemain sepakbola kampun kesukitan. Tapi cari Negatif	Negatif
10	@detikcom Akibat dijahah.	Positif

Gambar 3. 2 Pelabelan Dataset

### 3.2 Preprocessing

Pada tahap *preprocessing* data, dilakukan proses pengelolaan data setelah data terkumpul. Pada tahap ini, akan melakukan pemrosesan data untuk mengubah data mentah menjadi data terstruktur. Terdapat beberapa tahapan yang harus dilalui untuk melakukan pemrosesan pada teks. Tahapan pemrosesan data pada penelitian ini adalah sebagai berikut [20]:



Gambar 3. 3 Tahapan *Preprocessing*

Gambar 3.3 menampilkan serangkaian langkah dalam pemrosesan data. Terdapat tahapan yang terurut mulai dari *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.

#### 1. Case Folding

Tahapan *case-folding* berfungsi untuk memberikan perubahan pada teks yang memiliki huruf kapital, agar diubah menjadi huruf non kapital atau huruf kecil. Berikut merupakan hasil dari *case folding*:

Tabel 3. 1 Hasil *Case Folding*

N	Sebelum <i>Case-folding</i>	Hasil <i>Case-folding</i>
1	@detikcom Kalian punya 300 Juta Penduduk kenapa pengurus Danantara dari asing? <a href="https://t.co/uMUhj7LNvA">https://t.co/uMUhj7LNvA</a>	@detikcom kalian punya 300 juta penduduk kenapa pengurus danantara dari asing? <a href="https://t.co/umuhj7lInva">https://t.co/umuhj7lInva</a>
2	@detikcom Karena PSSI nya bagus <a href="https://t.co/2M2yEpsSIJ">https://t.co/2M2yEpsSIJ</a>	@detikcom karena pssi nya gizinya tak bagus <a href="https://t.co/2m2yepsslj">https://t.co/2m2yepsslj</a>
3	@detikcom Demi popularitas nasionalisme abaikan nasionalisme	@detikcom demi popularitas abaikan nasionalisme
4	@detikcom Apa bapak gak tau kalo timnas kita ini adalah timnas	@detikcom apa bapak gak tau kalo timnas kita ini adalah timnas

	Belanda asia	cabang	belanda asia	cabang
5	@idextratime orang gada 20orang bener2 jago bola?	270jt Indonesia yg	@idextratime 270jt indonesia 20orang bener2 bola?	orang gada yg jago

## 2. *Cleaning*

Tahap *cleaning* berfungsi untuk menghapus dan juga menghilangkan karakter khusus seperti url akun, emoji, angka, dan sebagainya. Berikut merupakan hasil dari tahap *cleaning*:

Tabel 3. 2 Hasil *Cleaning*

No	Sebelum <i>Cleaning</i>	Hasil <i>Cleaning</i>
1	@detikcom kalian punya 300 juta penduduk kenapa pengurus danantara dari https://t.co/umuhj7lnva	kalian punya juta penduduk kenapa pengurus danantara dari asing?
2	@detikcom karena pssi nya gizinya tak bagus https://t.co/2m2yepsslj	karena pssi nya gizinya tak bagus
3	@detikcom popularitas nasionalisme	demi popularitas abaikan nasionalisme
4	@detikcom apa bapak gak tau kalo timnas kita ini adalah timnas belanda cabang asia	apa bapak gak tau kalo timnas kita ini adalah timnas belanda cabang asia
5	@idextratime orang indonesia 270jt gada 20orang yg jago bola?	jt orang indonesia gada orang yg jago bola

## 3. *Tokenizing*

Tahap *tokenizing* akan memecah sebuah kalimat menjadi unit-unit dasar yang disebut dengan token. Sebuah token dapat berupa kata individual, kombinasi, dan juga karakter tunggal. Dengan *tokenizing*, akan mempermudah untuk diproses dan dianalisa. Berikut ini merupakan hasil dari tahap *tokenizing*:

Tabel 3. 3 Hasil *Tokenizing*

No	Sebelum <i>Tokenizing</i>	Hasil <i>Tokenizing</i>
1	kalian punya juta penduduk kenapa pengurus danantara dari asing	kalian, punya, juta, penduduk, kenapa, pengurus, danantara, dari, asing
2	karena pssi nya gizinya tak bagus	karena, pssi, nya, gizinya, tak, bagus
3	demi popularitas abaikan nasionalisme	demi, popularitas, abaikan, nasionalisme
4	apa bapak gak tau kalo timnas kita ini adalah timnas belanda cabang asia	apa, bapak, gak, tau, kalo, timnas, kita, ini, adalah, timnas, belanda, cabang, asia
5	jt orang indonesia gada orang yg jago bola?	jt, orang, indonesia, gada, orang, yg, jago, bola

## 4. *Stopword Removal*

Pada tahap *stopword removal* digunakan untuk menghilangkan beberapa kata umum yang tidak memiliki arti penting dan minim kontribusi dalam memahami konten teks secara keseluruhan. Contoh kata yang dihilangkan dalam *stopword removal* seperti “yang”, “dan”. Berikut merupakan hasil dari *stopword removal*:

Tabel 3. 4 Hasil *Stopword Removal*

No	Sebelum <i>Stopword-removal</i>	Hasil <i>Stopword-removal</i>
----	---------------------------------	-------------------------------

1	kalian, punya, juta, penduduk, kenapa, pengurus, danantara, dari, asing	kalian, punya, juta, penduduk, pengurus, danantara, asing
2	karena, pssi, nya, gizinya, tak, bagus	pssi, nya, gizinya, tak, bagus
3	demi, popularitas, abaikan, nasionalisme	popularitas, abaikan, nasionalisme
4	apa, bapak, gak, tau, kalo, timnas, kita, ini, adalah, timnas, belanda, cabang, asia	apa, bapak, gak, tau, kalo, timnas, timnas, belanda, cabang, asia
5	jt, orang, indonesia, gada, orang, yg, bener, jago, bola	jt, orang, indonesia, gada, orang, yg, bener, jago, bola

### 5. Stemming

Pada tahap *stemming*, berguna untuk mengubah sebuah kata yang memiliki imbuhan menjadi kata dasar atau kata awal. *Stemming* akan memperpendek sebuah kata dengan mnghilangkan awalan dan akhiran. Berikut ini merupakan hasil dari *stemming*:

Tabel 3. 5 Hasil *Stemming*

No	Sebelum <i>Stemming</i>	Hasil <i>Stemming</i>
1	kalian, punya, juta, penduduk, pengurus, danantara, asing	kalian, punya, juta, duduk, urus, danantara, asing
2	pssi, nya, gizinya, tak, bagus	pssi, nya, gizi, tak, bagus
3	popularitas, abaikan, nasionalisme	popularitas, abai, nasionalisme

4	apa, bapak, gak, tau, kalo, timnas, timnas, belanda, cabang, asia	apa, bapak, gak, tau, kalo, timnas, timnas, belanda, cabang, asia
5	jt, orang, indonesia, gada, orang, yg, bener, jago, bola	jt, orang, indonesia, gada, orang, yg, bener, jago, bola

### 3.3 Ekstraksi Fitur

Setelah tahap pemrosesan teks selesai, selanjutnya dilakukan ekstraksi fitur. Ekstraksi fitur dilakukan dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF umum digunakan untuk pembobotan kata dalam kasus penelitian *text mining* dan analisis sentimen. Tujuan dari ekstraksi fitur ini adalah untuk memberikan bobot yang tinggi terhadap beberapa kata dalam dokumen, serta menurunkan bobot kata yang sering muncul dalam dokumen.

Dalam ekstraksi fitur, untuk membuat sebuah *word vector*, dilakukan perhitungan dengan TF-IDF. Pada TF-IDF terdiri dari dua kata, yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF diartikan sebagai jumlah sebuah kata dari setiap dokumen, sedangkan IDF digunakan untuk mengurangi bobot sebuah kata apabila kata tersebut sering muncul diseluruh dokumen. Berikut ini merupakan hasil perhitungan TF-IDF:

#### a. Hasil Perhitungan *Term Frequency*

Tabel 3. 6 Hasil Perhitungan TF

No	Term	TF (D1)	TF (D2)	TF (D3)	TF (D4)	TF (D5)
1	main	0,2	0	0	0,14	0
2	lokal	0,2	0	0	0	0
3	jadi	0,2	0	0	0	0
4	minoritas	0,2	0	0	0	0
5	wkwkwk	0,2	0	0	0	0
6	pssi	0	0,2	0	0	0
7	nya	0	0,2	0	0	0

8	gizi	0	0,2	0	0	0
9	tak	0	0,2	0	0	0
10	bagus	0	0,2	0	0	0
11	popularitas	0	0	0,33 3	0	0
12	abai	0	0	0,33 3	0	0
13	nasionalis me	0	0	0,33 3	0	0
14	butuh	0	0	0	0,14 2	0
15	naturalisasi	0	0	0	0,14 2	0,2 5
16	indo	0	0	0	0,14 2	0
17	passing	0	0	0	0,14 2	0
18	aja	0	0	0	0,14 2	0
19	salah	0	0	0	0,14 2	0
20	stop	0	0	0	0	0,2 5
21	pake	0	0	0	0	0,2 5
22	istilah	0	0	0	0	0,2 5

b. Hasil Perhitungan IDF

Tabel 3. 7 Hasil Perhitungan IDF

No	Term	IDF
1	main	0,39794
2	lokal	0,69897
3	jadi	0,69897

4	minoritas	0,69897
5	wkwkwk	0,69897
6	pssi	0,69897
7	nya	0,69897
8	gizi	0,69897
9	tak	0,69897
10	bagus	0,69897
11	popularitas	0,69897
12	abai	0,69897
13	nasionalisme	0,69897
14	butuh	0,69897
15	naturalisasi	0,39794
16	indo	0,69897
17	passing	0,69897
18	aja	0,69897
19	salah	0,69897
20	stop	0,69897
21	pake	0,69897
22	istilah	0,69897

c. Hasil Perhitungan TFIDF

Tabel 3. 8 Hasil Perhitungan TFIDF

No	Term	TFI DF (D1)	TFI DF (D2)	TFID F (D3)	TFI DF (D4)	TFI DF (D5)
1	main	0,07 9588	0	0	0,056 5074 8	0
2	lokal	0,01 3979 4	0	0	0	0

3	jadi	0,01 3979 4	0	0	0	0
4	minoritas	0,01 3979 4	0	0	0	0
5	wkwk wk	0,01 3979 4	0	0	0	0
6	pssi	0	0,01 3979 4	0	0	0
7	nya	0	0,01 3979 4	0	0	0
8	gizi	0	0,01 3979 4	0	0	0
9	tak	0	0,01 3979 4	0	0	0
10	bagus	0	0,01 3979 4	0	0	0
11	popul aritas	0	0	0,023 27570 1	0	0
12	abai	0	0	0,023 27570 1	0	0
13	nasion alisme	0	0	0,023 27570 1	0	0
14	butuh	0	0	0	0,099 2537 4	0
15	natura lisasi	0	0	0	0,056 5074 8	0,056 5074 8

16	indo	0	0	0	0,099 2537 4	0
17	passin g	0	0	0	0,099 2537 4	0
18	aja	0	0	0	0,099 2537 4	0
19	salah	0	0	0	0,099 2537 4	0
20	stop	0	0	0	0	0,174 7425
21	pake	0	0	0	0	0,174 7425
22	istilah	0	0	0	0	0,174 7425

### 3.4 Pelatihan Model

Setelah proses ekstraksi fitur selesai, selanjutnya masuk ke tahap pelatihan model. Pada pelatihan model, dataset yang berjumlah 1002 data dibagi menjadi data latih (*train data*) dan data uji (*testing data*). Presentase pembagian data latih dan data uji adalah 80% untuk data latih, serta 20% untuk data uji. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk mengukur kinerja dari kedua model. Berikut merupakan data pelatihan model:

Tabel 3. 9 Data Latih

No	Teks	Label
1	cocok bener	Positif
2	bener tu kiper lb gelandang	Positif
3	arti timnas indonesia tuai hasil muas andal kualitas main bina tim atau liga domestik pr pssi banyak baik sistem sepakbola indonesia andal transfer naturalisasi	Positif
4	padahal udah nunggu kuartet lini blakang main bareng walsh baggott pattynama	Positif

5 caraka kek nama hokky bgt pdhl Negatif  
club jg jadi camat

### 3.5 Evaluasi Model

Setelah proses pelatihan model selesai dilakukan, langkah selanjutnya adalah melakukan evaluasi performa dari kedua model. Tujuan dari evaluasi model adalah untuk mengetahui performa terhadap model yang digunakan. Evaluasi model dilakukan dengan menggunakan data uji, lalu menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, serta *f1-score*. Berikut merupakan tabel perbandingan hasil evaluasi dari kedua model yang digunakan :

Tabel 3. 10 Hasil Evaluasi Model

Aspek	Naïve Bayes	SVM
<b>Accuracy</b>	76,23%	70,29%
<b>Precision</b>	78,43%	76,74%
<b>Recall</b>	75,47%	62,26%
<b>F1-Score</b>	76,93%	68,73%

Pada tabel 3.10 menunjukkan hasil perbandingan evaluasi model. Berdasarkan data yang telah disajikan, terlihat bahwasannya model *Naïve Bayes* terlihat memiliki performa yang lebih unggul dibandingkan *Support Vector Machine* dari semua segi perhitungan metrik evaluasi yang dilakukan. Berdasarkan hasil pengujian terhadap data uji, menunjukkan model *Naïve Bayes* memiliki nilai akurasi sebesar 76,23%. Nilai tersebut jauh lebih tinggi dibandingkan model *Support Vector Machine* yang hanya mendapatkan nilai sebesar 70,29%. Sementara itu, hasil perhitungan presisi, *recall*, dan *f1-score* secara berturut-turut *Naïve Bayes* menghasilkan nilai 78,43%, 75,47, dan 76,93. Hal itu juga meghasilkan nilai yang lebih tinggi dibandingkan *Support Vector Machine* yang hanya menghasilkan presisi sebesar 76,74%, *recall* sebesar 62,26%, dan *f1-score* sebesar 68,73%.

### 3.6 Hasil Prediksi

Setelah berhasil melakukan evaluasi model terhadap data uji yang telah diberikan, model akan melakukan prediksi data cuitan terhadap data latih. Tujuan dari prediksi ini adalah mengetahui keakuratan kedua model terhadap data uji. Dari hasil prediksi menunjukkan model mampu melakukan prediksi terhadap data uji, meskipun hasil prediksi dengan data aslinya berbeda. Pada tabel 3.11 berikut menampilkan 10 dari 101 data hasil prediksi dari kedua model:

Tabel 3. 11 Hasil Prediksi

No	Teks Asli	Label Asli	Prediksi NB	Prediksi SVM
1	semua main turun .....turun	negatif	positif	positif
2	kalo mau benah sepak bola .....latih	negatif	positif	positif
3	jangan bilang bunuh .... kapan	positif	negatif	negatif
4	konsep nya .... naturalisasi	negatif	negatif	negatif
5	proses instan	negatif	negatif	negatif
6	politis pdip neng banget ngerecoki bola paok gak ketulungan	negatif	negatif	negatif
7	kunci hilang budaya .... manajemen	positif	negatif	negatif
8	anjay si mees emang gak boong kalo orang indo mukanye kek om timur gw	positif	negatif	positif
9	perlu puluh main belakang tengah amp striker	positif	positif	negatif
10	khayal anda boaz amp gonzales umur tahun	negatif	positif	positif

### 3.6 Implementasi Sistem

Implementasi sistem bertujuan untuk membangun antarmuka analisis sentimen berbasis web yang akan memudahkan dalam proses analisis sentimen. Beberapa proses analisis sentimen yang ada dalam sistem ini antara lain *upload file*, *preprocessing*, *training* model, hasil prediksi model, serta visualisasi hasil. Berikut merupakan tampilan dari sistem yang telah dibangun:

a. Halaman beranda

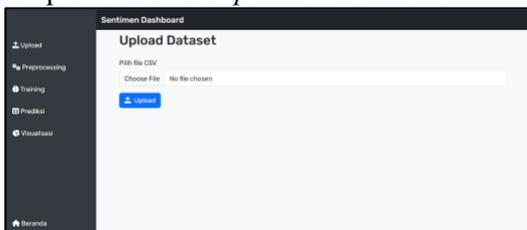
Pada halaman beranda menampilkan *landing page* sebagai *homescreen* ketika sistem baru pertama kali dibuka. Berikut merupakan tampilan dari halaman beranda:



Gambar 3. 4 Halaman Beranda

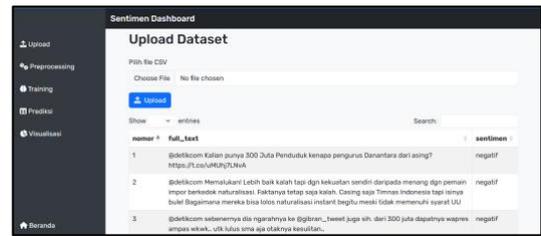
b. Halaman *upload file*

Halaman *upload file* berperan sebagai halaman utama. Pada halaman ini, berguna untuk mengunggah file dataset yang akan dilakukan analisis sentimen. File yang diunggah harus file dengan format csv. Berikut ini merupakan tampilan halaman *upload file* dataset:



Gambar 3. 5 Halaman *Upload File*

Gambar 3.5 merupakan tampilan halaman *upload file* dataset. Ketika pengguna sudah berhasil mengunggah file, maka bisa menekan tombol *upload*. Ketika tombol berhasil di klik, akan menampilkan *preview* dari dataset yang telah diunggah. Berikut merupakan tampilan halaman *upload* dataset setelah pengguna menunggah file:

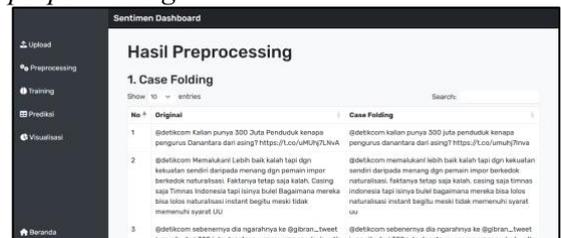


Gambar 3. 6 *Preview Dataset*

Terlihat pada gambar 3.6, ketika dataset berhasil diunggah, akan menampilkan *preview* dataset.

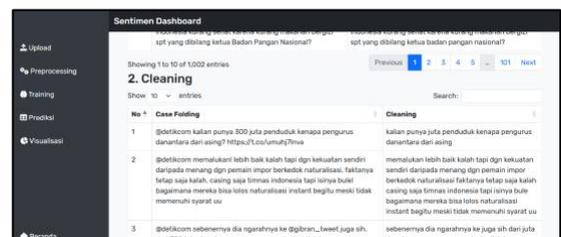
c. Halaman *Preprocessing*

Pada halaman *preprocessing* berguna untuk melakukan tahapan pemrosesan teks dan juga menampilkan hasilnya dalam bentuk tabel. Halaman *preprocessing* akan menampilkan hasil dari setiap tahapan pemrosesan seperti *case folding*, *cleaning*, *tokenzing*, *stopword removal*, serta *stemming*. Berikut merupakan tampilan dari setiap tahap halaman *preprocessing*:



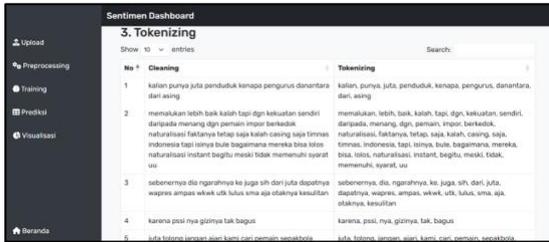
Gambar 3. 7 Halaman *Preprocessing*

Pada gambar 3.7 menampilkan halaman *preprocessing*. Dalam *preprocessing* ini, terdapat langkah atau tahapan yang dijalankan agar data bisa digunakan untuk analisis. Dalam tampilan menu *preprocessing*, menampilkan hasil pemrosesan dalam bentuk tabel, dan setiap langkah mulai dari *case-folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming* disajikan dalam tabel yang terpisah. Berikut merupakan tampilan lainnya hasil dari *preprocessing*:



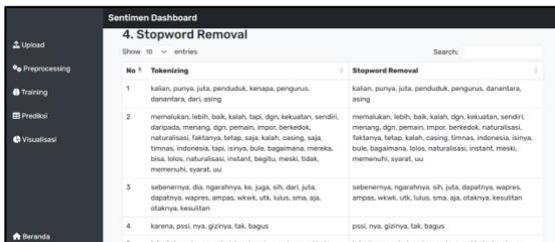
Gambar 3. 8 Tampilan Hasil *Cleaning*

Dalam tampilan gambar 3.8 menampilkan tabel yang menampilkan data sebelum dan sesudah *cleaning*. Kolom hasil *cleaning* menampilkan data yang telah dihilangkan beberapa huruf yang tidak diperlukan seperti nomor, url, *mention* akun, dan emoji.



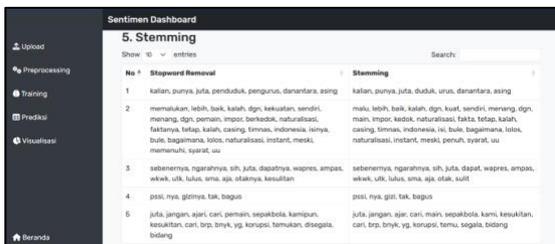
Gambar 3. 9 Tampilan Hasil *Tokenizing*

Gambar 3.9 merupakan hasil pemrosesan pada tahap *tokenizing*. Pada tampilan hasil *tokenizing* menampilkan sebuah tabel yang berisi data sebelum dan sesudah *tokenizing*. Kolom hasil *tokenizing* menampilkan data teks yang setiap hurufnya dipisah oleh tanda koma, atau bisa dibilang hurufnya terpecah.



Gambar 3. 10 Hasil *Stopword removal*

Gambar 3.10 merupakan hasil pemrosesan pada tahap *stopword removal*. Pada tampilan hasil *stopword removal* menampilkan data sebelum dan sudah *stopword removal*.



Gambar 3. 11 Hasil *Stemming*

Gambar 3.11 merupakan hasil pemrosesan tahap *stemming*. Pada tampilan hasil *stemming*, menampilkan data sebelum dan sesudah *stemming*.

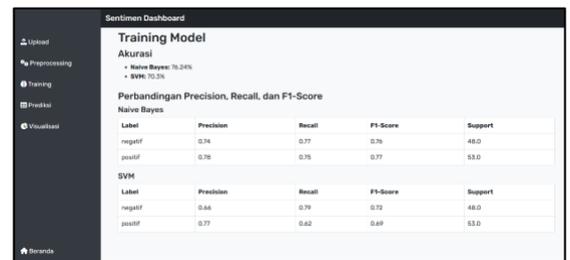


Gambar 3. 12 Hasil Final *Clean Text*

Gambar 3.12 merupakan hasil dataset yang telah bersih, atau data yang telah melalui semua tahapan *preprocessing*.

d. Halaman *Training Model*

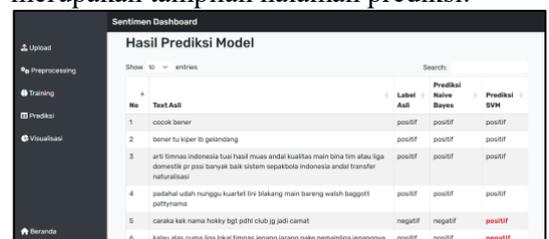
Pada halaman ini, akan menampilkan hasil akurasi model *Naive Bayes* dan *Support Vector Machine*. Pada menu *training model*, langkah ekstraksi fitur dijalankan sekaligus dalam menu ini. Berikut merupakan tampilan halaman *training model*:



Gambar 3. 13 Halaman *Training Model*

e. Halaman *Prediksi*

Pada halaman prediksi, sistem menampilkan hasil prediksi model *naive bayes* dan *Support Vector Machine* terhadap label sentimen dataset asli. Kedua model mampu melakukan prediksi terhadap dataset yang ada. Pada tampilan halaman prediksi, akan menampilkan data dalam bentuk tabel. Berikut ini merupakan tampilan halaman prediksi:



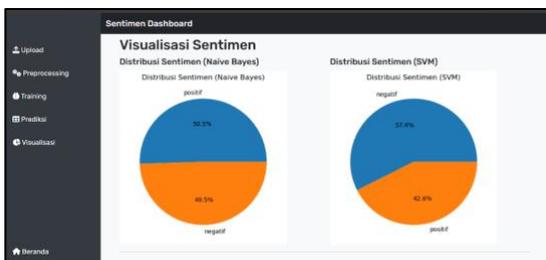
Gambar 3. 14 Halaman *Prediksi*

Pada gambar 3.14 terlihat model memprediksi sentimen dengan label dataset asli. Pada tampilan hasil prediksi, terlihat jika hasil prediksi model tidak sama dengan label asli, maka teks label sentimen akan berwarna

merah. Sementara itu, jika prediksi sesuai dengan label asli, maka warna teks akan tetap hitam.

f. Halaman visualisasi

Menu terakhir dalam sistem ini adalah menu visualisasi. Dalam menu ini, akan menampilkan tiga grafik visualisasi hasil analisis sentimen yang dilakukan oleh kedua model. Grafik visualisasi yang ditampilkan antara lain seperti grafik distribusi sentimen kedua model, visualisasi *confusion matrix*, serta grafik tingkat kesesuaian antara model dengan label asli. Berikut ini merupakan tampilan menu visualisasi dari sistem yang dibuat:



Gambar 3.15 Halaman Visualisasi

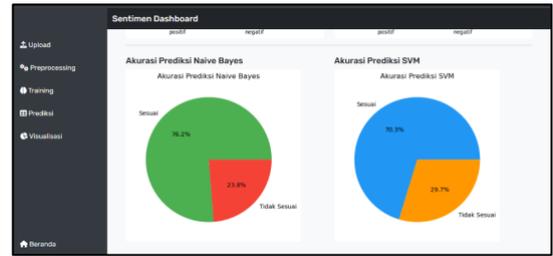
Gambar 3.15 menampilkan *pie chart* hasil distribusi sentimen model *Naive Bayes* dan *Support Vector Machine*. Dalam grafik tersebut menunjukkan jumlah presentase hasil prediksi sentimen kedua model. Model *Naive Bayes* memiliki nilai presentase sentimen positif sebanyak 51.7%, sedangkan sentimen negatif berjumlah 48.3%. Sedangkan model *Support Vector Machine* memiliki nilai presentase sebesar 51.7% untuk sentimen negatif, sedangkan 48.3% untuk sentimen positif.



Gambar 3.16 Hasil *Confusion Matrix*

Pada gambar 3.16 menampilkan perbandingan *confusion matrix* dari kedua model. *Confusion matrix* tersebut menampilkan jumlah hasil data uji yang dilakukan kedua model. Kedua model

menunjukkan nilai *confusion matrix* yang berbeda.



Gambar 3.17 Grafik Kesesuaian Model

Pada gambar 3.17 menampilkan grafik akurasi kesesuaian dan ketidaksesuaian kedua model terhadap label asli. Model *Naive Bayes* menunjukkan tingkat kesesuaian sebesar 67,2%. Nilai tersebut lebih tinggi jika dibandingkan dengan model *Support Vector Machine* yang hanya memiliki kesesuaian sebesar 63,7%.

## 4 Kesimpulan

Pada penelitian ini berhasil melakukan analisis sentimen terkait pemain naturalisasi Timnas Indonesia menggunakan metode *Naive Bayes* dan *Support Vector Machine*. Tujuan utama dari analisis ini adalah untuk mengetahui tingkat akurasi dari model yang digunakan. Dari hasil percobaan, model *Naive Bayes* memiliki tingkat akurasi yang lebih tinggi dibandingkan *Support Vector Machine*. Akurasi *Naive Bayes* sebesar 76,23%, sedangkan SVM memiliki tingkat akurasi sebesar 70,29%.

Hasil perhitungan evaluasi model menunjukkan model *Naive Bayes* juga lebih unggul dibandingkan SVM. Dari aspek presisi, *recall*, dan *f1-score* model *Naive Bayes* memiliki nilai 78,43%, 75,47%, dan 76,93%. Sedangkan *Support Vector Machine* memiliki nilai presisi sebesar 76,74%, *recall* sebesar 62,26%, dan *f1-score* sebesar 68,73%.

### References:

[1] G. Gunawan and I. Mahfud, "PENGARUH LATIHAN DRIBBLE DENGAN METODE BERMAIN TERHADAP HASIL DRIBBLE SEPAK BOLA SSB MITRA UTAMA LAMPUNG SELATAN," *J. Phys. Educ.*, vol. 3, no. 2, pp. 49–58, Dec. 2022.

[2] Bola.com, "Rekor Timnas Indonesia di Sepanjang Sejarah Kualifikasi Piala Dunia: 15 Kali Berpartisipasi, Bagaimana Hasilnya?," bola.com. Accessed: Dec. 06, 2024. [Online]. Available:

<https://www.bola.com/indonesia/read/5620179/rekor-timnas-indonesia-di-sepanjang-sejarah-kualifikasi-piala-dunia-15-kali-berpartisipasi-bagaimana-hasilnya>

Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naïve Bayes,” *J. Inf. Syst. Res. JOSH*, vol. 4, no. 4, Art. no. 4, Jul. 2023, doi: 10.47065/josh.v4i4.3737.

[3] G. K. Annas and N. M. Hazzar, “Analisis Persamaan Hak Kewarganegaraan bagi Pemain Naturalisasi Sepakbola di Indonesia,” *WICARANA*, vol. 2, no. 2, Art. no. 2, Oct. 2023, doi: 10.57123/wicarana.v2i2.37.

[4] “Kritik Keras Naturalisasi, Anggota DPR Fraksi Demokrat Dirujuk Netizen Sampai Dilaporkan ke AHY,” *merdeka.com*. Accessed: Dec. 06, 2024. [Online]. Available: <https://www.merdeka.com/trending/kritik-keras-naturalisasi-anggota-dpr-fraksi-demokrat-dirujuk-netizen-sampai-dilaporkan-ke-ahy-228654-mvk.html>

[5] D. Atika, S. Styawati, and A. A. Aldino, “TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TEKANAN MENTAL PADA MEDIA SOSIAL TWITTER,” *J. Teknol. Dan Sist. Inf.*, vol. 3, no. 4, pp. 86–97, 2022, doi: 10.33365/jtsi.v3i4.2054.

[6] “Apa itu Analisis Sentimen? - Penjelasan tentang Analisis Sentimen - AWS,” Amazon Web Services, Inc. Accessed: Jul. 18, 2025. [Online]. Available: <https://aws.amazon.com/id/what-is/sentiment-analysis/>

[7] D. N. Novianti, D. F. Shiddieq, F. F. Roji, and W. Susilawati, “Komparasi Algoritma Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Metaverse,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, Art. no. 1, Dec. 2023, doi: 10.57152/malcom.v4i1.1061.

[8] S. Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, “Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm,” *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, Art. no. 1, Jul. 2023, doi: 10.57152/predatecs.v1i1.814.

[9] M. U. Albab, Y. K. P, and M. N. Fawaiq, “Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic,” *J. Transform.*, vol. 20, no. 2, Art. no. 2, Jan. 2023, doi: 10.26623/transformatika.v20i2.5374.

[10] R. Sulastiyono, A. Setiawan, and S. Nugroho, “Sentimen Analisis Pembatalan Indonesia Menjadi