

Analisis Sentimen Terhadap Naturalisasi Pemain pada Youtube Menggunakan *Decision Tree* dan *Naive Bayes*

Sentiment Analysis of Player Naturalization on Youtube Using Decision Trees and Naive Bayes

Billy Franko¹, Nicholas Wilyanto², Hafiz Irsyad³

Informatika, Universitas Multi Data Palembang, Palembang, Indonesia^{1,2,3,4}

Email: billy.franko@mhs.mdp.ac.id¹, nicholas.wilyanto@mhs.mdp.ac.id², hafizirsyad@mdp.ac.id³

Abstract

This research was conducted with the aim of knowing the sentiment of the community on the topic of naturalization which is currently popular regarding the Indonesian national football team players. Researchers chose the *Decision Tree* algorithm and the *Naive Bayes* algorithm as the algorithms to be used. *Decision Tree* and *Naive Bayes* are some algorithms that are commonly used in sentiment analysis because they provide good results. The test results show that the *Decision Tree* algorithm with depth 3 has an accuracy of 70%, depth 4 has an accuracy of 71.8% and depth 5 has an accuracy of 70.9% and the *Naive Bayes* algorithm has an accuracy of 85.4%. The results of this study indicate that the *Naive Bayes* algorithm is one algorithm that is quite effective in building a classification model in the case of sentiment analysis. By utilizing the *Decision Tree* and *Naive Bayes* algorithms in building models for sentiment analysis, it is hoped that it will be able to provide a clearer picture of public opinion on naturalization so that the government can find out whether the public supports or rejects the addition of naturalized players to the Indonesian national football team.

Keywords: *Sentiment Analysis, Decision Tree, Naive Bayes, Naturalization*

Abstrak

Penelitian ini dilakukan dengan tujuan untuk mengetahui sentimen masyarakat mengenai topik naturalisasi yang sedang hangat belakangan ini mengenai pemain Timnas Sepakbola Indonesia. Peneliti memilih algoritma *Decision Tree* dan algoritma *Naive Bayes* sebagai algoritma yang akan digunakan. *Decision Tree* dan *Naive Bayes* merupakan beberapa algoritma yang umumnya dipakai dalam analisis sentimen karena memberikan hasil yang baik. Hasil pengujian menunjukkan algoritma *Decision Tree* dengan *depth* 3 memiliki akurasi sebesar 70%, *depth* 4 memiliki akurasi sebesar 71,8% dan *depth* 5 memiliki akurasi sebesar 70,9% dan algoritma *Naive Bayes* memiliki akurasi sebesar 85,4%. Hasil penelitian ini menunjukkan bahwa algoritma *Naive Bayes* merupakan salah satu algoritma yang cukup efektif dalam membangun suatu model klasifikasi dalam kasus analisis sentimen. Dengan memanfaatkan algoritma *Decision Tree* dan *Naive Bayes* dalam membangun model guna analisis sentimen diharapkan mampu memberi gambaran yang lebih jelas mengenai pendapat masyarakat terhadap naturalisasi sehingga pemerintah dapat mengetahui apakah masyarakat mendukung atau justru menolak adanya penambahan pemain naturalisasi dalam Timnas sepakbola Indonesia.

Kata Kunci: Analisa Sentimen, *Decision Tree, Naive Bayes, Naturalisasi*

I. PENDAHULUAN

Sepak bola tetap menjadi salah satu olahraga yang sangat populer, terutama di Indonesia, yang terbukti dari adanya kompetisi yang terus berlangsung baik di dalam negeri maupun internasional. Akarnya di Indonesia dapat ditelusuri kembali hingga tahun 1914 pada masa penjajahan Belanda, dengan munculnya klub-klub sepak bola yang diperkenalkan oleh berbagai kelompok etnis di Hindia Belanda, termasuk orang Belanda, Arab, dan Tionghoa [1].

Salah satu debat sentral yang berkaitan dengan kewarganegaraan adalah tentang siapa yang berhak atasnya. [2]. Di

Indonesia, ada suatu istilah yang dikenal sebagai naturalisasi untuk memperoleh status kewarganegaraan bagi penduduk non-pribumi (asing) setelah memenuhi semua persyaratan dan prosedur yang ditetapkan [3].

Beberapa hari ini, sering terdapat masalah dalam bentuk hujatan terdapat pemain-pemain Timnas Indonesia hasil naturalisasi. Sinulingga sebagai Staf Khusus Menteri BUMN dan Anggota Exco PSSI, yang dikutip dari bola.net, "Sebab, bisa menghambat proses naturalisasi yang dikerjakan. Pemain-pemain yang tadinya ingin membela Timnas Indonesia, karena netizen suka menyerang seperti itu, jadi takut". Dapat disimpulkan bahwa naturalisasi merupakan

salah satu topik sensitif dan bersifat kontroversial bagi masyarakat Indonesia [4].

Analisis sentimen berfokus terhadap pendapat serta sikap, evaluasi, dan emosi seseorang. Bidang ini telah berkembang pesat karena kemajuan dalam pemrosesan bahasa alami, didorong oleh perkembangan dalam penambahan teks, penambahan data, penambahan web, dan pembelajaran mesin. Analisis sentimen sering dipakai dan diperlukan untuk menganalisis suatu produk, konsumen, jasa, kesehatan, industri dan keuangan [5].

Berbagai pendekatan analisis sentimen seperti *Naive Bayes*, *Decision Tree*, dan *K-nearest Neighbor* (KNN) umumnya digunakan dalam kasus penambahan data untuk menghasilkan hasil yang optimal. [6]. Algoritma KNN mengklasifikasikan data berdasarkan jarak terpendek ke objek data. Meskipun memiliki berbagai keunggulan, KNN juga memiliki kekurangan seperti kebutuhan akan nilai k yang ditentukan, kompleksitas komputasi, batasan memori, dan kerentanan terhadap atribut yang tidak relevan. [7]. *Decision tree* adalah representasi berbentuk pohon yang terdiri dari simpul yang mewakili keputusan dan cabang yang mewakili hasil dari setiap keputusan tersebut. Keuntungannya meliputi interpretabilitas, penanganan data kategorikal dan numerik, serta kecenderungan untuk tidak memerlukan normalisasi atau standarisasi data. [8]. *Naive Bayes Classifier* adalah algoritma *data mining* yang memanfaatkan probabilitas dan statistik [9]. *Naive Bayes Classifier* ini merupakan algoritma yang didasarkan pada teorema Bayes dengan asumsi yang kuat tentang independensi antara fitur-fitur [10]. *Naive Bayes* menawarkan keuntungan seperti berlaku untuk data kuantitatif dan kualitatif, memerlukan volume data minimal, dan proses pelatihan serta pengujian yang efisien [11].

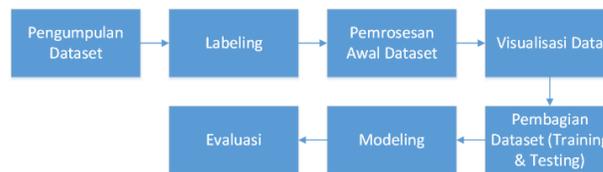
Media sosial memiliki beragam bentuk yang memiliki ciri khas masing-masing, seperti blog, konten audio, visual, dan audiovisual. Salah satu *platform* media sosial yang populer di negara Indonesia adalah Youtube. Youtube memiliki keunggulan tertentu, salah satunya adalah interaktivitasnya yang tinggi. Interaksi yang aktif di Youtube didukung oleh fitur komentar, di mana pengguna dapat memberikan pendapat atau bertanya tentang konten yang disampaikan dalam video, sehingga memicu terbentuknya diskusi. Diskusi ini sering kali memunculkan beragam pendapat dari berbagai sudut pandang yang mencerminkan nilai-nilai sentimen. Dalam kolom komentar Youtube yang kompleks, tidak jarang terdapat opini yang berseberangan. Oleh karena itu, kolom komentar *Youtube* menjadi sumber data yang penting untuk dilakukan analisis sentimen [1].

Dengan dimulainya *Asian Football Confederation* (AFC) U-23, banyak dari masyarakat yang memberikan respon dan komentar melalui *platform* media sosial terkait penggunaan pemain naturalisasi dalam tim nasional Indonesia. Berdasarkan hal tersebut, peneliti berkeinginan untuk melakukan analisis sentimen terhadap tanggapan masyarakat terkait isu tersebut di media sosial, khususnya Youtube.

Penelitian ini bertujuan untuk menganalisa sentimen masyarakat terhadap adanya pemain hasil naturalisasi pada

Timnas Indonesia menggunakan algoritma *Decision Tree* dan *Naive Bayes*.

II. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Penelitian ini dilaksanakan dalam beberapa tahapan seperti yang ditunjukkan pada Gambar 1. Tahapan yang dilaksanakan terdiri dari:

A. Pengumpulan Dataset

Dalam tahapan ini, dilaksanakan pengumpulan *dataset*. *Dataset* dikumpulkan menggunakan layanan *Netlytic*, yang memungkinkan ekstraksi komentar dari video Youtube secara otomatis. Proses pengumpulan data dilakukan dengan memasukkan URL video yang akan dianalisis ke dalam *platform Netlytic*, kemudian data diekstraksi dan disimpan dalam format CSV. Video Youtube yang diambil untuk analisis sentimen berjudul “Pemain Keturunan Baru Timnas Indonesia yang Bakal Bungkam Semua Kritik” dari URL berikut:

https://www.youtube.com/watch?v=7zm3Ke_bFc&lc=Ug_yMx6uYCoAR1sTXMh14AaABAg

Video tersebut diambil pada tanggal 5 Mei 2024 pada pukul 11 siang. *Dataset* yang diambil berisi 1.019 komentar yang berisi dari unsur huruf, emoji, dan angka. *Dataset* berukuran 153 KB. *Dataset* ini memiliki 11 atribut, yang terdiri dari:

1. *id* merupakan atribut yang menunjukkan komentar beberapa.
2. *author* merupakan atribut yang menunjukkan *username* pengirim komentar.
3. *description* merupakan atribut yang menunjukkan isi dari komentarnya.
4. *guid* merupakan atribut yang menunjukkan id global identifikasi dari setiap komentar.
5. *to* merupakan atribut yang menunjukkan kepada siapa komentar itu tertuju. Atribut *to* akan berisi *username* seseorang dalam berbentuk tag seperti “@contoh”, jika komentar mereka tertuju kepada seseorang.
6. *likecount* merupakan atribut yang menunjukkan berapa jumlah *like* yang diperoleh oleh suatu komentar.
7. *link* merupakan atribut yang menunjukkan alamat web komentar tersebut diposting di Youtube.
8. *pubdate* merupakan atribut yang menunjukkan kapan komentar tersebut diposting.
9. *replycount* merupakan atribut yang menunjukkan berapa jumlah balasan bagi suatu komentar.
10. *title* merupakan atribut yang menunjukkan isi dari komentarnya seperti atribut *description*.
11. *authorchannelurl* merupakan atribut yang menunjukkan

Tabel 1. Contoh Labeling

Contoh Kalimat	Sentimen
banyak main naturalisasi baik tim tapi benar buruk kembang putra daerah sempat putra daerah makin kecil	negatif
naturalisasi solusi hebat kuat vamos timnas indonesia	positif
timnas masuk piala dunia aku bakal kawin	netral

alamat web *channel* Youtube pemilik komentar.

Untuk tahap selanjutnya, atribut yang digunakan dari *dataset* yang telah dikumpulkan ini, hanyalah atribut *description* yang berisi komentar dari video yang menjadi fokus utama.

B. Labeling

Langkah selanjutnya setelah *dataset* sudah dibersihkan adalah proses *labeling*. Proses *labeling* dilakukan secara manual terhadap setiap komentar dengan tiga jenis sentimen berupa positif, negatif, dan netral. Penjelasan tiga sentimen yang akan diberikan adalah seperti berikut:

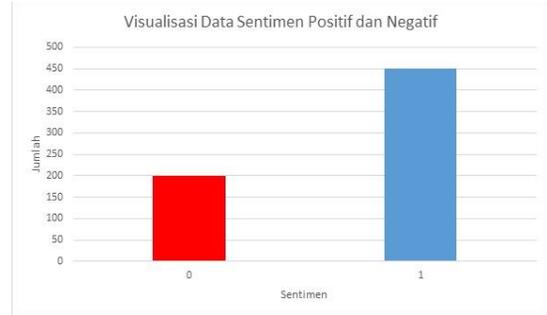
1. Sentimen Positif
Komentar yang menunjukkan perasaan positif seperti kebahagiaan, pujian, atau dukungan.
2. Sentimen Negatif
Komentar yang menunjukkan perasaan negatif seperti kekecewaan, kritik, atau kemarahan.
3. Sentimen Netral
Komentar yang bersifat netral, tidak sesuai dengan topik yang dibahas, atau tidak mengandung emosi yang jelas.

Nantinya komentar dengan kategori sentimen netral akan dihapus dan difokuskan terhadap kategori sentimen positif dan negatif saja. Tabel 1 menunjukkan contoh *labeling* yang dilaksanakan dalam penelitian ini.

C. Pemrosesan Awal Dataset

Setelah dilaksanakan pengumpulan *dataset*, dilanjutkan dengan beberapa proses seperti:

1. Case folding
Proses *case folding* bertujuan untuk mengubah semua huruf menjadi huruf kecil. Untuk mengubah semua huruf menjadi huruf kecil, proses *case folding* memeriksa ukuran setiap karakter dari awal hingga akhir. Jika ada huruf kapital, huruf tersebut akan diubah menjadi huruf kecil. [10].
2. Tokenisasi
Tokenisasi adalah salah satu proses yang membagi data hingga menjadi bagian-bagian kata atau karakter sesuai kebutuhan sistem [5].
3. Stopword removal
Stopword removal bertujuan untuk menghilangkan karakter, tanda baca, dan kata-kata umum yang tidak memberikan informasi yang signifikan. Contohnya adalah kata-kata seperti "dan", "atau", "tapi", dan lainnya [5].
4. Stemming



Gambar 2. Contoh Bar Plot pada Python

Stemming adalah bagian dari pra-pemrosesan data di mana sufiks, prefiks, dan infiks dihilangkan sehingga kata yang dihasilkan berbentuk dasar [12].

Contoh hasil pemrosesan awal *dataset* ditunjukkan pada Tabel 2.

Tabel 2. Sebelum dan Sesudah Pemrosesan Awal Dataset

Sebelum	Sesudah
Bersama pemain keturunan Indonesia menuju Piala Dunia	sama main turun indonesia tuju piala dunia

D. Visualisasi Data

Proses visualisasi data dilaksanakan setelah proses *preprocessing* data selesai dilakukan. Visualisasi dilakukan dengan alasan agar proses analisis sentimen juga dapat memberikan gambaran visual yang jelas. Pada penelitian ini, visualisasi dilakukan dengan cara membuat diagram *plot* dari data yang telah di-*preprocessing* sebelumnya. Dua kategori sentimen yang ingin divisualisasikan adalah *Positive* dan *Negative*. Untuk membuat diagram *plot*, peneliti menggunakan metode *matplotlib* dari Python. Visualisasi yang dihasilkan berbentuk *bar plot*. Contoh visualisasi data menggunakan *bar plot* dapat dilihat pada Gambar 2.

Selain itu, *word cloud* dapat juga digunakan untuk visualisasi data. *Word cloud* adalah gambar visual berdasarkan frekuensi kemunculan kata-kata pada suatu kumpulan teks, dengan ukuran huruf menentukan jumlah kata yang muncul. Dengan demikian semakin besar ukuran huruf menunjukkan semakin banyak kata yang muncul [13]. Pada Gambar 3 bisa dilihat contoh visualisasi data berbentuk *word cloud*.



Gambar 3. Contoh Word Cloud

E. Pembagian Dataset (Training & Testing)

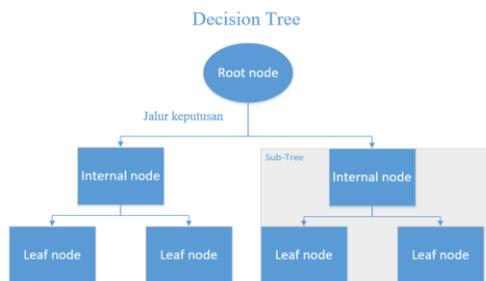
Pada tahap ini, akan dilaksanakan pembagian *dataset* untuk proses *training* dan *testing* model klasifikasi, dimana penelitian ini membagi *dataset* menggunakan *library scikit-learn* bagian *model_selection* menjadi 85% *dataset* menjadi data *training*, dan sisa 15% menjadi data *testing*.

F. Modeling

Pada tahap ini, akan dilaksanakan pemodelan klasifikasi. Data teks yang sudah melalui proses *cleaning* akan digunakan sebagai data *training* model serta data *testing*. Proses *modeling* dilaksanakan menggunakan *hardware* “Notebook Lenovo Ideapad Slim 3i-14igl05”, dan menggunakan *software* Google Colab dengan tujuan untuk menghasilkan sistem klasifikasi. Penelitian ini menggunakan 2 jenis model *machine learning*, yaitu:

1. *Decision Tree*

Decision Tree merupakan metode analisis data yang digunakan untuk memproyeksikan atau menentukan hasil di waktu mendatang dengan cara membentuk model klasifikasi atau regresi dalam bentuk struktur pohon. Proses ini melibatkan pembagian data menjadi bagian-bagian yang lebih kecil dan secara bertahap membangun struktur pohon keputusan. Pohon ini terdiri dari simpul keputusan dan simpul daun [14]. Struktur umum dari algoritma *Decision Tree* dapat dilihat pada Gambar 4.



Gambar 4. Struktur Umum *Decision Tree*

Secara umum cara jalannya algoritma *Decision Tree* adalah sebagai berikut [15]:

- 1) Persiapkan data pelatihan yang umumnya berasal dari data sebelumnya yang telah dikategorikan dan diklasifikasikan ke dalam kategori serta kelas tertentu.
- 2) Selanjutnya, akar pohon dipilih berdasarkan atribut yang memiliki nilai *gain* tertinggi. Dimana, setelah menghitung nilai *gain* dari setiap atribut. Atribut yang memiliki nilai *gain* tertinggi akan menjadi akar pohon pertama.
- 3) Sebelum menghitung nilai *gain* dari atribut, hitung terlebih dahulu nilai entropi. Rumus untuk menghitung nilai entropi ditunjukkan pada (1):

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (1)$$

Keterangan:

- S = himpunan kasus
- n = jumlah partisi S

p_i = proporsi S_i terhadap S

- 4) Hitunglah nilai *gain* menggunakan rumus yang ditunjukkan pada (2):

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * entropy(S_i) \quad (2)$$

Keterangan:

- S = himpunan kasus
- A = fitur
- n = jumlah partisi atribut A
- $|S_i|$ = proporsi S_i terhadap S
- $|S|$ = jumlah kasus dalam S

- 5) Ulangi langkah ke-2 hingga semua data (*record*) terbagi ke dalam partisi yang sesuai.
- 6) Proses partisi pohon akan berhenti ketika:
 - a. Semua data (*record*) dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam data (*record*) yang dipartisi lagi.
 - c. Tidak ada data (*record*) di dalam cabang yang kosong.

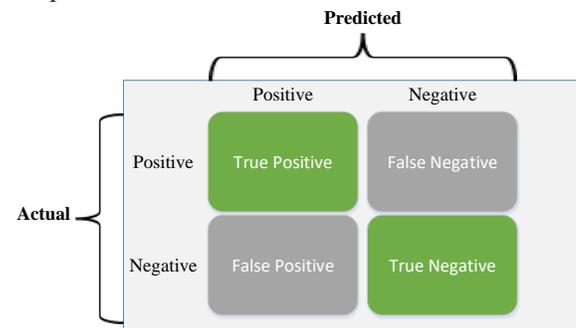
2. *Naive Bayes*

Naive Bayes Classifier adalah suatu metode klasifikasi yang berasal dari teorema Bayes. Metode ini menggunakan probabilitas dan statistik untuk memprediksi peluang berdasarkan pengalaman masa sebelumnya, dengan asumsi yang sangat sederhana tentang ketergantungan antara kondisi atau kejadian [10]. Dasar dari *Naive Bayes* yang dipakai dalam pemrograman adalah rumus Bayes seperti pada (3) ini.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (3)$$

A. Evaluasi

Pada tahap ini, dilakukan evaluasi terhadap model yang sudah dirancang. Evaluasi dilakukan bertujuan untuk mengukur kualitas dan kinerja dari model klasifikasi. Evaluasi dilakukan dengan menggunakan metode *confusion matrix* yang merupakan suatu alat evaluasi metrik yang digunakan untuk melakukan perhitungan akurasi pada kasus *data mining* [12]. Bentuk umum *confusion matrix* dapat dilihat pada Gambar 5.



Gambar 5. Bentuk Umum *Confusion Matrix*

Dalam *confusion matrix*, terdapat beberapa variabel yang digunakan dalam proses evaluasi seperti [6]:

1. TN
TN merupakan singkatan dari *true negative*, yaitu data negatif yang diklasifikasikan sebagai negatif.
2. TP
TP merupakan singkatan dari *true positive*, yaitu data positif yang diklasifikasikan sebagai positif.
3. FN
FN merupakan singkatan dari *false negative*, yaitu data positif yang diklasifikasikan sebagai negatif.
4. FP
FP merupakan singkatan dari *false positive*, yaitu data negatif yang diklasifikasikan sebagai positif.

Dalam *confusion matrix*, terdapat tiga parameter yang digunakan dalam perhitungan evaluasi metrik yaitu [10]:

1. *Accuracy*
Klasifikasi mengacu pada persentase ketepatan data yang diklasifikasikan secara tepat setelah pengujian dilakukan terhadap hasil klasifikasi. Rumus dari *accuracy* dapat dilihat pada (4) berikut.

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

2. *Precision*
Precision adalah rasio antara jumlah kasus yang diprediksi positif dengan benar dibandingkan dengan total kasus yang diprediksi positif. Rumus dari *precision* dapat dilihat pada (5) berikut.

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

3. *Recall*
Recall mengacu pada proporsi kasus positif yang benar-benar positif yang berhasil diprediksi secara benar. Rumus dari *recall* dapat dilihat pada (6) berikut.

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

III. HASIL DAN PEMBAHASAN

B. Visualisasi Data

Visualisasi data dilaksanakan dua kali menggunakan *word cloud*, dimana hasilnya berupa visualisasi kata yang bersifat positif dan negatif. *Word cloud* positif didominasi oleh kata “indonesia”, “timnas”, “naturalisasi”, dan “main”. Sedangkan yang negatif didominasi oleh kata “timnas”, “main”, “yang”, dan “tidak”. Hasil visualisasi data yang memakai *word cloud* dapat dilihat pada Gambar 6. Visualisasi juga dilaksanakan menggunakan *bar plot*, dimana hasilnya menunjukkan bahwa terdapat lebih banyak sentimen positif daripada yang negatif. Hasil visualisasi menggunakan *bar plot* dapat dilihat pada Gambar 7.



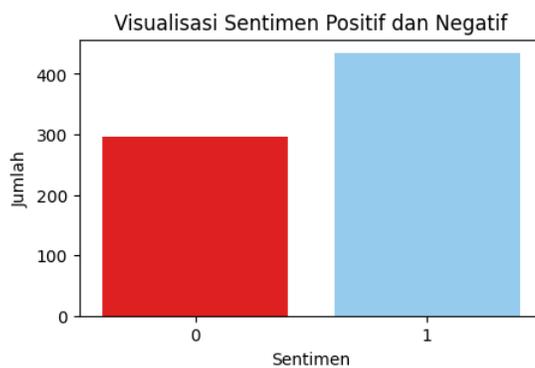
Gambar 6. Hasil Visualisasi Kata Positif dan Negatif

C. Evaluasi Klasifikasi

Untuk model klasifikasi yang menggunakan algoritma *Decision Tree*, pengujian dilakukan menggunakan beberapa *depth* yang berbeda yaitu *max_depth* = 3, 4, dan 5. *Depth* merupakan salah satu parameter dalam pelatihan model *Decision Tree*, dimana *depth* yang terlalu rendah dapat mengakibatkan model yang kurang mampu menangkap pola data dengan baik (*underfitting*), dan *depth* yang terlalu besar dapat mengakibatkan model yang terlalu kompleks dan rentan terhadap *noise* (*overfitting*). Sehingga peneliti memutuskan untuk menggunakan parameter *depth* yang bernilai 3, 4, dan 5 sebagai *depth* yang paling cocok dengan model yang ingin dilatih. Hasil performa model *Decision Tree* dapat dilihat pada Tabel 3.

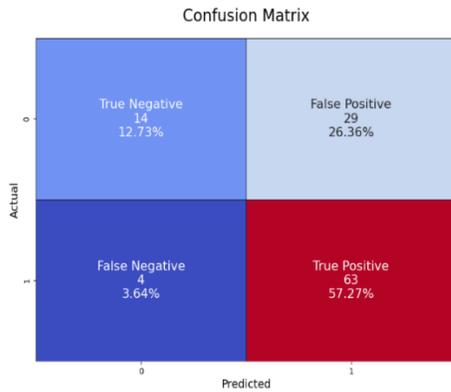
Tabel 3. Tabel Performa *Decision Tree*

<i>max_depth</i>	3		4		5	
	P	N	P	N	P	N
<i>Accuracy</i>	0.7		0.718		0.709	
<i>Precision</i>	0.68	0.78	0.71	0.75	0.70	0.74
<i>Recall</i>	0.94	0.33	0.91	0.42	0.91	0.40
<i>F1-score</i>	0.79	0.46	0.80	0.54	0.79	0.52
<i>Support</i>	67	43	67	43	67	43



Gambar 7. Hasil Visualisasi Sentimen Positif dan Negatif

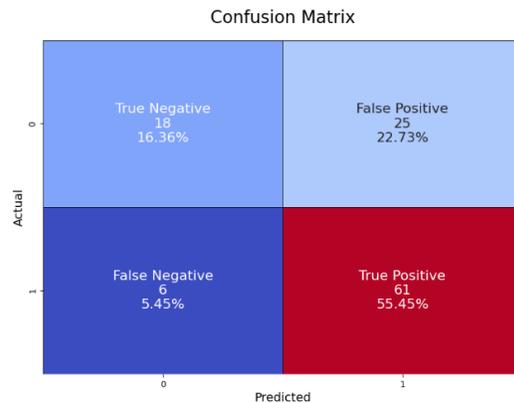
Visualisasi dari *Decision Tree* yang terbentuk dalam *depth* yang berbeda dibangun dengan menggunakan fungsi dari *library scikit-learn*. Visualisasi *Decision Tree* dalam *depth* 3 dapat dilihat pada Gambar 8. Evaluasi dari hasil proses pengujian dari *Decision Tree* dapat dilihat dalam bentuk *confusion matrix*. Hasil dari *confusion matrix* dari *Decision Tree* dalam *depth* 3 dapat dilihat pada Gambar 9:



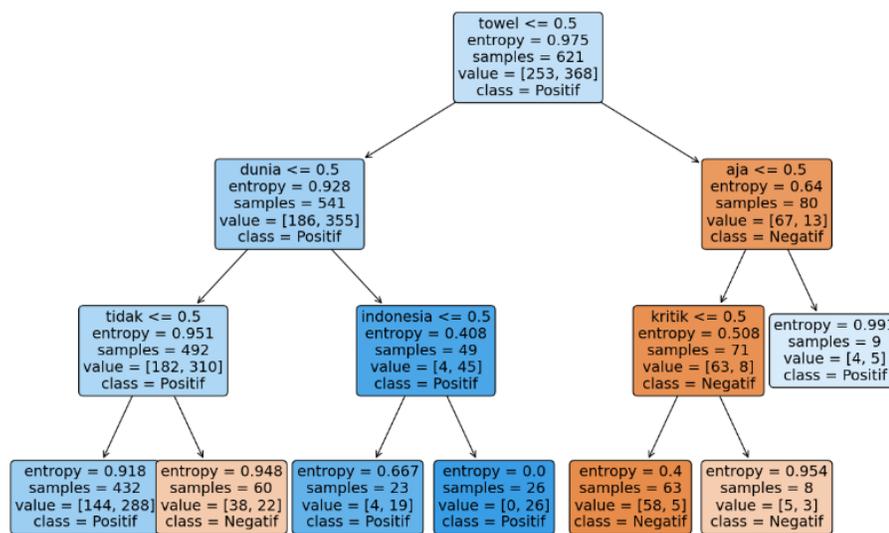
Gambar 9. Confusion Matrix Decision Tree Dalam Depth 3

Hasil dari *confusion matrix* menggunakan *depth* 3 menunjukkan bahwa nilai *true negative* bernilai 12.73%, nilai *false positive* bernilai 26.36%, nilai *false negative* bernilai 3.64%, dan nilai *true positive* bernilai 57.27%. Hasil ini menunjukkan bahwa model berhasil mengklasifikasikan 63 sampel positif dan 14 sampel negatif dengan benar. Modelnya juga salah mengklasifikasi 29 sampel negatif sebagai positif dan 4 sampel positif sebagai negatif.

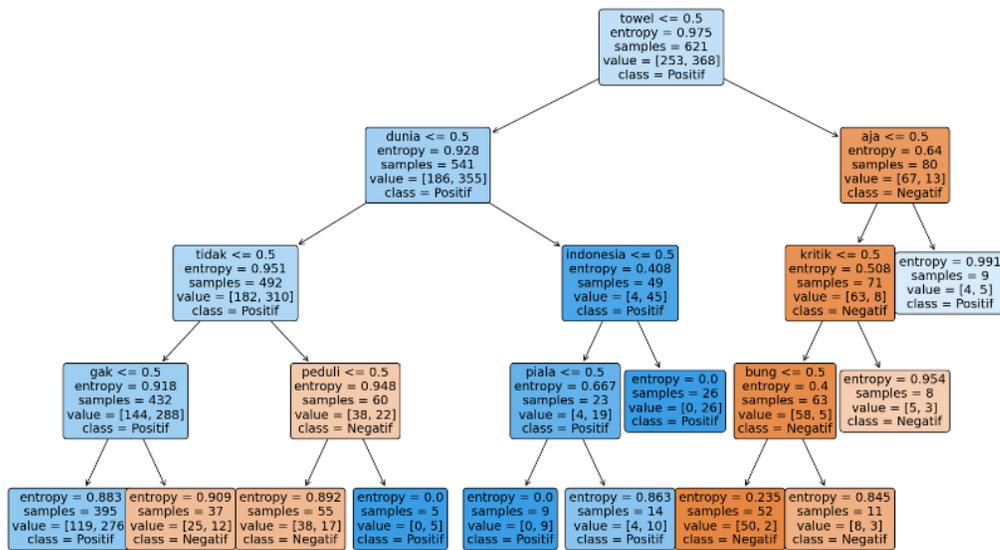
Visualisasi *Decision Tree* dalam *depth* 4 dapat dilihat pada Gambar 10. Sedangkan hasil *confusion matrix* dari *Decision Tree* dalam *depth* 4 dapat dilihat pada Gambar 11.



Gambar 11. Confusion Matrix Decision Tree Dalam Depth 4



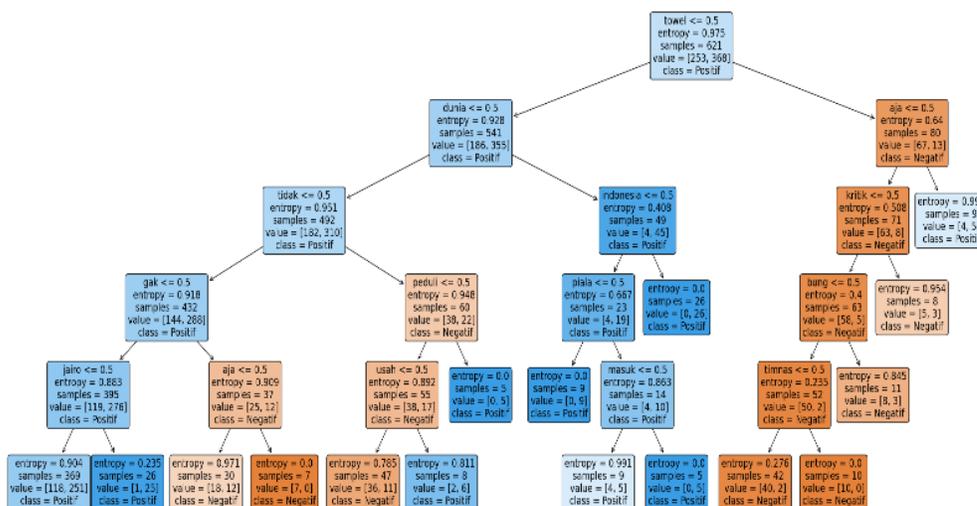
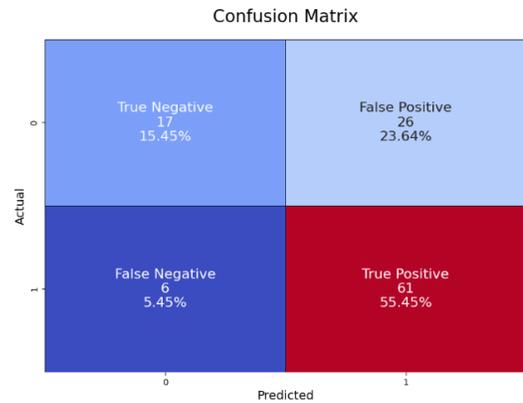
Gambar 8. Visualisasi Decision Tree Dalam Depth 3



Gambar 10. Visualisasi Decision Tree Dalam Depth 4

Hasil dari *confusion matrix* menggunakan *depth 4* menunjukkan bahwa nilai *true negative* bernilai 16.36%, nilai *false positive* bernilai 22.73%, nilai *false negative* bernilai 5.45%, dan nilai *true positive* bernilai 55.45%. Hasil ini menunjukkan bahwa model berhasil mengklasifikasikan 61 sampel positif dan 18 sampel negatif dengan benar. Modelnya juga salah mengklasifikasi 25 sampel negatif sebagai positif dan 6 sampel positif sebagai negatif.

Visualisasi *Decision Tree* dalam *depth 5* dapat dilihat pada Gambar 12. Sedangkan hasil *confusion matrix* dari *Decision Tree* dalam *depth 5* dapat dilihat pada Gambar 13.



Gambar 12. Visualisasi Decision Tree Dalam Depth 5

Gambar 13. Confusion Matrix Decision Tree Dalam Depth 5

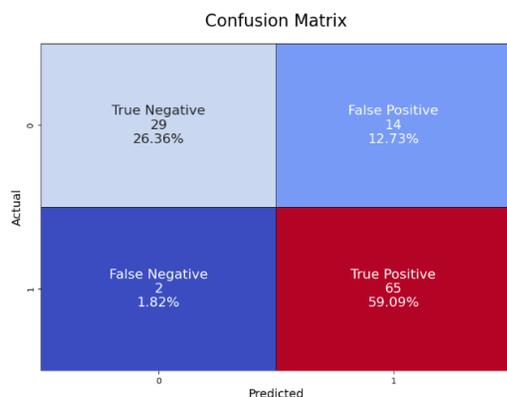
Hasil dari *confusion matrix* menggunakan *depth* 5 menunjukkan bahwa nilai *true negative* bernilai 15.45%, nilai *false positive* bernilai 23.64%, nilai *false negative* bernilai 5.45%, dan nilai *true positive* bernilai 55.45%. Hasil ini menunjukkan bahwa model berhasil mengklasifikasikan 61 sampel positif dan 17 sampel negatif dengan benar. Modelnya juga salah mengklasifikasi 26 sampel negatif sebagai positif dan 6 sampel positif sebagai negatif.

Untuk model klasifikasi kedua menggunakan algoritma *Naive Bayes*, hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Tabel Performa *Naive bayes*

	Positif	Negatif
<i>Accuracy</i>	0.854	
<i>Precision</i>	0.82	0.94
<i>Recall</i>	0.97	0.67
<i>F1-score</i>	0.89	0.78
<i>Support</i>	67	43

Hasil *confusion matrix* dari *Naive Bayes* dapat dilihat pada Gambar 14.



Gambar 14. Confusion Matrix *Naive Bayes*

Hasil dari *confusion matrix* menggunakan *Naive Bayes* menunjukkan bahwa nilai *true negative* bernilai 26.36%, nilai *false positive* bernilai 12.73%, nilai *false negative* bernilai 1.82%, dan nilai *true positive* bernilai 59.09%. Hasil ini menunjukkan bahwa model berhasil mengklasifikasikan 65 sampel positif dan 29 sampel negatif dengan benar. Modelnya juga salah mengklasifikasi 14 sampel negatif sebagai positif dan 2 sampel positif sebagai negatif.

Berdasarkan hasil performa dari kedua model algoritma yang dilatih, dapat dikatakan dalam kasus ini, performa yang didapatkan dari model *Naive Bayes* lebih tinggi daripada performa dari model *Decision Tree*. Hal ini bisa terjadi karena *dataset* yang dipakai di dalam penelitian ini berjumlah kecil dan memiliki fitur yang bersifat sederhana dan sedikit. Selain itu, pemilihan parameter pada model *Decision Tree* seperti *max_depth* dan *min_sample_leaf* juga dapat mempengaruhi tingkat akurasi dari model yang dilatih.

IV. KESIMPULAN

Berdasarkan hasil dari pengujian yang dilaksanakan oleh peneliti, dapat disimpulkan bahwa arah sentimen masyarakat terhadap naturalisasi pemain Timnas Indonesia lebih condong ke arah positif dibandingkan negatif. Model klasifikasi yang dilatih menggunakan algoritma *Decision Tree* memiliki akurasi sebesar 70% pada *depth* bernilai 3, 71,8% pada *depth* bernilai 4, dan 70,9% pada *depth* bernilai 5. Hal ini menunjukkan bahwa dengan 3 nilai parameter *depth* berbeda yang diuji, *depth* bernilai 4 memberikan tingkat akurasi yang terbaik. Sedangkan model klasifikasi menggunakan algoritma *Naive Bayes* memiliki akurasi sebesar 85,4%. Hasil ini menunjukkan bahwa algoritma *Naive Bayes* merupakan salah satu algoritma yang cukup efektif dalam membangun suatu model klasifikasi dalam kasus analisis sentimen. Melalui penelitian ini, peneliti berharap hasil dari penelitian ini mampu menolong tim administrasi sepak bola beserta pemainnya dalam memahami persepsi dan emosi publik terhadap keputusan-keputusan yang diciptakan. Selain itu, hasil penelitian ini juga diharapkan mampu membantu tim administrasi sepak bola menciptakan kebijakan mengenai naturalisasi yang lebih bersifat transparan dan memegang akuntabilitas.

REFERENSI

- [1] M. H. Joviansyah, L. N. Alyssa, I. S. Rahadatul'aisyi, G. N. Bakry, and N. Aristi, "Analisis Sentimen Dan Jaringan Komentar Video Youtube Najwa Shihab 'piala Dunia U-20 Gagal Digelar Di Indonesia. Mari Lihat Dari Dua Perspektif,'" Universitas Katolik Soegijapranata, vol. 4, no. 1, pp. 2–4, Nov. 2023, doi: 10.24167/jkm.v4i1.10351.
- [2] M. A. Bimasakti, "Problematisasi Konstitusionalitas Naturalisasi di Indonesia," Pancasila: Jurnal Keindonesiaan, vol. 3, no. 1, pp. 103–104, Apr. 2023, doi: 10.52738/pjk.v3i1.149.
- [3] G. K. Annas and N. M. Hazzar, "Analisis Persamaan Hak Kewarganegaraan Bagi Pemain Naturalisasi Sepakbola Di Indonesia," WICARANA, vol. 2, no. 2, p. 130, Oct. 2023, doi: 10.57123/wicarana.v2i2.37.
- [4] D. A. Prasetya, "PSSI Minta Netizen tak Lagi Suka Menghujat Pemain Timnas Indonesia: Nanti Bikin Takut Calon Pemain Naturalisasi," Bola.net, May 09, 2024. Accessed: May 29, 2024. [Online]. Available: https://www.bola.net/tim_nasional/pssi-minta-netizen-tak-lagi-suka-menghujat-pemain-timnas-indonesia-nanti-bikin-takut-calon-pe-49d005.html
- [5] M. F. Asshiddiqi and K. M. Lhaksmana, "Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI," Universitas Telkom Bandung, vol. 7, no. 3, p. 9938, Dec. 2020.
- [6] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree,knn, Dan Naive Bayes," INTI Nusa Mandiri, vol. 15, no. 1, p. 88, Aug. 2020, doi: 10.33480/inti.v15i1.1433.
- [7] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," IJCIT (Indonesian Journal on Computer and Information Technology), vol. 6, no. 2, p. 119, Jul. 2021, doi: 10.31294/ijcit.v6i2.10438.
- [8] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," Universitas Djuanda, vol. 3, no. 2, pp. 1861–1863, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.

-
- [9] F. Afshoh, "Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter," UNIVERSITAS MUHAMMADIYAH SURAKARTA, 2017.
- [10] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," INOVTEK Polbeng - Seri Informatika, vol. 3, no. 1, p. 51, Jun. 2018, doi: 10.35314/isi.v3i1.335.
- [11] A. F. Watratan, A. Puspita. B, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," Journal of Applied Computer Science and Technology, vol. 1, no. 1, p. 9, 2020, doi: 10.52158/jacost.v1i1.9.
- [12] E. A. Lisangan, A. Gormantara, and R. Y. Carolus, "Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia," KONSTELASI: Konvergensi Teknologi dan Sistem Informasi, vol. 2, no. 1, p. 26, Apr. 2022, doi: 10.24002/konstelasi.v2i1.5609.
- [13] I. T. Julianto and Lindawati, "Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut," Jurnal Algoritma, vol. 19, no. 1, p. 459, 2022, doi: 10.33364/algoritma/v.19-1.1112.
- [14] O. P. Moerdyanto and I. K. D. Nuryana, "Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma Decision Tree," JINACS, Universitas Negeri Surabaya, vol. 5, no. 1, p. 91, 2023.
- [15] A. Novandya, "Klasifikasi Data Cuaca Dengan Optimasi Decision Tree Berbasis Particle Swarm Optimization Dan Adaboost," SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER NUSA MANDIRI JAKARTA, 2017.