

Perbandingan Metode *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* pada Analisis Sentimen terhadap Naturalisasi Timnas Berdasarkan Komentar di Instagram

Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes Methods in Sentiment Analysis of National Team Naturalization Based on Comments on Instagram

Andrianus Dwi Haryo¹, J.B. Budi Darmawan^{2*}

¹ Jurusan Informatika, Universitas Sanata Dharma

^{2*} Jurusan Informatika, Universitas Sanata Dharma

Email : ddwi1327@gmail.com

*Email Korespondensi : b.darmawan@usd.ac.id

ABSTRAK

Sepak bola merupakan olahraga paling populer di Indonesia, dengan Tim Nasional sebagai simbol kebanggaan nasional. Salah satu isu yang memicu perdebatan publik adalah kebijakan naturalisasi pemain asing untuk memperkuat Timnas. Opini publik terkait isu ini banyak ditemukan di media sosial, terutama Instagram. Untuk mengetahui sentimen ekspresi pendapat masyarakat tentang berbagai topik, termasuk isu-isu terkait sepak bola dan Timnas digunakan analisis sentimen. Penelitian ini bertujuan untuk menganalisis sentimen dari 4.258 komentar Instagram terhadap isu naturalisasi dengan membandingkan unjuk kerja metode *Multinomial Naïve Bayes* dan algoritma *Bernoulli Naïve Bayes*. Proses analisis meliputi *preprocessing*, penerjemahan data menggunakan Google Translate API, pelabelan dengan *VADER*, pembobotan TF-IDF, serta penyeimbangan data menggunakan *SMOTE*. Evaluasi dilakukan menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa *Multinomial Naïve Bayes* memberikan unjuk kerja lebih baik dengan *accuracy* 74,65%, *precision* 75,96%, *recall* 74,65%, dan *F1-score* 74,32% pada skenario *k-fold* 10 dan *alpha* 0,01 dibandingkan menggunakan *Bernoulli Naïve Bayes*.

Kata kunci: analisis sentimen, instagram, *multinomial naïve bayes*, *bernoulli naïve bayes*, naturalisasi pemain asing

ABSTRACT

Football is the most popular sport in Indonesia, with the National Team serving as a symbol of national pride. One of the most debated issues is the naturalization of foreign players to strengthen the team. Public opinions on this topic are widely expressed on social media, especially Instagram. This study aims to analyze sentiment from 4,258 Instagram comments related to the naturalization issue by comparing the performance of *Multinomial Naïve Bayes* and *Bernoulli Naïve Bayes* Methods. The analysis process includes *preprocessing*, translation using the Google Translate API, sentiment labeling with *VADER*, TF-IDF vectorization, and data balancing using *SMOTE*. The classification performance is evaluated using a *confusion matrix* with metrics such as *accuracy*, *precision*, *recall*, and *F1-score*. The results show that *Multinomial Naïve Bayes* achieves better performance with 74.65% accuracy, 75.96% precision, 74.65% recall, and 74.32% F1-score under 10-fold cross-validation and alpha 0.01 compared to *Bernoulli Naïve Bayes*.

Keyword : sentiment analysis, instagram, *multinomial naïve bayes*, *bernoulli naïve bayes*, naturalization of foreign players

PENDAHULUAN

Sepak bola adalah salah satu olahraga paling populer di Indonesia, dengan Tim Nasional Sepakbola Indonesia (Timnas) menjadi simbol kebanggaan bagi jutaan penggemarnya. Setiap pertandingan Timnas selalu menjadi sorotan, baik di media massa maupun platform daring. Ulasan serta komentar dari penggemar dan publik umum sering kali membanjiri media sosial, forum, serta platform diskusi lainnya. Sebagai salah satu negara dengan populasi terbesar di dunia, Indonesia juga memiliki penggemar sepak bola yang sangat antusias. Sebanyak 90,8% dari populasi Indonesia dapat mengidentifikasi olahraga sepak bola, dan dalam kelompok tersebut, sekitar 46,7% menunjukkan minat aktif terhadap olahraga sepak bola di Indonesia [1].

Dalam beberapa tahun terakhir, isu naturalisasi pemain asing untuk memperkuat Timnas Indonesia menjadi topik yang hangat diperbincangkan. Kebijakan naturalisasi ini bertujuan untuk meningkatkan kualitas dan daya saing tim nasional di tingkat internasional. Namun, kebijakan tersebut menimbulkan beragam reaksi dari masyarakat, mulai dari dukungan penuh hingga penolakan keras. Perdebatan ini mencerminkan kompleksitas sentimen publik terhadap identitas nasional, prestasi olahraga, dan strategi pengembangan sepak bola Indonesia [2].

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat mengekspresikan pendapat mereka. Media sosial kini menjadi platform utama untuk berbagi opini, kritik, dan dukungan terhadap berbagai isu, termasuk naturalisasi pemain Timnas. Instagram sebagai layanan jejaring sosial berbagi foto dan video sangat banyak digunakan oleh masyarakat Indonesia untuk mengekspresikan pendapat mereka tentang berbagai topik, termasuk isu-isu terkait sepak bola dan Timnas. Karakteristik platform ini yang memungkinkan interaksi melalui komentar dan *caption* menjadikannya sumber data yang kaya untuk analisis sentimen. Analisis sentimen adalah proses mengkategorikan opini tekstual ke dalam kelompok-kelompok seperti "positif", "negatif". Analisis subjektivitas, penggalan opini, dan ekstraksi evaluasi adalah beberapa nama lain untuk itu [3].

Perkembangan teknologi memungkinkan sentimen publik ini dianalisis menggunakan metode klasifikasi *machine learning*. Salah satunya adalah metode klasifikasi *Naïve Bayes*. Varian klasifikasi *Naïve Bayes* yang digunakan dalam pengolahan teks diantaranya *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes*. Model *Multinomial Naïve Bayes* menggunakan jumlah kemunculan kata dalam dokumen, sedangkan model *Bernoulli Naïve Bayes* menggunakan kemunculan kata dalam dokumen dan mengabaikan jumlah kemunculannya [4].

Beberapa penelitian menggunakan algoritma *machine learning Naïve Bayes* dan *K-Means* untuk menganalisis opini publik terkait naturalisasi pemain sepak bola Timnas Indonesia dengan hasil yang menunjukkan efektivitas metode ini dalam mengklasifikasikan sentimen masyarakat ke dalam kelompok negatif, netral, dan positif [5][6]. Penelitian oleh Adji menunjukkan analisis sentiment opini perkuliahan yang dilakukan secara daring menggunakan algoritma *machine learning Multinomial Naïve Bayes* menghasilkan *Accuracy* sebesar 79,54% [7]. Penelitian mengenai analisis sentiment pemindahan ibu kota negara menghasilkan *sensitivity* 93,45 menggunakan algoritma *Bernoulli Naïve Bayes* dan *sensitivity* 90,19% menggunakan *Multinomial Naïve Bayes* [8].

Oleh karena itu, peneliti tertarik melakukan perbandingan metode *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* dalam melakukan analisis sentimen terhadap kebijakan naturalisasi pemain Timnas Indonesia di media sosial Instagram dengan pelabelan secara otomatis menggunakan pustaka *VADER*. Untuk mengetahui unjuk kerja kedua varian metode *Naïve Bayes* tersebut, setiap algoritma diuji unjuk kerjanya melalui metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score* untuk mengetahui tingkat efektivitas model dalam mengklasifikasikan opini publik.

METODE PENELITIAN

Tahapan-tahapan untuk memperoleh hasil dari klasifikasi analisis sentiment terhadap kebijakan naturalisasi pemain Timnas Indonesia ditampilkan pada Gambar 1.

Pengumpulan Data

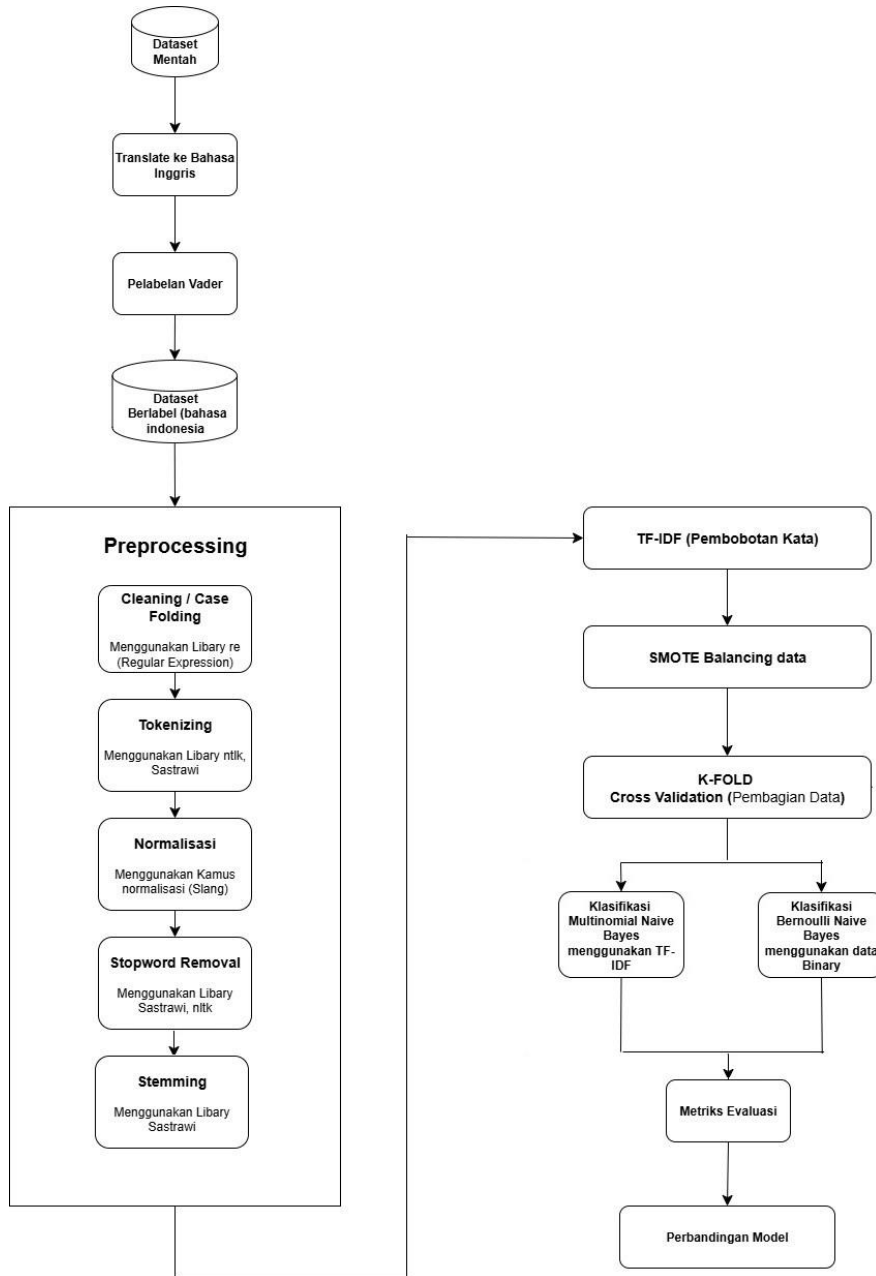
Proses ini dilakukan dengan menggunakan *google IG Comment Exporter* dengan cara menyalin *link* postingan Instagram yang komentarnya ingin diambil. Data yang diambil adalah data komentar mengenai tanggapan masyarakat tentang naturalisasi pemain Timnas Indonesia di laman Instagram Timnas Indonesia, dan data komentar yang dikumpulkan pada tanggal 10 maret 2025 sampai dengan 25 maret 2025 dengan menggunakan tehnik *scraping* dan mendapatkan total sebanyak 4.999 komentar pada postingan @timnasindonesia tentang naturalisasi pemain.

Translate ke Bahasa Inggris

Proses ini menggunakan pustaka *pandas* dan *deep_translator* dengan *Google Translate* untuk menerjemahkan teks dalam kolom komentar dari bahasa Indonesia ke bahasa Inggris.

Pelabelan VADER

Data Instagram yang akan digunakan belum mempunyai label sentimen sehingga data perlu diberikan label terlebih dahulu. Proses labelling dilakukan dengan bantuan pustaka *NLTK (Natural Language Toolkit)* pada Python yaitu *VADER* dengan mencari nilai polaritas untuk menentukan nilai positif dan negatif. Di mana nilai positif akan bernilai > 0.5 dan nilai negatif akan bernilai < 0.5 .



Gambar 1. Diagram alur penelitian

Preprocessing

Tahapan pada *preprocessing* pada penelitian ini yaitu *cleaning*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Pada tahap *cleaning* dilakukan untuk menghapus karakter atau elemen yang tidak diperlukan seperti angka, tanda baca, simbol, emoji, dan URL. Tahap *case folding* mengubah karakter huruf dalam dokumen ke dalam huruf kecil (*lowercase*) agar format teks menjadi sama. Tahap *tokenizing* merupakan proses membagi kalimat menjadi token-token berupa kata. Selanjutnya adalah tahap normalisasi untuk memperbaiki ejaan kata yang disingkat atau diperpanjang menjadi kata baku yang sesuai dengan KKBI. Normalisasi sangat penting dalam konteks analisis media sosial karena banyaknya penggunaan bahasa tidak formal. Tahap *stopword removal* untuk menghapus kata-kata umum yang dianggap tidak memiliki makna signifikan atau relevansi dalam konteks analisis teks dengan menggunakan pustaka *nltk*. Selanjutnya tahap *stemming* mengubah kata-kata menjadi bentuk dasarnya dengan menggunakan pustaka Sastrawi.

Pembobotan Kata TFIDF

Untuk mengubah teks menjadi representasi numerik, digunakan metode *Term Frequency Inverse Document Frequency* (TFIDF) menggunakan pustaka *scikit-learn*. Metode ini memberi bobot lebih besar pada

kata yang sering muncul di sebuah dokumen namun jarang muncul di seluruh korpus, sehingga membantu memfokuskan model pada kata-kata penting. Rumus untuk mencari TF-IDF dapat disajikan pada persamaan (1) [9].

$$w_{i,j} = t_{f_{i,j}} \times idf = t_{f_{i,j}} \times \log\left(\frac{N+1}{df_{i+1}}\right) + 1 \quad (1)$$

Keterangan :

$w_{i,j}$: Bobot kata- i pada dokumen ke- j
$t_{f_{i,j}}$: Jumlah kata- i pada dokumen ke- j
idf_i	: <i>Inverse document frequency</i> dari kata i
N	: Jumlah keseluruhan dokumen
df_i	: Jumlah dokumen yang mengandung kata ke- i

Balancing Data Menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*

SMOTE merupakan teknik *over-sampling* yang digunakan untuk mengatasi permasalahan *dataset* yang tidak seimbang. *SMOTE* memodifikasi kumpulan data yang tidak seimbang dan menghasilkan kumpulan data yang seimbang dari kumpulan data yang tidak seimbang. *SMOTE* mendistribusikan *instance* kelas mayoritas dan kelas minoritas secara merata. Teknik *SMOTE* meningkatkan *accuracy* prediksi pada kelas minoritas dengan membuat *instance* sintetik dari kelas minoritas tersebut [10].

Klasifikasi *Multinomial Naïve Bayes*

Model *Multinomial* dalam klasifikasi *Naïve Bayes* merupakan model yang memperhatikan frekuensi kata dalam dokumen. Dalam *Multinomial Naïve Bayes*, perhitungan probabilitas *posterior* untuk suatu kelas c setelah melihat data d dapat menggunakan persamaan (2) [4].

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (2)$$

Keterangan :

$P(c d)$: Probabilitas <i>posterior</i> bahwa sampel d termasuk dalam kelas c .
$P(c)$: Probabilitas <i>prior</i> dari kelas c .
$P(t_k c)$: Probabilitas bersyarat dari kata t_k yang muncul di dokumen yang dimiliki kelas c .

Kategori terbaik dalam klasifikasi *Multinomial Naïve Bayes* adalah kategori yang memiliki nilai maksimal atau *maximum a posterior (MAP) class* persamaan C_{map} menggunakan persamaan 3.

$$C_{map} = \arg \max P(c|d) \quad (3)$$

Persamaan (4) adalah untuk menghitung probabilitas *prior* dengan N_c merupakan jumlah sampel dalam kelas c dan N merupakan jumlah sampel dalam *dataset*.

$$P(c) = \frac{N_c}{N} \quad (4)$$

Probabilitas suatu *term* (kata) t diberikan kelas c menggunakan *Laplace Smoothing* dengan nilai $\alpha = 1$ dapat dituliskan pada persamaan (5) [4].

$$P(t|c) = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'} + B')} \quad (5)$$

Keterangan :

$P(t c)$: Probabilitas suatu <i>term</i> (kata) t diberikan kelas c (disebut <i>likelihood probability</i>).
B'	: Total kata pada keseluruhan kelas.
T_{ct}	: Jumlah kemunculan <i>term</i> t dalam semua dokumen yang berlabel kelas c .
$\sum_{t' \in V} T_{ct'}$: Total frekuensi semua kata dalam <i>vocabulary</i> V yang muncul di kelas c (disebut total kata dalam kelas c).
\propto	: Nilai <i>smoothing</i> . Penambahan \propto (α)=1, disebut dengan <i>Laplace Smoothing</i> .

Klasifikasi *Bernoulli Naïve Bayes*

Model *Bernoulli* dalam klasifikasi *Naïve Bayes* merupakan model yang menggunakan kemunculan kata dalam dokumen dan mengabaikan jumlah kemunculannya. Dalam *Bernoulli Naïve Bayes*, perhitungan probabilitas *posterior* untuk suatu kelas c setelah melihat data d dapat menggunakan persamaan (6) [4].

$$P(c|d) \propto P(c) \prod_{1 \leq i \leq M} P(d_i|c) \quad (6)$$

Keterangan :

$P(d_i|c)$: *Conditional probability*.

$P(c)$: *Prior probability* dokumen yang berada di kategori c .

Conditional probability dihitung pada dokumen latih yang terdapat pada kategori c yang mengandung *term* (pada sembarang posisi yang dapat muncul beberapa kali). Perhitungan *Conditional probability* dapat dilihat pada persamaan (7) [4].

$$P(d_i|c) = \prod_{1 \leq i \leq M} ((B_{it}P(e_i|c)) + ((1 - B_{it})(1 - P(e_i|c))) \quad (7)$$

Keterangan :

$P(d_i|c)$: Probabilitas dokumen d diberikan dokumen tersebut termasuk dalam kelas c

$P(e_i|c)$: Merupakan probabilitas dalam kategori c , dimana *term* (dapat muncul beberapa kali dan dalam sembarang posisi)

B_{it} : Jika *term* bernilai 0 maka tidak muncul dalam dokumen dan jika *term* bernilai 1 akan muncul pada dokumen.

Dalam model *Bernoulli* ketidakmunculan suatu *term* menjadi faktor yang diperhitungkan sehingga akan mempengaruhi hasil klasifikasi. Persamaan untuk menghitung $P(e_i|c)$ disajikan pada persamaan (8) [4].

$$P(e_i|c) = \frac{df_i + 1}{df' + 2} \quad (8)$$

Keterangan :

df_i : Merupakan jumlah dokumen latih pada suatu kategori c yang mengandung *term*.

df' : Merupakan jumlah seluruh dokumen pada kategori c .

1 : Merupakan *Laplace smoothing* agar terhindar dari *zero probability*

2 : Nilai konstan yang menunjukkan kejadian muncul tidak munculnya *term*.

Metrik Evaluasi

Evaluasi unjuk kerja dengan *confusion matrix* dilakukan dengan cara membandingkan hasil aktual dengan hasil prediksi klasifikasi. Terdapat empat elemen dalam *confusion matrix* untuk merepresentasikan nilai hasil klasifikasi, yaitu *True Positive* (TP) merupakan prediksi benar untuk kelas positif, *True Negative* (TN) merupakan prediksi benar untuk kelas negatif, *False Positive* (FP) merupakan prediksi salah sebagai positif (seharusnya negatif) dan *False Negative* (FN) merupakan prediksi salah sebagai negatif (seharusnya positif) [11]. Dari *confusion matrix*, kita dapat menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score*.

Accuracy merupakan persentase prediksi yang benar dari total prediksi. *Accuracy* diperoleh menggunakan persamaan (9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

Precision merupakan proporsi prediksi positif yang benar dari semua prediksi positif. *Precision* diperoleh menggunakan persamaan (10).

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

Recall merupakan proporsi kasus positif yang berhasil diprediksi dengan benar. *Recall* diperoleh menggunakan persamaan (11).

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

F1-score merupakan rata-rata harmonik dari *precision* dan *recall*, mengukur keseimbangan antara keduanya. *F1-score* diperoleh menggunakan persamaan (12).

$$F1 - score = \frac{2(Recall \times Precision)}{Recall + Precision} \quad (12)$$

Skenario Pengujian

Dalam tahapan pengujian ini dilakukan perbandingan kinerja pada *dataset* untuk melihat model mana yang lebih baik. Model *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* dievaluasi menggunakan *k-fold cross validation* dengan berbagai variasi *k* (1, 3, 5, 7, 9, 10) dan berbagai variasi nilai *alpha* (0.01, 0.1, 1, 10). Selanjutnya, dilakukan evaluasi unjuk kerja untuk setiap kombinasi *k-fold* dan *alpha* menggunakan metrik evaluasi *accuracy*, *precision*, *recall* dan *F1-score* untuk menentukan model dan variasi parameter yang optimal.

HASIL DAN ANALISIS

Pada bagian ini dijelaskan analisis hasil dari proses pengolahan data yang meliputi tahap-tahap seperti pengumpulan data, terjemahan data, *labeling VADER*, *Preprocessing*, *TF-IDF*, *SMOTE*, Pengujian metode dengan membagi data menjadi data latih dan data uji menggunakan (*k-fold cross validation*), klasifikasi model, evaluasi dengan menggunakan *confusion matrix* dan analisis hasil yang diperoleh.

Setelah mendapatkan *dataset* dengan kolom komentar, *dataset* ini disimpan dalam format file CSV. Selanjutnya kolom komentar yang awalnya berbahasa Indonesia diterjemahkan ke dalam bahasa Inggris menggunakan pustaka *googletrans*. Setelah diperoleh data yang telah diterjemahkan selanjutnya dilakukan pelabelan untuk menentukan label positif atau negatif pada *dataset* ini menggunakan pustaka *VADER* di *python*. Setelah mendapatkan label menggunakan *VADER*, dilakukan *preprocessing* meliputi data *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Setelah melewati pelabelan dan *preprocessing*, *dataset* yang semula berjumlah 4999 berkurang menjadi 4258 dikarenakan terdapat adanya komentar emoji atau kosong sehingga data tersebut dihilangkan. Selanjutnya *dataset* melalui proses pembobotan kata menggunakan *TF-IDF*. Pada tahap ini data berlabel positif sekitar 70 % dan data berlabel negatif sekitar 30 % yang menunjukkan data tidak seimbang. Selanjutnya data dibuat seimbang dengan teknik *SMOTE* sehingga diperoleh *dataset* sejumlah 5922 dengan masing-masing label positif dan negatif menjadi 2961.

Dari pengujian klasifikasi *dataset* yang telah seimbang dengan metode *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* menggunakan skenario pengujian di atas diperoleh hasil yang disajikan pada Tabel 1.

Dari hasil pengujian variasi *k-fold* dengan nilai *k* = 1, 3, 5, 7, 9, 10 mempengaruhi *accuracy* yang diperoleh, untuk metode *Multinomial Naïve Bayes accuracy* optimal diperoleh saat nilai *k* = 10 sedangkan untuk metode *Bernoulli Naïve Bayes accuracy* optimal diperoleh saat nilai *k* = 7. Selain itu variasi nilai *alpha* = 0,01, 0,1, 1, 10 juga mempengaruhi *accuracy* yang diperoleh, untuk kedua metode *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes accuracy* optimal diperoleh saat nilai *alpha* = 0.01.

Tabel 1. Perbandingan unjuk kerja dengan *accuracy* terbaik untuk metode *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes*

Variabel	<i>Multinomial Naïve Bayes terbaik</i>	<i>Bernoulli Naïve Bayes terbaik</i>
<i>k-fold</i>	10	7
<i>alpha/ var_smoothing</i>	0,01	0,01
<i>accuracy</i> (%)	74,65	72,09
<i>precision</i> (%)	75,96	73,17
<i>recall</i> (%)	74,65	72,09
<i>F1-score</i> (%)	74,32	71,75

Dapat dilihat dari Tabel 1. bahwa *accuracy* optimal diperoleh dengan menggunakan metode *Multinomial Naïve Bayes* dengan nilai terbaik untuk *accuracy* 74,65 % dan *F1-score* 74,32% mengungguli metode

Bernoulli Naïve Bayes dengan nilai terbaik untuk *accuracy* 72,09 % dan *F1-score* 71,75%. Hal ini menunjukkan bahwa untuk analisis sentimen terhadap kebijakan naturalisasi pemain Timnas Indonesia di media sosial Instagram menggunakan metode *Multinomial Naïve Bayes* dengan kemampuan memodelkan jumlah kemunculan (frekuensi) kata dalam dokumen menghasilkan *accuracy* yang lebih baik dibandingkan metode *Bernoulli Naïve Bayes* yang hanya menggunakan kemunculan kata dalam dokumen.

KESIMPULAN

Berdasarkan hasil penelitian perbandingan metode algoritma *Multinomial Naïve Bayes* dan *Bernoulli Naïve Bayes* pada analisis sentimen terhadap kebijakan naturalisasi pemain Timnas Indonesia di media sosial Instagram menggunakan data sebanyak 4258, metode *Multinomial Naïve Bayes* menunjukkan unjuk kerja lebih optimal dengan *accuracy* sebesar 74,65% *precision* 75,96% *recall* 74,65% *F1-score* sebesar 74,32% pada *k-fold* 10 dan *alpha* 0.01 dibandingkan dengan dengan metode *Bernoulli Naïve Bayes* dengan *accuracy* sebesar 72,09% *precision* 73,17% *recall* 72,09% *F1-score* sebesar 71,75% pada *k-fold* 7 dan *alpha* 0,01.

DAFTAR PUSTAKA

- [1] B. Pradiksa and D. A. Prianto, "Persepsi masyarakat terhadap sepakbola Indonesia di masa pandemi Covid-19," *Jurnal Prestasi Olahraga*, vol. 5, no. 2, pp. 53–59, 2022.
- [2] D. R. P. Jaya and S. Lestari, "Analisis Sentimen Naturalisasi Tim Nasional Indonesia U-23 di Era Shin Tae-yong Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 3262–3277, 2024.
- [3] J. Zhao, K. Liu, and L. Xu, "Sentiment analysis: Mining opinions, sentiments, and emotions." MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ..., 2016.
- [4] C. D. Manning, *An introduction to information retrieval*. 2009.
- [5] A. Tegar, N. E. Fresia, and M. Habibi, "Pemetaan Opini Publik Menggunakan Data Mining: Studi Kasus Naturalisasi Pemain Sepak Bola dengan K-Means dan Naïve Bayes Classifier," *Indonesian Journal On Data Science Ypedumelu: Universitas Jenderal Achmad Yani Yogyakarta*, vol. 2, no. 1, pp. 43–50, 2024.
- [6] A. S. Adhityas, *Analisis Sentimen terhadap Program Naturalisasi Timnas Indonesia pada X Menggunakan Algoritma Naïve Bayes*, eProceedings of Engineering, vol. 12, no. 2, pp. 1–9, 2025.
- [7] B. R. Adji and J. B. B. Darmawan, *Analisis Sentimen terhadap Kuliah Daring Menggunakan Multinomial Naïve Bayes dengan Seleksi Fitur Information Gain*, In The 2nd Seminar Nasional dan Prosiding Scitech 2023, 2023.
- [8] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara dengan Klasifikasi Naïve Bayes untuk Model Bernoulli dan Multinomial," *Jurnal Gaussian*, vol. 9, no. 3, pp. 237–246, 2020.
- [9] S. Jansen, *Hands-On Machine learning for Algorithmic Trading: Design and implement investment strategies based on smart algorithms that learn from data using Python*. Packt Publishing Ltd, 2018.
- [10] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique," *Decision Analytics*, vol. 2, no. 1, p. 1, 2015.
- [11] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.