



Pemetaan Opini Publik Menggunakan Data Mining: Studi Kasus Naturalisasi Pemain Sepak Bola dengan K-Means dan Naive Bayes Classifier

Tegar Agustian^{a,1,*}, Emilia Fresia Nandela^{b,2}, Stani .A. Sinay^{b,3}, Muhammad Habibi^{b,4}

^{a,b} Informatika, FTTI Unjaya, Yogyakarta, Indonesia

¹ agustiantegarr@gmail.com*; ² emeliafresianandela@gmail.com; ³ rstapat@gmail.com,

⁴ Muhammadhabibi17@gmail.com

* corresponding author

ABSTRAK

ARTICLE INFO

Naturalisasi merupakan salah satu proses yang dilakukan oleh warga asing agar menjadi Warga Negara Indonesia (WNI) yang sah di mata hukum. Saat ini Timnas Indonesia memiliki beberapa pemain naturalisasi. Beberapa kalangan menyambut positif kehadiran mereka, melihatnya sebagai langkah strategis untuk meningkatkan kualitas dan daya saing tim. Namun, ada pula yang merasa skeptis dan meragukan keberlanjutan dukungan terhadap pemain lokal. Data yang diambil dari 3584 komentar YouTube melalui YouTube Data API mencerminkan keragaman opini yang dapat memberikan gambaran lebih mendalam tentang dinamika pandangan publik. Penelitian ini penting dalam konteks pemahaman pandangan masyarakat terhadap naturalisasi pemain sepak bola Timnas. Dengan menggunakan teknik *Data Mining*, terutama *K-Means Clustering* dan *Naive Bayes Classifier*, penelitian ini memberikan wawasan mendalam tentang kelompok-kelompok masyarakat dengan perspektif serupa atau berbeda terkait isu tersebut. Hasil dari proses *K-Means Clustering* digunakan sebagai label awal untuk melatih model *Naive Bayes Classifier*. Evaluasi kinerja model dilakukan menggunakan *confusion matrix*, yang menghasilkan akurasi sebesar 93,17% dan *error rate* sebesar 6,83%. Proses ini dilakukan terhadap dataset komentar YouTube yang telah diberi label melalui *K-Means Clustering*. Hasil klasifikasi menggunakan metode *Naive Bayes* menunjukkan bahwa 3328 data komentar setuju dengan adanya naturalisasi pemain dan 256 data komentar tidak setuju.

Article history

Received: 29 February 2024

Revised: 14 May 2024

Accepted: 9 August 2024

Keywords

Data Mining

K-Means Clustering

Naive Bayes Classifier

Naturalisasi Pemain Sepak Bola

Opini Masyarakat.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

1. Pendahuluan

Sepak bola adalah salah satu olahraga yang paling digemari di Indonesia, dengan prestasi luar biasa Timnas Indonesia di bawah bimbingan pelatih yang berkualitas tentu menjadi pendorong bagi masyarakat di negeri ini [1]. Namun, keberhasilan Timnas Indonesia menyebabkan perbedaan pendapat di kalangan masyarakat mengenai keberadaan pemain-pemain yang naturalisasi, seperti yang terjadi pada kasus-kasus lainnya di mana teknologi informasi mempengaruhi perilaku masyarakat dalam melakukan kegiatan sehari-hari [2].

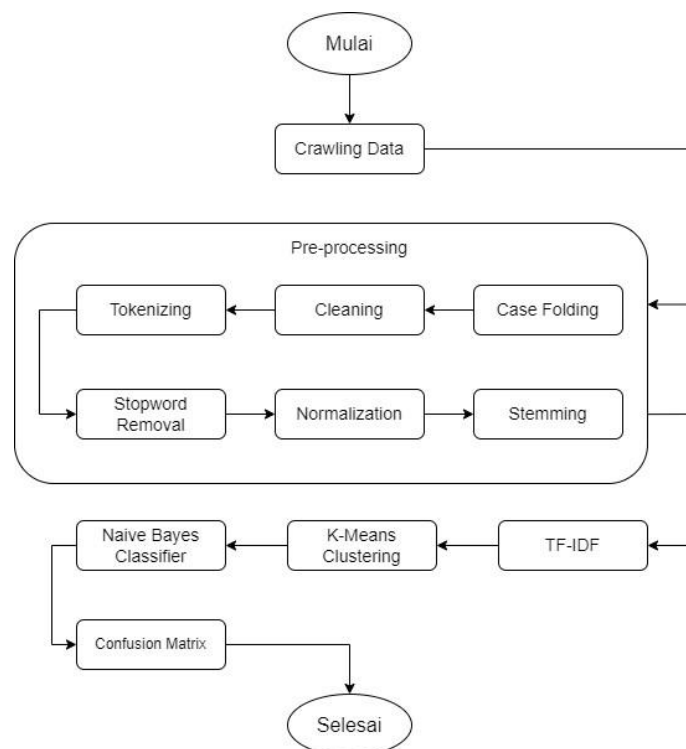


Setiap tahun, muncul pertanyaan etis dan identitas nasional di berbagai pemberitaan yang memicu eskalasi perdebatan tentang pemain sepak bola naturalisasi, yang dapat diatasi dengan menggunakan analisis sentimen untuk memahami opini masyarakat [3]. Sementara sekumpulan pendapat mengharapkan agar pemain-pemain tersebut tetap bermain demi meningkatkan wibawa dan kualitas Timnas, beberapa lainnya lebih skeptis terhadap keberlanjutan dukungan terhadap pemain lokal, yang dapat dianalisis menggunakan metode Naive Bayes untuk mengklasifikasikan opini masyarakat [4]. Penolakan naturalisasi pemain Timnas adalah alasan utama dari penulisan ini, yang dapat diatasi dengan menggunakan kombinasi Naive Bayes dan K-Means untuk mengelompokkan komentar yang memiliki sentimen serupa [5]. Dengan mengerti pandangan umum, kita memahami proses perdebatan dan landasan budaya dan sosial yang telah memicu persepsi dari masyarakat, yang dapat dianalisis menggunakan metode text mining [6]. Setelah memahami hal ini, kita dapat menggeser strategi komunikasi dan olahraga agar lebih bijak, dengan menggunakan teknologi informasi untuk meningkatkan kualitas pelayanan [7].

Penggunaan kombinasi Naive Bayes dan K-Means dalam analisis yang dilakukan memiliki dasar pembuktian yang kuat, seperti yang terbukti pada penelitian-penelitian sebelumnya [8]. Algoritme K-Means Clustering digunakan untuk mengelompokkan komentar yang memiliki sentimen serupa, yang akhirnya memberikan label-data, seperti yang dilakukan pada penelitian tentang analisis sentimen pada media sosial [9]. Selanjutnya, setelah elompokan penerapannya, Naive Bayes digunakan untuk melakukan klasifikasi terhadap data, seperti yang dilakukan pada penelitian tentang klasifikasi opini masyarakat pada media sosial [10].

2. Metode Penelitian

Penelitian ini merinci serangkaian langkah-langkah metodologis yang dilibatkan dalam menganalisis sentimen masyarakat terkait naturalisasi pemain sepak bola Timnas Indonesia. Dengan cermatnya pendekatan ini, penelitian dapat memberikan gambaran yang komprehensif terhadap dinamika opini publik.



Gambar 1. Tahap-tahap Penelitian

2.1. Crawling atau Pengumpulan Data

Data diperoleh melalui pemanfaatan *YouTube Data API* yang diotentikasi dengan kunci *API YouTube Developer*. Komentar-komentar yang berasal dari platform *YouTube* dianggap sebagai objek analisis, memastikan keberagaman pandangan masyarakat terkait isu kontroversial naturalisasi pemain Timnas Indonesia dalam dataset ini.

Proses pengumpulan data dimulai dengan pemilihan video yang akan menjadi fokus analisis dengan kriteria yang difokuskan pada relevansi terhadap topik naturalisasi pemain sepak bola, yang diharapkan dapat memberikan gambaran yang lebih komprehensif tentang dinamika opini publik.

Selanjutnya, dilakukan ekstraksi data komentar dari video tersebut melalui implementasi kode *Python* di lingkungan *Jupyter Notebook*. Video yang dipilih memiliki ID: "HWhPHTffIp4". Hasilnya, terdapat 3584 data komentar yang kemudian disimpan dalam format *Excel* untuk kemudahan pengolahan dan analisis selanjutnya.

2.2. Preprocessing Data

Komentar yang telah berhasil diambil masih dalam bentuk data mentah yang perlu melalui tahap *pre-processing* sebelum dapat diolah lebih lanjut. Proses *pre-processing* diperlukan untuk mempersiapkan data agar siap untuk tahapan analisis selanjutnya. Teknik *pre-processing* teks diimplementasikan untuk melakukan segmentasi pada teks, di mana hanya melalui karakteristik segmentasi teks tersebut dapat dinilai, dianalisis, dan diklasifikasikan. Beberapa langkah *pre-processing* yang dilakukan melibatkan segmentasi teks untuk memperoleh data yang lebih terstruktur dan siap digunakan dalam analisis lebih lanjut. Berikut ini adalah tahapan-tahapannya:

1. *Case Folding*

Case folding adalah proses mengonversi semua karakter dalam teks ke dalam huruf kecil atau huruf besar agar konsistensi teks terjaga. Hal ini dilakukan untuk mencegah ambiguitas yang dapat timbul jika terdapat variasi huruf besar dan kecil dalam analisis teks.

2. *Cleaning*

Cleaning atau pembersihan adalah tahap di mana data mentah atau teks awal dibersihkan dari elemen-elemen yang tidak diinginkan atau *noise*, seperti karakter khusus, angka, atau tanda baca. Tujuan dari tahap ini adalah untuk menyederhanakan dan membersihkan teks agar lebih mudah diolah.

3. *Tokenizing*

Tokenizing merupakan langkah memecah atau mengolah teks menjadi unit-unit yang lebih kecil serta jelas dan terdefinisi, yang disebut sebagai token. Token dapat berupa kata, frasa, atau bahkan karakter.

4. *Stopword Removal*

Stopword removal adalah proses menghapus kata-kata umum atau stopwords dari sebuah teks. *Stopwords* merupakan kata-kata yang seringkali muncul namun kurang memberikan makna signifikan dalam analisis, sehingga dianggap sebagai gangguan yang dapat dihilangkan agar perhatian lebih terfokus pada kata-kata kunci.

5. *Normalization*

Normalization adalah proses menyamakan atau mengonversi variasi bentuk kata ke dalam bentuk standar. Misalnya, mengonversi kata-kata ke bentuk dasar atau mengganti sinonim dengan kata-kata yang setara, sehingga analisis dapat lebih konsisten.

6. *Stemming*

Stemming adalah proses menghapus atau memotong akhiran kata agar hanya menyisakan bentuk dasarnya (*stem*). Hal ini dilakukan untuk mengurangi variasi kata dan menghasilkan representasi yang lebih sederhana. Contohnya, mengubah kata-kata ke bentuk dasar seperti "berlari" menjadi "lari".

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

TF-IDF, atau Term Frequency-Inverse Document Frequency, merupakan sebuah pendekatan statistik yang digunakan untuk menilai sejauh mana signifikansi suatu kata dalam suatu dokumen di dalam rangkaian dokumen atau korpus. Metode memberikan nilai bobot pada kata-kata dalam dokumen, dengan mempertimbangkan dua faktor utama:

1. *TF (Term Frequency)*

Mengukur seberapa sering suatu kata muncul dalam suatu dokumen. Rumusnya adalah jumlah kemunculan kata tersebut dalam dokumen dibagi dengan total kata dalam dokumen.

$$Tf_{t,d} = 1 + {}^{10}\text{Log } tf \quad (1)$$

Keterangan:

- tf = banyak kata pada dokumen.
- $Tf_{t,d}$ = banyak kata t pada dokumen d atau pembobotan lokal.

2. *IDF (Inverse Document Frequency)*

Mengukur seberapa jarang suatu kata muncul di seluruh korpus dokumen. Rumusnya adalah logaritma dari total dokumen dibagi dengan jumlah dokumen yang mengandung kata tersebut.

$$Idf_t = \frac{{}^{10}\text{Log } n}{df_t} \quad (2)$$

Keterangan:

- Idf_t = pembobotan global dokumen t .
- n = banyak dokumen.
- df_t = banyak dokumen dengan kata t .

3. *TF-IDF Weighting*

Proses penghitungan bobot *TF-IDF* suatu kata dalam suatu dokumen dihasilkan dengan mengalikan nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*.

$$Tf - Idf_{t,d} = Tf_{t,d} \times Idf_t \quad (3)$$

Keterangan:

- $Tf_{t,d}$ = banyak kata t pada dokumen d atau pembobotan lokal.
- Idf_t = pembobotan global dokumen t .
- $Tf - Idf_{t,d}$ = nilai akhir atau bobot akhir dari kata.

2.4. *K-Means Clustering*

K-Means Clustering adalah algoritma dalam analisis data dan pengelompokan (*clustering*) yang bertujuan untuk membagi sekelompok data menjadi k kelompok yang berbeda. Tujuan utama dari algoritma ini adalah untuk mengelompokkan data ke dalam cluster atau kelompok sedemikian rupa sehingga setiap titik data dalam kelompok memiliki kesamaan dengan titik lain dalam kelompok yang sama, dan sejauh mungkin berbeda dari titik-titik dalam kelompok lain.

Langkah-langkah algoritma *K-Means Clustering*:

1. Tentukan jumlah cluster k yang diinginkan.
2. Pilih nilai acak sebagai titik pusat awal (*centroid*) untuk setiap cluster sebanyak k .
3. Mengukur jarak setiap data ke setiap *centroid*, kemudian menentukan *centroid* terdekat untuk setiap data ke *centroid* dengan *euclidean distance*:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (4)$$

Keterangan:

- d = Jarak terdekat antara data kriteria
- x_i = Data kriteria ke- i , $i=1,2,\dots,n$.
- μ_j = Centroid kriteria pada cluster ke- j , $j=1,2,\dots,n$.

4. Mengelompokkan data berdasarkan seberapa dekatnya dengan centroid. Data yang memiliki nilai paling mendekati *centroid* akan diatributkan ke dalam *cluster* tersebut.
5. Memperbaharui nilai *centroid*.

$$C_j = \frac{1}{N_k} \sum_{j=1}^{N_k} X_j \quad (5)$$

Keterangan:

- C_j = Centroid baru.
 - N_k = Jumlah total data dalam cluster.
 - j = Data kriteria centroid pada cluster ke- j .
 - X_j = Data dalam cluster ke- $j, j=1,2,\dots,m$.
6. Iterasi dilakukan dari langkah 3 hingga langkah 5 hingga tidak ada perbedaan antara pusat centroid terakhir dan centroid sebelumnya. Iterasi terakhir menjadi acuan untuk menentukan pembagian cluster pada data.

2.5. Naive Bayes Classifier

Naive Bayes Classifier adalah algoritma klasifikasi yang merujuk pada teorema probabilitas Bayes. Teorema Bayes adalah suatu prinsip fundamental dalam teori probabilitas yang memberikan cara untuk memperbarui probabilitas suatu peristiwa berdasarkan informasi baru yang diperoleh. Dalam implementasinya, *Naive Bayes* memanfaatkan probabilitas prior dan *likelihood* untuk memprediksi kelas data baru. Persamaan teorema bayes dapat dinyatakan sebagai berikut :

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (6)$$

Keterangan:

- P = Probabilitas kelas.
- A = Hipotesis data kelas yang spesifik.
- B = Data yang akan diklasifikasikan.
- $P(A|B)$ = Probabilitas hipotesis A berdasarkan kondisi B .
- $P(B|A)$ = Probabilitas data B dari kondisi hipotesis A .
- $P(A)$ = Probabilitas dari kelas A .
- $P(B)$ = Probabilitas dari data B (*evidence*).

2.6. Confusion Matrix

Confusion matrix adalah pendekatan yang digunakan untuk menghitung akurasi dalam konsep penambangan data atau untuk mengukur kinerja proses klasifikasi. Menghitung akurasi menggunakan *confusion matrix* menghasilkan data dari model klasifikasi berupa *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*. Persamaan berikut untuk menghitung akurasinya.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (8)$$

$$Error\ rate = \frac{FP+FN}{TP+TN+FP+FN} \times 100\% \quad (9)$$

3. Hasil dan Pembahasan

3.1. Labeling Menggunakan K-Means Clustering

Tahap ini merupakan tahap mengelompokkan atau memberia kategori dimana kesamaan karakteristik data akan dikelompokkan menjadi *cluster* atau kelompok yang berbeda. Disini pengelompokannya berupa pendapat (setuju atau tidak setuju) terhadap maraknya naturalisasi pemain timnas.

	comment	pendapat
0	naturalisasi main maroko masuk piala dunia amin	setuju
1		aamiin setuju
2		aminn setuju
3		amiin yarobbal alamiin setuju
4	tidak naturalisasinegara inidk cipta bibit yang bgus kah	setuju
5		bibit bagus nya dengkul om setuju
6	pakstjanganlagiasalasanpilihpemainuntukkelasasiacontohkemarinlawanirakkamakalautimnassuksesbisabisapakstjakandiliriklupkuperopadikemudianharitidakmenutupkemungkinan	setuju
7		setujusengpenting masuk pildun setuju
8	lihat profil ide gila sintayong tim bentuk naturalisasi sintayong indonesia jawara asia lolos pal piala dunia katar	setuju
9	yth ketua pssi mohon kuwalifikasi piala asia gulir naturalisasi yang bagus tsb selesai timnas berani saing	setuju
10		ok kt masuk tim kelas dunia setuju
11		tuiu kalah neara banokit latih vano alam main vano alam bravo indonesia setuju

Gambar 2. Labeling

3.2. Model Naive Bayes Classifier

Dalam upaya memperhitungkan frekuensi kemunculan setiap kata dalam dokumen untuk menentukan kontribusinya terhadap kelas tertentu, seperti setuju atau tidak setuju, maka diperlukan analisis yang lebih mendalam. Oleh karena itu, dengan menggunakan metode Multinomial Naive Bayes (MultinomialNB) untuk mengukur dan memprediksi sentimen pada dokumen baru, fokus utama adalah pada pemahaman kontribusi kata-kata terhadap polaritas sentimen dalam tugas analisis teks. Selain itu, penggunaan teknik Term Frequency-Inverse Document Frequency (TF-IDF) juga diterapkan untuk menentukan bobot pentingnya setiap kata dalam dokumen, sehingga dapat membantu meningkatkan akurasi prediksi sentimen. Dengan demikian, analisis sentimen dapat dilakukan dengan lebih efektif dan efisien. Berikut ini perhitungan TF-IDF untuk data training:

Data Training - TF-IDF Representation:	
(0, 615)	0.6455952375249149
(0, 653)	0.7636797688070233
(1, 851)	0.1291459828256561
(1, 789)	0.23856176284076722
(1, 744)	0.186737831860953
(1, 704)	0.24527761195253547
(1, 487)	0.2584158011363192
(1, 504)	0.0812008847728811
(1, 562)	0.2303422278211976
(1, 226)	0.14968091277435783
(1, 946)	0.2708098197657289
(1, 234)	0.24527761195253547
(1, 778)	0.2356133745279121
(1, 625)	0.27600124610840343
(1, 874)	0.23856176284076722
(1, 102)	0.19070242778271893
(1, 685)	0.21459724649378195
(1, 992)	0.10361377501246267
(1, 291)	0.2303422278211976
(1, 948)	0.15253778504114474
(1, 980)	0.2179482091917879
(1, 411)	0.2640939706539606
(1, 324)	0.1915479751577652
(1, 194)	0.19241600137859446
(2, 14)	0.6225774626715125

Gambar 3. Hasil TF-IDF data training

Data Testing - TF-IDF Representation:	
(0, 690)	0.41665092016832256
(0, 689)	0.4513262311921612
(0, 504)	0.1418182987760908
(0, 488)	0.19122424535938698
(0, 365)	0.2619013698906107
(0, 357)	0.4555857478188203
(0, 154)	0.3261398161575278
(0, 121)	0.42838022950444427
(1, 980)	0.268586745112957
(1, 915)	0.5334509876571696
(1, 851)	0.15915202653045607
(1, 783)	0.25717869018502276
(1, 594)	0.22389050600528584
(1, 504)	0.10006726562386534
(1, 468)	0.17446451147995073
(1, 423)	0.3254541078586515
(1, 365)	0.18479811261504372
(1, 357)	0.16073109194960641
(1, 324)	0.3540789394740771
(1, 146)	0.268586745112957
(1, 91)	0.21080230969321126
(1, 34)	0.2310608671315565
(3, 941)	1.0
(4, 997)	0.10381678581733833
(4, 915)	0.2511989461401597

Gambar 4. Hasil TF-IDF data testing

3.3. Hasil Evaluasi Naive Bayes Classifier

Pada tahap ini, mendapatkan hasil klasifikasi sentimen dari komentar yang dianalisis. Memahami pola sentimen (positif dan negatif) dapat memberikan wawasan tentang bagaimana masyarakat merespons isu naturalisasi pemain. Penting untuk mengeksplorasi faktor-faktor yang mempengaruhi sentimen ini, seperti konteks sosial, budaya, atau bahkan peristiwa terkini yang relevan.

	comment	pendapat	prediksi nbc
0	bukti jika mng mnjdi juara jika pke pmain nturlasimnurt lbih bgus main lokal	setuju	setuju
1	warga indonesia latih sty perbanyaklah naturalisasi jika timnas indonesia pingin juara khusus timnas anda timnas senior main naturalisasi banyak bravo timnas indonesia	setuju	setuju
2		bungkus	setuju
3		tuju	tidak setuju
4	ide gila hasil kalah melulu shin tae yong gaya melulu main naturalisasi kalah melulu main latih orang asia timnas indonesia majumaju belakang bidang sepak bola urut peringkat dunia dunia sungguh milu sedih negara indonesia negeri dunia urut rangking dunia platih eropa kelas dunia yose morinho zinadin zidan timnas maju masa negara lho tanggungtanggung cari latih kelas dunia timnas indonesia masuk piala dunia sesuai mumpi rakyat bangsa indonesia timnas indonesia piala dunia tuan rumah timnas senior indonesia piala dunia junior	setuju	setuju
5		why not	setuju
6		hoaks hoaks	setuju
7		setuju	setuju

Gambar 5. Hasil Prediksi

3.4. Confusion Matrix

Ketika sudah mendapatkan hasil prediksinya, kita akan menguji *confusion matrix* untuk memeriksa nilai data aktual dan prediksi guna mengevaluasi keakuratan model yang telah dibuat. Ini membantu kita memahami sejauh mana model kita mampu memprediksi dengan benar dan memahami sejauh mana kesalahan prediksi yang mungkin terjadi.

	Predicted Negative	Predicted Positive
Actual Negative	668	0
Actual Positive	49	0

Gambar 6. Hasil Confusion Matrix

Hasil *confusion matrix* mendapatkan nilai TP=0, TN=668, FP=0, dan FN= 49. Kemudian lakukan perhitungan akurasi (*accuracy*) dan tingkat kesalahan (*error rate*).

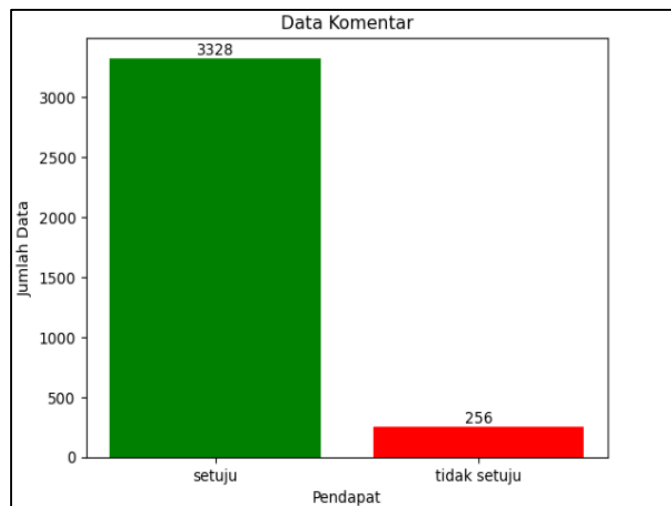
Akurasi: 93.17%
Error Rate: 6.83%

Gambar 7. Hasil Akurasi dan tingkat kesalahan

Dari perhitungannya mendapatkan hasil akurasi (*accuracy*) sebesar 93.17% dan tingkat kesalahan (*error rate*) sebesar 6.83%.

3.5. Hasil Penelitian

Hasil penelitian menunjukkan pola ekspresi positif pada kelompok 'setuju' dan kritik negatif pada kelompok 'tidak setuju', memberikan wawasan tentang perbedaan pendapat dalam komunitas.



Gambar 8. Hasil Penelitian

4. Kesimpulan

Penelitian ini mengaplikasikan teknik Data Mining, khususnya *K-Means Clustering* dan *Naive Bayes Classifier*, untuk mengelompokkan pendapat publik terkait naturalisasi pemain di Timnas Indonesia. Data komentar *YouTube* sebanyak 3584 dianalisis menggunakan *K-Means Clustering*, lalu hasilnya digunakan sebagai label untuk melatih model *Naive Bayes Classifier*. Model ini memiliki akurasi sebesar 93.17% dan tingkat kesalahan 6.83%, memberikan wawasan mendalam tentang pola pendapat masyarakat terhadap isu tersebut. Hasil penelitian dapat menjadi dasar untuk pengambilan keputusan yang informasional dalam merancang kebijakan dan strategi komunikasi responsif terhadap dinamika opini publik.

Penelitian ini membawa kontribusi signifikan dalam pemahaman pandangan masyarakat terhadap naturalisasi pemain di Timnas Indonesia. Melalui penerapan algoritma *K-Means Clustering* dan *Naive Bayes Classifier*, penelitian ini berhasil mengungkap pola ekspresi positif pada kelompok 'setuju' dan kritik negatif pada kelompok 'tidak setuju'. Analisis mendalam terhadap 3584 data komentar *YouTube* memberikan gambaran yang lebih komprehensif tentang perbedaan pendapat dalam komunitas. Hasil penelitian ini tidak hanya menjadi sumber informasi yang berharga untuk pengambilan keputusan, tetapi juga dapat membentuk dasar untuk perancangan kebijakan dan strategi komunikasi yang responsif terhadap dinamika opini publik.

References

- [1] H. Al Rasyid Harpizon et al., 'Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes', *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [2] N. H. Alfianty and S. Mulyati, 'Penerapan Naïve Bayes untuk Klasifikasi Data Penyakit Pada Anak'.
- [3] R. A. Fauzan, '(SENAFTI) 30 Agustus 2023-Jakarta', 2023.
- [4] A. Rahman, F. Rahmat, M. Y. Fariqi, and S. Adi, 'Metode Naive Bayes untuk Menganalisis Akurasi Sentimen Komentar di Youtube', 2020. [Online]. Available: <http://bit.ly/2u802Pe>
- [5] M. A. Akbar, F. Fatimah, and J. Jaenudin, 'Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering'.
- [6] D. Septiani and I. Isabela, 'SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS'.
- [7] M. A. Akbar, F. Fatimah, and J. Jaenudin, 'Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering'.
- [8] N. Silalahi and Guidio Leonarde Ginting, 'Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF', *Bulletin of Computer Science Research*, vol. 3, no. 4, pp. 276–282, Jun. 2023, doi: 10.47065/bulletincsr.v3i4.266.
- [9] P. Y. Saputra, D. H. Subhi, F. Zain, and A. Winatama, 'IMPLEMENTASI SENTIMEN ANALISIS KOMENTAR CHANNEL VIDEO PELAYANAN PEMERINTAH DI YOUTUBE MENGGUNAKAN ALGORITMA NAÏVE BAYES'.
- [10] F. Syofiani, S. Alam, and M. I. Sulistyono, 'Analisis Sentimen Penilaian Masyarakat Terhadap Childfree Berdasarkan Komentar di Youtube Menggunakan Algoritma Naïve Bayes', *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 2, pp. 688–703, Sep. 2023, doi: 10.37012/jtik.v9i2.1661.