

Opini Public Terhadap Berita Ijazah Presiden RI Ke-7 Menggunakan Random Forest Dan Naive Bayes

**Immanuel Bunawan^{1)*}, Peter Raynard Susanto²⁾, Hafidz Irsyad³⁾,
Abdul Rahmad⁴⁾**

^{1,2,3)}Prodi Informatika, Fakultas Ilmu Komputer dan Rekayasa,
Universitas Multi Data Palembang, Indonesia

⁴⁾Prodi Teknik Elektro, Fakultas Ilmu Komputer dan Rekayasa,
Universitas Multi Data Palembang, Indonesia

*Coresponding Email: ¹immanuelbuanwan_2226250040@mhs.mdp.ac.id,

²peterreynardsusanto_2226250020@mhs.mdp.ac.id, ³hafizirsyad@mdp.ac.id, ⁴arahman@mdp.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisa sentimen publik terhadap berita yang sedang tren saat ini di YouTube yaitu mengenai Ijazah Presiden ke-7 Bpk Joko Widodo. Kami menggunakan dua metode dalam pengklasifikasian ini yaitu Random Forest dan Naive Bayes agar dapat membandingkan metode mana yang lebih efektif dan dapat memberikan hasil akurasi yang lebih tinggi. Pada Metode Random Forest, hasil pengklasifikasian dari masing-masing pohon digabungkan untuk melihat sentimen mana yang lebih banyak dihasilkan apakah positif, negatif atau netral. Hasil decision tree yang diambil adalah sentimen dengan hasil tertinggi. Sementara, Naive Bayes adalah pendekatan yang umum dilakukan dalam pengklasifikasian analisis sentimen karena metode ini sederhana dan efisien. Data yang telah diproses sebelumnya kemudian dilatih dan hasil kinerjanya dievaluasi menggunakan confusion matrix. Analisis sentimen pada penelitian ini menggunakan empat berita yang ada di YouTube dengan tema yang sama yaitu mengenai Ijazah Presiden ke-7. Penelitian ini juga menerapkan visualisasi dengan SMOTE pada kedua metode. Hasil penelitian menunjukan bahwa metode Random Forest menghasilkan akurasi yang lebih tinggi yaitu 99% dibandingkan metode Naive Bayes yang hanya menghasilkan akurasi 93%. Pada penelitian ini dapat mengimplementasikan Naive Bayes dan Random Forest dengan menggunakan SMOTE untuk melakukan imbalance data dengan baik.

Kata Kunci: Ijazah, Naive Bayes, Random Forest, Sentimen, SMOTE, YouTube.

Abstract

This study aims to analyze public sentiment towards the current trending news on YouTube, namely regarding the 7th Presidential Diploma of Mr. Joko Widodo. We use two methods in this classification, namely Random Forest and Naive Bayes, to compare which method is more effective and can provide higher accuracy results. In the random forest method, the classification results from each tree are combined to see which sentiment is produced more, whether positive, negative, or neutral. The decision tree results taken are the sentiments with the highest results. While Naive Bayes is a common approach in sentiment analysis classification because this method is simple and efficient. The data that has been previously processed is then trained, and its performance results are evaluated using a confusion matrix. The sentiment analysis in this study uses four news items on YouTube with the same theme, namely regarding the 7th Presidential Diploma. This study also applies visualization with SMOTE to both methods. The results of the study show that the random forest method produces a higher accuracy

of 99% compared to the Naive Bayes method, which only produces an accuracy of 93%. In this study, Naive Bayes and Random Forest can be implemented using SMOTE to assess data imbalance properly.

Keywords: Diploma, Naive Bayes, Random Forest, Sentiment, SMOTE, YouTube.

PENDAHULUAN

Pada saat ini perkembangan teknologi sedang berkembang dengan sangat cepat yang telah melakukan perubahan terhadap gaya hidup masyarakat. Pada beberapa tahun sebelumnya untuk mendapatkan berita, masyarakat biasanya menggunakan media cetak seperti koran, dimana melalui media ini berita yang disampaikan satu atau dua hari setelah kejadian, sedangkan pada zaman sekarang dengan perkembang teknologi yang sangat pesat memungkinkan untuk mendapatkan berita dari berbagai penjuru dunia dalam waktu yang singkat. Untuk mendapatkan dan membaca berita dapat dilakukan melalui media baca berita seperti melalui website berita, artikel dan masih banyak lagi, serta berita juga bisa didapatkan melalui media sosial seperti Youtube, TikTok dan lain-lain.

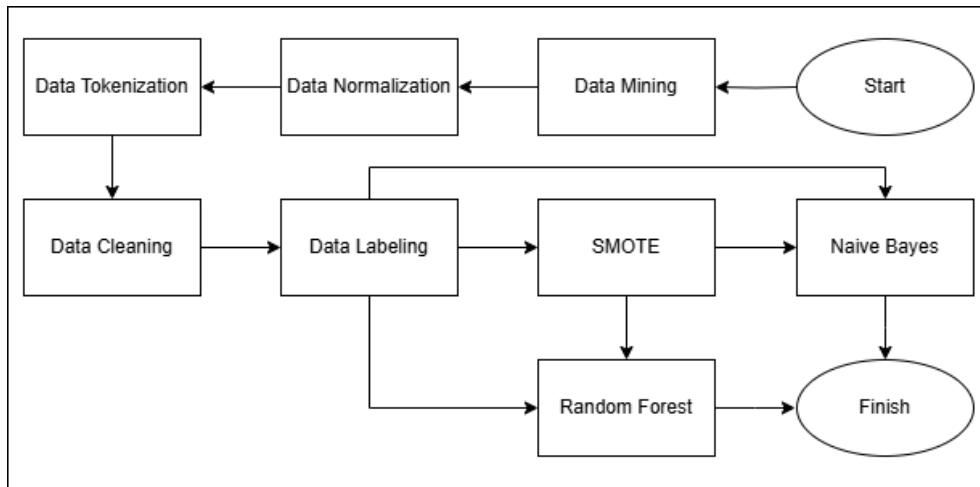
Youtube merupakan platform media sosial yang besar yang memberikan layanan berupa sebuah web yang dapat digunakan untuk mengupload, mendownload, dan juga menonton serta memberikan komentar terhadap video secara bebas. Youtube juga merupakan platform untuk berbagai video hiburan, akan tetapi memberikan video edukasi, sosial, dan juga berita yang terjadi. Youtube sudah digunakan banyak kalangan dengan berbagai tujuan, mulai dari hiburan, edukasi sampai berita (Hariyanto & Putra, 2022; Citra, Suroso, & Hermadi, 2023; Curran et al., 2020; Utomo & Baskoro, 2023). Dengan adanya kebebasan penonton YouTube untuk memberikan komentar, platform ini cocok digunakan untuk mengumpulkan sentimen. Sentimen adalah emosi penonton terhadap video yang ditonton dan dapat digunakan untuk menganalisa pendapat mereka atas video tersebut (Ali et al., 2025). Analisis sentimen adalah konsep yang mencakup banyak tugas, seperti ekstraksi sentimen, klasifikasi sentimen, ringkasan opini, analisis ulasan, deteksi sarkasme atau deteksi emosi, dan lain-lain(Ardiansyah, Alamsyah, & Arif, 2024; Cui et al., 2023; Larasati, Ratnawati, & Hanggara, 2022). Pada penelitian yang dilakukan oleh Putri, Nurhaliza, dan Vionanda (2025), penelitian tersebut menggunakan 995 ulasan YouTube yang diambil dengan bantuan google colab untuk menganalisis mengenai bagaimana pengguna di Indonesia memberikan rating serta umpan balik terhadap pengalaman mereka menggunakan aplikasi YouTube.

Pengkajian ini menggunakan dua metode yaitu Random Forest dan Naive Bayes. Random Forest adalah salah satu algoritma machine learning yang umum digunakan dalam permasalahan klasifikasi, Random Forest memiliki dua fungsi yaitu klasifikasi dan prediksi di mana hasil setiap decision tree belum tentu sama. Oleh karena itu hasil keputusan akhir diambil dari suara terbanyak dari decision tree (Kurniawan et al., 2024). Semakin banyak hasil decision tree yang diambil, semakin tinggi akurasi terutama ketika masing-masing pohon tidak memberikan hasil yang sama. Pada penelitian yang telah dilakukan menggunakan algoritma Random Forest berhasil mencapai akurasi sebesar 80% dalam mengklasifikasikan tweet tentang pinjaman online di Indonesia (Ardiansyah et al., 2024). Sementara metode Naive Bayes adalah algoritma klasifikasi probabilistik yang cepat dan sederhana (Hidayat et al., 2024). Algoritma ini digunakan untuk mengklasifikasikan data kedalam beberapa kelas seperti dalam memprediksi apakah sebuah komentar termasuk ke dalam sentimen positif, negatif, atau netral. Pada penelitian terdahulu yang menggunakan metode Naive Bayes berhasil mencapai akurasi tertinggi sebesar 88,89% dari dataset tidak seimbang dalam mengklasifikasikan sentimen publik di Twitter terhadap Puan Maharani (Hidayat et al., 2024), sedangkan pada penelitian lain menunjukkan bahwa Naive Bayes mampu menghasilkan akurasi sebesar 92% dalam melakukan analisis sentimen pada ulasan yang ada di aplikasi e-commerce seperti Shopee dan Tokopedia (Ramadhan, Adam, & Maulana, 2022).

Adapun tujuan dari penelitian ini adalah untuk membandingkan antara algoritma Naive Bayes dan Random Forest dengan menggunakan SMOOTH dan tanpa menggunakan SMOTE untuk memberikan kontribusi hasil perbandingan algoritma Naive Bayes dan Random Forest dengan optimasi SMOTE dan tanpa optimasi SMOTE pada analisis sentimen opini publik pada media Youtube.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menggunakan dua algoritma klasifikasi yaitu Random Forest dan Naive Bayes, serta data yang digunakan dibagi menjadi dua, menggunakan SMOTE dan tanpa menggunakan SMOTE. Terdapat alur kerja yang dapat dilihat pada Gambar 1.

**Gambar 1.** Alur Kerja

Adapun beberapa tahapan yang dilakukan seperti:

1. Pengumpulan Data

Kumpulan data yang digunakan dalam kajian ini adalah komentar pengguna yang menonton berita pada aplikasi YouTube yang diambil dengan bantuan *Google Colab* yang berjumlah empat berita dengan total 3952 komentar. Adapun berita yang kami ambil bertema mengenai ijazah palsu dengan judul berita sebagai berikut:

1. Merasa Dihina Jokowi Laporkan 5 Orang soal Ijazah Palsu, Bentuk Pembelajaran atau Kriminalisasi?
2. Roy Suryo, Kuasa Hukum Jokowi, Eks Ketua KPU & Prof Komaruddin Bicara Geger Ijazah UGM Jokowi
3. Teman Kuliah Jokowi Heran Banyak Orang "Termakan" Isu Ijazah Palsu
4. Jokowi & Teman Kuliah UGM Tertawa saat Disinggung Soal Ijazah Palsu (Arsip Kompas TV)

2. Preprocessing Data

Pada tahap ini dilakukan beberapa tahap agar teks yang telah dikumpulkan dapat dianalisis. Tahapan-tahapan yang dilakukan meliputi:

1. Case Folding

Teks yang telah dikumpulkan dilakukan *case folding* agar semua teks menggunakan huruf kecil. Hal ini dilakukan agar penggunaan huruf selalu konsisten.

2. Tokenisasi

Setelah *case folding*, tahap selanjutnya adalah tokenisasi. Tokenisasi dilakukan untuk mengetahui banyak kata. Tokenisasi dilakukan dengan memecah suatu teks menjadi kata, frasa, simbol atau elemen lain yang mempunyai makna.

3. Stopword Removal

Selanjutnya dilakukan penghilangan *stopword*. *Stopword* adalah kata-kata yang tidak memiliki makna penting.

4. Filtering

Pada tahap ini, simbol-simbol atau karakter khusus di hilangkan.

5. Stemming

Kata-kata yang tersisa dari proses *stopword removal* dilakukan *stemming*. *Stemming* adalah penyederhanaan kata-kata yang memiliki imbuhan menjadi kata dasar.

3. SMOTE

Synthetic Minority Oversampling Technique atau yang biasa disebut SMOTE merupakan teknik yang biasa digunakan untuk menyeimbangkan kelas data yang tidak seimbang, dalam hal ini sentimen positif, negatif atau netral. Jumlah data yang tidak seimbang dapat menyebabkan model yang dikembangkan tidak optimal dan kurang akurat karena model lebih cenderung kepada kelas data yang mayoritas, sehingga tidak dapat memprediksi dengan baik di kelompok data minoritas. SMOTE mampu menghasilkan data tambahan pada kelompok minoritas sehingga kedua grup memiliki jumlah data yang seimbang tanpa menambah jumlah data mayoritas. Hal ini membantu model yang dikembangkan untuk belajar secara adil sehingga dapat menghindari overfitting. SMOTE mengidentifikasi tetangga terdekat dari setiap data yang dimasukkan ke dalam kelompok yang lebih sedikit kemudian, SMOTE membuat data baru di antara data kelompok yang jumlahnya lebih sedikit dan tetangga terdekat yang dipilih *random* (Chawla et al., 2002).

4. Klasifikasi dengan Naive Bayes

Algoritma Naive Bayes adalah metode yang sering digunakan dalam sentimen analisis. Dari hasil ekstraksi fitur menggunakan SMOTE akan dilakukan klasifikasi sentimen menggunakan Naive Bayes. Algoritma ini bekerja dengan mengklasifikasi teks berdasarkan probabilitas kelas sentimen (positif, negatif atau netral) (Hidayatullah, Purwantoro, & Umaidah, 2023) melalui persamaan (1).

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

dimana:

C = kelas sentimen (positif, negatif atau netral)

X = fitur teks (Vektor TF-IDF)

$P(C|X)$ = Probabilitas bahwa teks X termasuk ke dalam kelas C (positif, negatif atau netral)

$P(X|C)$ = Probabilitas munculnya fitur teks X jika diketahui bahwa kelasnya adalah C yang dihitung berdasarkan distribusi kata dalam masing-masing kelas

$P(X)$ = Probabilitas fitur teks X muncul dalam semua kelas

$P(C)$ = Probabilitas kelas tersebut muncul dalam data

Kelas dengan probabilitas tertinggi $P(C|X)$ akan dipilih sebagai klasifikasinya (positif, negatif atau netral).

5. Klasifikasi dengan Random Forest

Random Forest adalah sebuah model *machine learning* yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini merupakan bentuk *ensemble learning*, yang berarti menggabungkan beberapa model prediktif yang lebih sederhana untuk mencapai prediksi yang lebih akurat dan stabil (Mahmuda, 2024).

Membentuk pohon keputusan pada metode Random Forest sama dengan proses pada *Classification and Regression Tree* (CART), hanya saja pada Random Forest tidak dilakukan *pruning* (pemangkasan). Indeks Gain digunakan untuk memilih fitur di setiap simpul internal dari pohon keputusan. Untuk mendapatkan nilai Indeks Gain dapat dicari melalui (2), (3), dan (4).

$$\text{Indeks Gain} = \text{Entropi}(D) - \text{entropi}_A(D) \quad (2)$$

Persamaan (2) digunakan dalam Information Gain (IG) untuk mengukur seberapa banyak ketidakpastian (entropi) berkurang setelah data dipisah berdasarkan suatu atribut A. Nilai IG yang tinggi berarti atribut tersebut membagi data dengan baik, dan itu yang dipilih oleh algoritma decision tree.

$$\text{Entropi}(D) = - \sum_{i=1}^c p_i^2 \log(p_i) \quad (3)$$

dimana:

: jumlah kelas target (variabel prediktor)

p_i : proporsi kelas ke-i partisi D

Persamaan (3) digunakan Entropi (D) dalam mengukur tingkat ketidakpastian dalam partisi D. Jika semua data berada dalam satu kelas, maka entropi = 0 (tidak ada ketidakpastian). Dan jika data terbagi rata ke semua kelas, entropi maksimum (paling tidak pasti).

$$\text{Entropi}_A(D) = \sum_{j=1}^v \frac{n_j}{n} \times \text{entropi}(D_j) \quad (4)$$

dimana:

: jumlah partisi (berapa banyak subset setelah pemisahan data)

: jumlah observasi

n_j : jumlah observasi j

D_j : partisi ke-j

Persamaan (4) digunakan dalam $Entropi_A D$ merupakan ketidakpastian setelah pemisahan berdasarkan atribut A.

$\frac{n_j}{n}$ adalah proporsi data di partisi ke-j, digunakan untuk pembobotan.

Proses yang dilakukan setiap iterasi yang ada di dalam algoritma pohon klasifikasi pada dasarnya digunakan untuk mencari metode partisi yang dapat memberikan informasi gain tertinggi.

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Pengumpulan data adalah proses peneliti dalam mengumpulkan data yang diperlukan untuk menjalankan pemisahan data yang telah dibuat di dalam google colab. Penulis melakukan pengambilan data secara langsung ke beberapa *channel* berita pada laman YouTube. Data yang digunakan berupa komentar pada video yang digunakan dalam kasus ini dengan judul "Ijazah Presiden RI Ke-7".

2. Tokenisasi Data

Tokenisasi Data merupakan proses dalam memecah kalimat menjadi beberapa bagian agar dapat terbaca menggunakan *coding*. Contoh tokenisasi data yang telah dilakukan dapat dilihat pada Tabel 1.

Tabel 1. Tabel Hasil Tokenisasi data

No	Support
1	[habisi, menuding, bapak]
2	[kamu, nda, mau, dihina, tunjukan, ijazamu, mul]
3	[pastas, hina, karna, pembohong]
4	[berhasil, ngibul, nyaa, wowo, haha]
5	[ternyata, kamu, ternyata, wartidak, rw, 024]

3. Pelabelan Komentar

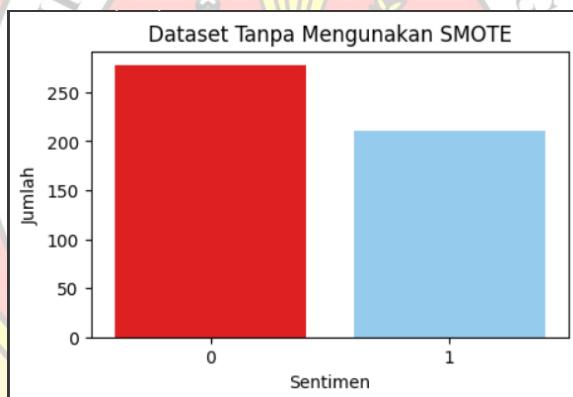
Pada proses ini dilakukan label data berupa angka 1 untuk komentar positif, angka 0 untuk komentar negatif, dan angka -1 untuk komentar yang bersifat netral. Contoh pelabelan komentar dapat dilihat pada Tabel 2.

Tabel 2. Tabel Hasil Pelabelan data

No	Support	Label
1	pastas hina karena bohong	0
2	kalian fitnah ijazahnya palsu lihat asli	1
3	jelas faktor benci	0
4	ijazah gelar bakatai downkamuad sah	1
5	mantap pk jokowi harus sabar hadap orang goblok	1

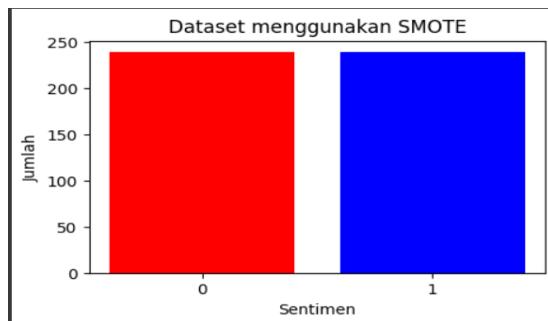
4. Dataset menggunakan SMOTE dan Tanpa SMOTE

Pada tahap ini dilakukan proses untuk menampilkan dataset dalam bentuk Gambar. Dataset yang dibuat tanpa menggunakan metode SMOTE dapat dilihat pada Gambar 2.



Gambar 2. Dataset tanpa menggunakan SMOTE

Sedangkan Pada Gambar 3 telah dilakukan *balancing* dataset menggunakan metode SMOTE. Terlihat perbedaan antara Gambar 2 dan Gambar 3 yaitu dari jumlah data yang ditampilkan.

**Gambar 3.** Dataset menggunakan SMOTE

Dari Gambar 2 di atas, dapat dilihat bahwa apabila tanpa menggunakan SMOTE, sentimen negatif mempunyai data yang lebih banyak dari sentimen positif. Tetapi setelah menggunakan SMOTE, seperti pada Gambar 3, dapat dilihat bahwa jumlah data sentimen positif dan sentimen negatif sama jumlahnya.

5. Algoritma Naive Bayes

Pada proses ini data yang telah disiapkan dilakukan *training* menggunakan Algoritma Naive Bayes yang telah disiapkan tanpa menggunakan SMOTE dan mendapatkan *Precision* sebesar 94%, *Recall* sebesar 93%, *F1-score* sebesar 93% serta akurasi sebesar 93%, sedangkan setelah dilakukan *training* dengan menggunakan SMOTE mendapatkan *Precision* sebesar 94%, *Recall* sebesar 93%, *F1-score* sebesar 93% serta akurasi sebesar 93%. Data-data lain seperti *Precision*, *Recall* dan *F1-score* dapat dilihat pada Tabel 3.

Tabel 3. Tabel Hasil Naive Bayes

SMOTE		Precision	Recall	F1-score	Support
YA	Positif	1.00	0.86	0.92	35
	Negatif	0.89	1.00	0.94	39
	Accuracy			0.93	74
	Macro avg	0.94	0.93	0.93	74
	Weight avg	0.94	0.93	0.93	74
Tidak	Positif	1.00	0.86	0.92	35
	Negatif	0.89	1.00	0.94	39
	Accuracy			0.93	74
	Macro avg	0.94	0.93	0.93	74
	Weight avg	0.94	0.93	0.93	74

6. Algoritma Random Forest

Pada proses ini data yang telah disiapkan dilakukan *training* terhadap model yang sudah dibuat pada algoritma Random Forest tanpa menggunakan SMOTE yang mendapatkan hasil *Precision* sebesar 96%, *Recall* sebesar 96%, *F1-score* sebesar 96% serta akurasi sebesar 96%, sedangkan setelah dilakukan *training* dengan bantuan SMOTE *Precision* sebesar 99%, *Recall* sebesar 99%, *F1-score* sebesar 99% serta akurasi sebesar 99%. Data lain seperti *Precision*, *Recall* dan *F1-score* dapat dilihat pada Tabel 4.

Tabel 4. Tabel Hasil Random Forest

SMOTE		Precision	Recall	F1-score	Support
Ya	Positif	1.00	0.97	0.99	35
	Negatif	0.97	1.00	0.99	39
	Accuracy			0.99	74
	Macro avg	0.99	0.99	0.99	74
	Weight avg	0.99	0.99	0.99	74
Tidak	Positif	1,00	0.91	0.96	35
	Negatif	0.93	1.00	0.96	39
	Accuracy			0.96	74
	Macro avg	0.96	0.96	0.96	74
	Weight avg	0.96	0.96	0.96	74

7. Hasil Perbandingan

Tabel 5 menunjukkan perbandingan hasil dari dua metode klasifikasi dengan menggunakan Naive Bayes dan Random Forest, dengan menggunakan dan tanpa menggunakan SMOTE. Terdapat akurasi sebesar 93%, *Precision* 0.94, *Recall* 0.93 serta *F1-score* 0.93 pada Algoritma Naive Bayes yang menggunakan SMOTE sedangkan Terdapat akurasi sebesar 93%, *Precision* 0.94, *Recall* 0.93 serta *F1-score* 0.93 pada Algoritma Naive Bayes tanpa menggunakan SMOTE, sedangkan terdapat akurasi sebesar 99%, *Precision* 0.99, *Recall* 0.99 serta *F1-score* 0.99 pada Algoritma Random Forest yang menggunakan SMOTE serta Terdapat akurasi sebesar 96%, *Precision* 0.96, *Recall* 0.96 serta *F1-score* 0.96 pada Algoritma Random Forest tanpa menggunakan SMOTE.

Tabel 5. Tabel Perbandingan Hasil

Metode	SMOTE	Accuracy	Precision (avg)	Recall (avg)	F1-score (avg)
Naive Bayes	Ya	93%	0.94	0.93	0.93
	Tidak	93%	0.94	0.93	0.93
Random Forest	Ya	99%	0.99	0.99	0.99
	Tidak	96%	0.96	0.96	0.96

SIMPULAN

Hasil dari penelitian ini menunjukkan perbedaan hasil akurasi dari kedua algoritma tersebut menunjukkan bahwa Algoritma Random Forest menghasilkan akurasi yang lebih bagus dibandingkan Algoritma Naive Bayes yang digunakan data yang telah diproses menggunakan SMOTE maupun data yang belum diproses menggunakan SMOTE sehingga mendapatkan hasil berupa 99% akurasi pada Algoritma Random Forest sedangkan pada Algoritma Naive Bayes mendapatkan akurasi sebesar 93% dengan kedua menggunakan SMOTE sedangkan pada pemrosesan tanpa menggunakan SMOTE mendapatkan akurasi sebesar 96% pada Algoritma Random Forest dan 93% pada Algoritma Naive Bayes.

DAFTAR PUSTAKA

- Ali, H. M. U., Farooq, Q., Imran, A., & el Hindi, K. (2025). A systematic literature review on sentiment analysis techniques, challenges, and future trends. *Knowledge and Information Systems*, 67(5), 3967–4034. <https://doi.org/10.1007/s10115-025-02365-x>
- Ardiansyah, M. A., Alamsyah, M., & Arif, M. F. (2024). Analisis sentimen Twitter tentang pinjaman online di Indonesia menggunakan metode Random Forest. *JUTISI: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 13(2), 1410. <https://doi.org/10.35889/jutisi.v13i2.2215>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Citra, S. O., Suroso, A. I., & Hermadi, I. (2023). Model Utaut 2: Penerimaan dan penggunaan YouTube di Indonesia. *Techno.Com*, 22(4), 1026–1036. <https://doi.org/10.33633/tc.v22i4.8817>

- Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: Evolution of research methods and topics. *Artificial Intelligence Review*, 56(8), 8469–8510. <https://doi.org/10.1007/s10462-022-10386-z>
- Curran, V., et al. (2020). YouTube as an educational resource in medical education: A scoping review. *Medical Science Educator*, 30(4), 1775–1782. <https://doi.org/10.1007/s40670-020-01016-w>
- Hariyanto, A., & Putra, A. (2022). Konten kreator YouTube sebagai sumber penghasilan (Telaah kritis hukum ekonomi syari'ah). *Al-Hukmi: Jurnal Hukum Ekonomi Syariah dan Keluarga Islam*, 3(2), 243–262. <https://doi.org/10.35316/alhukmi.v3i2.2325>
- Hidayat, R., Fikry, M., Yusra, Y., Yanto, F., & Cynthia, E. P. (2024). Penerapan Naïve Bayes Classifier dalam klasifikasi sentimen publik di Twitter terhadap Puan Maharani. *JUKI: Jurnal Komputer dan Informatika*, 6(1), 100–108. <https://doi.org/10.53842/juki.v6i1.479>
- Hidayatullah, H., Purwantoro, P., & Umaidah, Y. (2023). Penerapan Naïve Bayes dengan optimasi information gain dan SMOTE untuk analisis sentimen pengguna aplikasi ChatGPT. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1546–1553. <https://doi.org/10.36040/jati.v7i3.6887>
- Kurniawan, Y., Halim, E., Jennifer, E., Pribadi, F. A., Bhutkar, G., & Anwar, N. (2024). Understanding user engagement strategies for podcasts videos on YouTube in Indonesia: A study on content creation. *JOIV: International Journal on Informatics Visualization*, 8(2), 957. <https://doi.org/10.62527/joiv.8.2.2123>
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis sentimen ulasan aplikasi Dana dengan metode Random Forest. *J-PTIIK*, 6(9), 4305–4313.
- Mahmuda, S. (2024). Implementasi metode Random Forest pada kategori konten kanal YouTube. *Jurnal Jendela Matematika*, 2(1), 21–31. <https://doi.org/10.57008/jjm.v2i01.633>
- Putri, T., Nurhaliza, S., & Vionanda, D. (2025). Analisis sentimen penggunaan aplikasi YouTube menggunakan metode Naïve Bayes. *UNP Journal of Statistics and Data Science*, 3(1), 60–66. <https://doi.org/10.24036/ujsds/vol3-iss1/343>
- Ramadhan, B. Z., Adam, R. I., & Maulana, I. (2022). Analisis sentimen ulasan pada aplikasi e-commerce dengan menggunakan algoritma Naïve Bayes. *Journal of Applied Informatics and Computing*, 6(2), 220–225. <https://doi.org/10.30871/jaic.v6i2.4725>
- Utomo, F. A. P., & Baskoro, D. N. (2023). Pengaruh konten vlog di YouTube terhadap perilaku komunikasi siswa SMP Nusantara Plus. *Communicator Sphere*, 3(1), 13–31. <https://doi.org/10.55397/cps.v3i1.34>