

Penerapan Naive Bayes, Chi-Square dan SMOTE pada Opini Masyarakat Terhadap Fufufafa di YouTube

Lucky Tandoballa^{*1}, Andreyas², Novan Wijaya³

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

Email: ^{*1}lucky.tandoballa@mdp.ac.id, ²andreyas_2226250025@mhs.mdp.ac.id, ³novan.wijaya@mdp.ac.id

(Naskah masuk: 19 Juni 2025, diterima untuk diterbitkan: 31 Juli 2025)

Abstrak: Penelitian ini bertujuan untuk menganalisis opini masyarakat terhadap komentar terkait akun kaskus fufufafa pada platform YouTube dengan menggunakan metode Naive Bayes, Chi-square, dan SMOTE. Dalam penelitian ini meliputi beberapa metode tahapan utama: mining dataset, pelabelan dataset, preprocessing, dan penerapan metode SMOTE untuk mengatasi ketidakseimbangan pada kelas. Penambangan data dilakukan dengan cara mengumpulkan data komentar pengguna dari video YouTube terkait pembahasan akun kaskus fufufafa. Kemudian dilakukan langkah pelabelan untuk mengklasifikasikan komentar menjadi sentimen positif, negatif, atau netral. Tahap preprocessing meliputi pembersihan data dari unsur-unsur yang tidak diperlukan seperti tanda baca, angka, dan karakter khusus. Untuk mengatasi masalah ketidakseimbangan pada kelas, Kami kemudian menerapkan Synthetic Minority Oversampling Technique (SMOTE) dimana jumlah komentar dengan sentimen tertentu lebih sedikit dibandingkan jumlah komentar yang lain. Hasil penelitian ini menunjukkan bahwa akurasi model Naive Bayes mencapai 60,5%, sedangkan penggunaan seleksi fitur chi-square dengan SMOTE meningkatkan akurasi menjadi 68,3%. Dalam hal ini menunjukkan bahwa penggunaan chi-square dengan SMOTE dapat meningkatkan akurasi prediksi sentimen sebesar 7,8%. Kesimpulan dari penelitian ini adalah model Naive Bayes dengan pemilihan fitur chi-square dengan SMOTE lebih efektif dalam memprediksi opini masyarakat dibandingkan model Naive Bayes tanpa pemilihan fitur tersebut.

Kata Kunci – Opini Masyarakat, Chi-Square, Naive Bayes, SMOTE

Abstract: This study aims to analyze the sentiment of comments related to the kaskus fufufafa account on the YouTube platform using the Naive Bayes, Chi-square, and SMOTE methods. The research involves several main stages, including data mining, dataset labeling, preprocessing, and the application of the SMOTE method to address class imbalance. Data collection was carried out by gathering user comments from YouTube videos related to the targeted account. The labeling stage then classified the comments into positive, negative, or neutral sentiments. In the preprocessing stage, the data was cleaned by removing unnecessary elements such as punctuation, numbers, and special characters. To solve the problem of class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, especially when the number of comments in certain sentiment classes was lower than others. The results of this study show that the accuracy of the Naive Bayes model without feature selection reached 60.5%. However, after applying Chi-square feature selection combined with SMOTE, the accuracy increased to 68.3%. This indicates that using Chi-square and SMOTE can improve sentiment prediction accuracy by 7.8%. In conclusion, the Naive Bayes model with Chi-square feature selection and SMOTE is more effective in predicting user comment sentiments than the basic model.

Keywords – Chi-Square, Naive Bayes, Public Opinion, SMOTE

1. PENDAHULUAN

Media sosial telah berkembang menjadi platform penting bagi masyarakat untuk mengungkapkan pendapat dan perasaan mereka tentang berbagai masalah. Sebagai platform berbagi video terbesar, YouTube memungkinkan pengguna untuk meninggalkan komentar yang menunjukkan pendapat mereka tentang konten yang ditampilkan. Analisis komentar-komentar ini dapat memberikan informasi penting tentang persepsi publik tentang topik tertentu.

Namun, masalah utama dengan opini masyarakat adalah ketidakseimbangan kelas dalam data; ini terjadi ketika komentar dengan sentimen tertentu, seperti negatif, lebih banyak daripada komentar dengan sentimen lainnya, seperti positif. Karena model klasifikasi cenderung bias terhadap kelas mayoritas, ketidakseimbangan ini dapat mempengaruhi kinerja model. Untuk menyelesaikan masalah ini, sampel sintetis dari kelas minoritas dibuat dengan teknik *over-sampling* kelas minoritas sintetis (SMOTE). Dengan demikian, distribusi data menjadi lebih seimbang. [1].

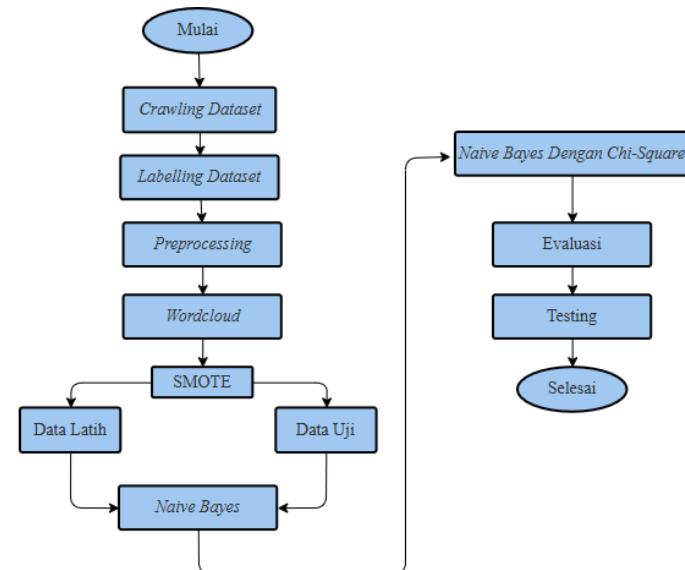
Algoritma *Naive Bayes* digunakan dalam klasifikasi teks karena mudah digunakan dan efektif dalam menangani data berukuran besar. Algoritma ini didasarkan pada prinsip probabilistik dan mengambil independensi antar fitur, yang memungkinkan proses pelatihan model yang cepat dan akurat [2]. *Naive Bayes* adalah salah satu algoritma yang paling efektif untuk menganalisis data berskala besar [3]. Namun, ketika digunakan pada data yang tidak seimbang, kinerjanya dapat menurun, sehingga untuk meningkatkan akurasi klasifikasi, algoritma ini harus dikombinasikan dengan teknik seperti SMOTE.

Beberapa penelitian di Indonesia telah menggabungkan *Naive Bayes* dan SMOTE dalam opini masyarakat. Misalnya, Kurnia et al. [1] melakukan analisis opini masyarakat terhadap komentar Twitter mengenai Bank BTN dengan menggabungkan *Naive Bayes*, SMOTE, dan Adaboost, yang menghasilkan akurasi sebesar 87,05%. Penelitian oleh Ramanda et al. [2] menganalisis opini masyarakat terhadap perilaku lawan arah yang diunggah di YouTube, dan menemukan bahwa penerapan SMOTE meningkatkan akurasi model dari 76,5% menjadi 78%. Dan penelitian lain oleh Hidayatullah et al. [4] menganalisis opini masyarakat terhadap pengguna aplikasi Chatgpt dengan pendekatan *machine learning* menggunakan algoritma *Naive Bayes* dengan optimasi *Information Gain* dan SMOTE meningkatkan akurasi model dari 76,12% menjadi 87,20%.

Penelitian ini bertujuan untuk mengembangkan model analisis sentimen yang menggabungkan algoritma *Naive Bayes* dan metode SMOTE untuk meningkatkan klasifikasi sentimen pada komentar YouTube. Dengan menggunakan metode seleksi fitur dengan *chi-square*, diharapkan tingkat kesalahan yang disebabkan oleh terlalu banyak atribut akan lebih rendah dan jumlah atribut yang dimasukkan dalam analisis akan lebih sedikit [5]. Diharapkan bahwa kombinasi ini akan meningkatkan kemampuan model untuk mengklasifikasikan sentimen pengguna dan mengatasi ketidakseimbangan kelas.

2. METODE PENELITIAN

Untuk melakukan klasifikasi sentimen dari data komentar YouTube, penelitian ini menggunakan pendekatan kuantitatif eksperimental yang melibatkan tahapan sistematis. Dataset yang digunakan berasal dari komentar yang ditemukan pada video platform YouTube terkait akun Fufufafa. Algoritma *Naive Bayes* digunakan untuk klasifikasi, teknik *Chi-Square* untuk pemilihan fitur, dan SMOTE untuk menangani ketidakseimbangan kelas digunakan. Gambar 1 menunjukkan tahapan pada penelitian ini.



Gambar 1. Tahapan Penelitian

2.1. Crawling Dataset

Crawling adalah metode untuk mengumpulkan data melalui internet. *Crawling* bekerja secara otomatis, data didasarkan pada kata kunci yang diberikan pengguna. *Crawler* merupakan alat yang digunakan untuk *crawling*. *Crawler* adalah program yang diprogram menggunakan algoritma tertentu sehingga dapat memindai situs web menggunakan alamat URL dan kata kunci dari pengguna [6]. Proses ini bertujuan juga mengekstraksi fitur dan komponen dari teks, serta menentukan apakah komentar tersebut memiliki nada sentimen negatif atau positif [7].

2.2. Pengumpulan Data

Pada tahap ini, data dikumpulkan melalui proses *crawling dataset* menggunakan Google Colab. Semua komentar pada video YouTube yang terkait dengan akun Fufufafa menunjukkan pendapat atau perasaan pengguna tentang konten video tersebut. Total 1336 komentar mentah dikumpulkan dan kemudian disimpan dalam format CSV untuk pemrosesan tambahan [8].

2.3. Pelabelan Sentimen

Pada tahap ini, setiap komentar diberi label sentimen secara manual pada komentar-komentar YouTube yang ada dalam file csv. Tujuan dari label ini adalah untuk memberikan tiga kategori (Positif, Negatif, dan Netral) pada dataset sesuai dengan konteks dan isi komentar tersebut. Pelabelan dilakukan secara manual dengan menggunakan kamus sentimen Bahasa Indonesia yang telah disesuaikan sebagai referensi. [9].

2.4. Preprocessing Data

Tahapan ini dilakukan untuk membersihkan data komentar agar siap dianalisis. Proses yang dilakukan meliputi [9]:

1. *Cleaning & Case Folding*,
2. *Tokenizing*,
3. *Filtering*,
4. *Stemming*.

2.5. Wordcloud

Wordcloud merupakan sebuah metode penambangan teks, menampilkan grafik frekuensi kata yang menunjukkan kata-kata yang paling sering ditemukan dalam teks sumber. Besar gambar menunjukkan frekuensi kata dalam dokumen. [10].

2.6. Penyeimbangan Data dengan SMOTE

Karena data yang tidak cocok atau tidak terstruktur dengan baik, kelas yang tidak seimbang sering terjadi dalam penelitian dengan banyak data [11]. Teknik Pengambilan Sampel Minoritas Sintetis (SMOTE), di sisi lain, meningkatkan keakuratan algoritme yang digunakan dan memungkinkan algoritme untuk menangani kelas yang tidak seimbang [12].

2.7. Proporsi data latih dan data uji

Prinsip Pareto, juga dikenal sebagai aturan 80:20, adalah dasar dari pemilihan persentase untuk membagi data penelitian ini. Pada dasarnya, teori ini mengatakan bahwa 20 persen keluaran atau koordinasi masukan dan keluaran yang tidak seimbang membentuk 80 persen hasil [13].

2.8. Klasifikasi dengan Naive Bayes

Dengan asumsi bahwa setiap elemen atau variabel independen, metode *Naive Bayes* mengklasifikasikan data dengan menggunakan klasifikasi *Bayesian Classification*. Thomas Bayes, seorang ilmuwan Inggris, memperkenalkan teknik ini, yang memungkinkan untuk memprediksi kemungkinan keanggotaan dalam sebuah kelas berdasarkan pengalamannya [14]. *Naive Bayes* berhasil melakukan klasifikasi dengan data pelatihan yang relatif kecil, sementara data uji akan digunakan untuk mengevaluasi keakuratan prediksi

dan keputusan yang dibuat [15]. Rumus atau persamaan Teorema Bayes dapat dihitung menggunakan Persamaan 1 [16].

$$P(N) = \frac{P(M).P(M)}{P(N)} \quad (1)$$

2.9. Seleksi Fitur dengan Chi-Square

Seleksi fitur dilakukan untuk mengurangi dimensi data teks yang tinggi. *Chi-Square* digunakan untuk memilih fitur-fitur yang paling signifikan terhadap kelas target. Fitur dengan nilai *Chi-Square* tertinggi dipertahankan dan digunakan dalam tahap pelatihan model [9]. Uji *Chi-Square* dirumuskan sebagai berikut.

$$X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

2.10. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F-measure* untuk mengukur performa klasifikasi pada tiap kelas sentimen. Hasil evaluasi menunjukkan peningkatan akurasi setelah penerapan *Chi-Square* dan SMOTE dibandingkan dengan penggunaan model *Naive Bayes* standar [17]. Rumus dalam hitung *accuracy*, *precision*, dan *recall* sebagai berikut.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3) \quad \text{Precision} = \frac{TP}{TP+FP} \quad (4) \quad \text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

2.11. Testing

Tahapan testing merupakan percobaan yang dilakukan pada data uji yang bertujuan untuk menentukan kata yang kita masukkan merupakan sentimen positif atau negatif.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data Awal

Dataset awal yang dikumpulkan terdiri dari 1336 komentar pengguna YouTube yang berkaitan dengan akun Fufufafa. Dataset berisi Kumpulan komentar youtube dengan berbagai pendapat tentang akun fufufafa. Hasil dari pengumpulan data dapat dilihat pada table 1.

Tabel 1. Hasil Dataset yang Dikumpulkan

<i>publishedAt</i>	<i>authorDisplayName</i>	<i>textDisplay</i>	<i>likeCount</i>
2025-05-04T00:19:59Z	@ryodtra966	Memang indonesia ini banyak yg pintar untuk memfitnah.. ingat teknologi sudang angat canggih.. bnyak sdra/i yg berkompeten d dunia informatika.. knp dtk gunakan mereka trus buka takbir ketidakpastian ini.. jgn hny bicara pintar tp sesungguhnya menutupi kebodohan sendiri.. Ingat Indonesia ini Negara bukan Kelompok!!!! kekuasaan tertinggi itu ada d tangan rakyat, jd apa yg sudah d pilih oleh kami masyarakat jd kalian kelompok dg kepentingan kalian mulai plintir. Tolong hargai kami masyarakat kecil ini..	0
2025-05-03T03:44:46Z	@pitrasyapitra4138	Masih anak anak ..belum diajarkan adab	0
2025-04-29T06:01:46Z	@hadipurwanto8904	Yg membuat akun fufu fafa akan ketahuan, Gibran bukan pemiliknya, gibran tau etika sopan satun, diam berarti mengakui, karena tak ada gunanya membahas fufu fafa.	1

2025-05-07T06:28:59Z	@muhamadarif9854	Didepan media akan tampak sopan santun lemah lembut plonga plongo, tapi gaada yg tau pribadi seseorang saat berselancar di dunia maya	0
2025-03-27T12:23:12Z	@Redmi13c-e8p	Kan sdh tahu gibran produk nepotisme produk paman usman.. Tp msh aja dipilih.. Jujur saya heran.. Otak yg mana yg digunakan berpikir.. Otak kanan. Otak kiri atau otak terbelakang.. Pak prabowo tetap kita hormati, krn jd capres dg cara yg benar.. Lhi si gibran jd cawapres lewat jalur MK.. Bg saya prabowo oke.. Dan wajib kita jaga.. Jika pak prabowo jatuh . Maka yg naik adalah gibran.. Mau minta gibran mundur.. Sampai bu iriana matipun gibran gak akan mundur.. Kita keras pada pak prabowo agar beliau tidak jatuh..	1

Untuk digunakan sebagai perbandingan selama pengujian, pelabelan data secara manual antara positif, negatif, dan netral. Setelah dikumpulkan didapatkan 263 ulasan negatif, 340 ulasan positif, dan 628 ulasan netral didistribusikan menurut kelas. Karena algoritma cenderung lebih "berpihak" pada kelas mayoritas, ketimpangan distribusi ini dapat menyebabkan bias pada model klasifikasi.

3.2. Preprocessing

Preprocessing dimulai dengan *cleaning*, *case folding*, menghapus kolom yang tidak diperlukan, normalisasi teks, *tokenizing*, *filtering*, *stemming*, *cleaning stop-words*, penggabungan data, label *encoding*. Hasil dari *preprocessing* dapat dilihat pada table 2.

Tabel 2. Hasil Preprocessing

authorDisplayName	textDisplay	Sentimen
@ryodtra966	memang indonesia ini banyak yang pintar untuk fitnah ingat teknologi sudah sangat canggih banyak saudara saudari yang kompeten di dunia informatika kenapa tidak guna mereka terus buka takbir tidak pasti ini jangan hanya bicara pintar tapi sungguh tutup bodoh sendiri ingat indonesia ini negara bukan kelompok kuasa tinggi itu ada di tangan rakyat jadi apa yang sudah di pilih oleh kami masyarakat juga kalian kelompok dengan penting kalian mulai pelintir tolong harga kami masyarakat kecil ini	Negatif
@pitrasyapitra4138	masih anak anak belum ajar adab	Netral
@hadipurwanto8904	yang buat akun fufu fafa akan tahu gibran bukan milik gibran tahu etika sopan satun diam arti akui karena tidak ada guna bahas fufu fafa	Netral
@muhamadarif9854	depan media akan tampak sopan santun lemah lembut tercengang tapi tidak ada yang tau pribadi orang saat selancar di dunia maya	Positif
@Redmi13c-e8p	kan sdh tahu gibran produk nepotisme produk paman usman tapi msh aja pilih jujur saya heran otak yang mana yg guna pikir br otak kanan otak kiri atau otak belakang pak prabowo tetap kita hormat krn jadi capres dengan cara yg benar lah si gibran jadi cawapres lewat jalur mk bagi saya prabowo oke dan wajib kita jaga jika pak prabowo jatuh maka yg naik adalah gibran mau minta gibran mundur sampai	Netral

bu iriana mati gibran tidak akan mundur kita keras pada pak prabowo agar beliau tidak jatuh

3.3. Wordcloud

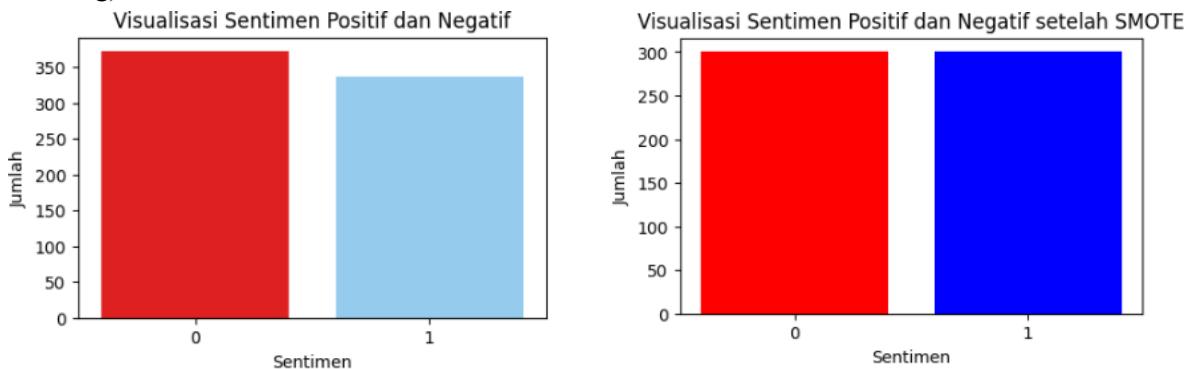
Wordcloud adalah visualisasi kata-kata yang paling sering muncul dalam dataset yang memiliki sentimen negatif maupun sentimen positif dapat dilihat seperti Gambar 7.



Gambar 2. Visualisasi Sentimen Positif dan Sentimen Negatif

3.4. Penerapan SMOTE

Jumlah komentar untuk setiap kelas diseimbangkan menjadi 300 setelah metode SMOTE diterapkan. Teknik ini dirancang untuk memastikan bahwa model tidak terlalu condong ke salah satu kelas dan memiliki kemampuan untuk belajar dari representasi data yang lebih setara. Pada Gambar 3, warna biru menunjukkan sentimen positif, sedangkan warna merah menunjukkan sentimen negatif. Gambar 3 juga menunjukkan perbandingan antara sentimen yang menggunakan SMOTE (data sentimen positif dan negatif dengan kondisi yang seimbang) dan yang tanpa SMOTE (data sentimen positif dan negatif dengan kondisi yang tidak seimbang).



Gambar 3. Visualisasi Sentimen Positif dan Sentimen Negatif

3.5. Pembagian Data Latih dan Data Uji

Penelitian ini memecah data dengan proporsi data latih sebesar 80% dan data uji sebesar 20%.

3.6. Evaluasi Model

Tabel 3 menunjukkan hasil evaluasi penelitian yang menghasilkan akurasi, presisi, skor *f1-score*, dan dukungan untuk *Naive Bayes* dan *Naive Bayes* dengan seleksi fitur *chi-square*. Percobaan menggunakan *Naive*

Bayes menghasilkan akurasi sebesar 60,5%, dan *Naïve Bayes* dengan *Chi-Square* menghasilkan akurasi sebesar 68,3%, yang menunjukkan peningkatan akurasi sebesar 7,8%.

Tabel 3. Hasil Evaluasi Model *Naïve Bayes* dan *Naïve bayes* menggunakan *Chi-Square*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>		<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
negatif	0.57	0.90	0.70	72	negatif	0.63	0.89	0.74	72
positif	0.75	0.30	0.43	70	positif	0.80	0.47	0.59	70
<i>accuracy</i>			0.605	142	<i>accuracy</i>			0.683	142

4. KESIMPULAN

Penelitian ini menunjukkan bahwa klasifikasi sentimen komentar pengguna YouTube lebih akurat dengan menggunakan algoritma *Naïve Bayes*, seleksi fitur *Chi-Square*, dan teknik SMOTE. Hasil evaluasi menunjukkan bahwa, dibandingkan dengan model dasar *Naïve Bayes* tanpa seleksi fitur, akurasi model meningkat sebesar 7,8% setelah penggunaan *Chi-Square* dan SMOTE. Sementara *Chi-Square* memainkan peran penting dalam menghilangkan fitur yang tidak relevan, meningkatkan kualitas representasi data. ketidakseimbangan kelas yang sebelumnya menjadi kendala untuk opini masyarakat berhasil. Oleh karena itu, metode ini dapat digunakan untuk membangun sistem klasifikasi opini publik yang otomatis dan akurat.

DAFTAR PUSTAKA

- [1] Kurnia, I. Purnamasari, and D. D. Saputra, “Analisis Sentimen Dengan Metode Naïve Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 235–242, 2023, doi: 10.35870/jtik.v7i3.707.
- [2] R. Md, R. D. Restiyan, and H. Irsyad, “Analisis Sentimen Masyarakat terhadap Perilaku Lawan Arah yang diunggah pada Media Sosial Youtube Menggunakan Naïve Bayes,” *BANDWIDTH J. Informatics Comput. Eng.*, vol. 02, no. 02, pp. 75–83, 2024.
- [3] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [4] H. Hidayatullah, P. Purwantoro, and Y. Umaidah, “Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, pp. 1546–1553, 2023, doi: 10.36040/jati.v7i3.6887.
- [5] Rosdiana, V. Novalia, I. Saputra, M. Ula, and M. Danil, “Application of Artificial Intelligence Chi-Square Model and Classification Of KNN in Heart Disease Detection,” vol. 6, no. July, pp. 180–188, 2022, doi: 10.31289/jite.v6i1.7343.
- [6] R. P. I. Putra, M. Akbar, and R. Amalia, “Analisis Sentimen Masyarakat Terhadap Kinerja Persatuan Sepakbola Seluruh Indonesia Menggunakan Metode Backpropagation,” *J. Inf. Technol. Ampera*, vol. 1, no. 2, pp. 106–118, 2020, doi: 10.51519/jurnalita.volume1.issuse2.year2020.page106-118.
- [7] E. L. W. Ninggrum and P. A. Widodo, “Implementasi Metode Multinomial Naive Bayes Classifier untuk Analisis Sentimen,” *Sustain.*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahhttps://www.researchgate.net/publication/305320484_SISTEM PEMBETUNGAN TERPUSAT STRATEGI MELESTARI
- [8] R. Hanifah and I. S. Nurhasanah, “Implementasi Web Crawling untuk Mengumpulkan Informasi Wisata Kuliner di Bandar Lampung,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 531, 2018, doi: 10.25126/jtiik.201855842.
- [9] T. Ernayanti, M. Mustafid, A. Rusgiyono, and A. R. Hakim, “Penggunaan Seleksi Fitur Chi-Square Dan Algoritma Multinomial Naïve Bayes Untuk Analisis Sentimen Pelanggan Tokopedia,” *J. Gaussian*, vol. 11, no. 4, pp. 562–571, 2023, doi: 10.14710/j.gauss.11.4.562-571.
- [10] A. Alamsyah and F. N. Zuhri, “Measuring Public Sentiment Towards Services Level in Online Forum using Naive Bayes Classifier and Word Cloud,” *CRS-ForMIND Int. Conf. Work. 2017*, no. October, 2017.
- [11] M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin

- Sinovac)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.
- [12] L. Amatullah, Y. Widiastiwi, and N. Chamidah, "Penerapan Klasifikasi Random Forest Terhadap Data Gangguan Spektrum Autisme (ASD) Pada Anak-Anak Menggunakan Seleksi Fitur Principal Component Analysis," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 3, no. 2, pp. 356–364, 2022.
- [13] I. Kurniawan, A. Lia Hananto, S. Shofia Hilabi, A. Hananto, B. Priyatna, and A. Yuniar Rahman, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [14] B. Paula, M. Fawzan, and H. Irsyad, "Analisis Sentiment Masyarakat Terhadap penyebaran Starlink di Indonesia Menggunakan Algoritma Naive Bayes," vol. 02, no. 2, pp. 141–148, 2024.
- [15] N. Asmiati and Fatmawati, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Pengaruh Negatif Game Online Bagi Remaja Milenial," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 3, pp. 141–149, 2020, doi: 10.35746/jtim.v2i3.102.
- [16] M. I. Fikri, T. S. Sabrina, and Y. Azhar, "Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020.
- [17] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.