



Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi *Naive Bayes*

Sella Adelia¹, Eva Milanda², Jena Santari³, Dea Talia Kesuma⁴, Echa Silvia⁵,
Fandi Kurniawan⁶

^{1,2,3,4,5,6}Sistem Teknologi informasi, Universitas Muhammadiyah Kotabumi, Kotabumi, Indonesia
Email: sella.2059201069@umko.ac.id¹, evami.2059201064@umko.ac.id²,
jenas.2059201048@umko.ac.id³, deaktb02@gmail.com⁴, echas.2059201042@umko.ac.id⁵,
fandi.kurniawan@umko.ac.id⁶

Abstrak

Di era digital yang terus berkembang, media sosial, khususnya YouTube, telah menjadi sumber utama informasi dan pembelajaran yang mudah diakses melalui telepon seluler. Bukan hanya sebagai alat interaksi, media sosial juga menawarkan ilmu pengetahuan dan pembelajaran yang berlimpah, termasuk tutorial pemrograman dan pengembangan perangkat lunak. YouTube khususnya menarik karena menyediakan konten yang mudah dipahami dan menarik, seperti penggunaan algoritma Naive Bayes, yang sering dibahas namun kurang dipahami secara luas. Penelitian ini berfokus pada analisis sentimen belajar pemrograman di YouTube, menggunakan algoritma klasifikasi Naive Bayes untuk mengklasifikasikan sentimen dalam komentar sebagai positif, negatif, atau netral. Hasil menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen dengan akurasi 100% untuk sentimen positif dan 66,39% untuk sentimen negatif. Analisis ini memberikan wawasan tentang pandangan dan respons pengguna terhadap konten pembelajaran pemrograman di YouTube, yang dapat menjadi dasar untuk perbaikan dan pengembangan konten edukasi lebih lanjut.

Kata Kunci: *Naive Bayes*, Media Sosial, *Pemrograman*, Analisa Sentiment, Youtube

1. PENDAHULUAN

Era digital saat ini ditandai oleh perkembangan internet yang luar biasa cepat, membawa dampak signifikan pada cara masyarakat global berinteraksi dan mengakses informasi. Fenomena ini khususnya terlihat dalam peran penting yang dimainkan oleh media sosial dalam kehidupan sehari-hari. Media sosial, yang muncul bersamaan dengan evolusi teknologi web 2.0, telah menjadi platform yang tidak hanya memungkinkan pengguna untuk berkomunikasi tetapi juga untuk berpartisipasi aktif dalam pembuatan dan pertukaran konten. Salah satu platform yang paling menonjol dalam hal ini adalah YouTube, yang telah berkembang



menjadi lebih dari sekadar alat hiburan, menjadi medium edukasi dan pertukaran informasi yang berharga [1].

YouTube, dengan popularitasnya yang luas, telah menjadi salah satu jaringan sosial yang sering digunakan untuk berbagai kegiatan, mulai dari hiburan hingga edukasi. Platform ini telah memainkan peran penting dalam bidang pendidikan, khususnya dalam pembelajaran online [2]. Salah satu area yang menarik perhatian khusus adalah pembelajaran pemrograman, dengan fokus pada penggunaan algoritma Naive Bayes, sebuah topik yang menarik bagi banyak pemuda dan pemudi yang mencari sumber pembelajaran tambahan melalui YouTube [3]. Materi yang disajikan oleh para kreator konten di YouTube sering kali dikemas dalam format yang mudah dipahami, dengan menggunakan bahasa yang santai dan pendekatan yang interaktif, membuatnya menjadi sumber yang sangat berharga bagi mahasiswa dan pelajar.

Konten-konten ini tidak hanya menawarkan pengetahuan tetapi juga memungkinkan pemuda dan mahasiswa untuk memahami berbagai konsep pemrograman dengan cara yang lebih efektif. Para YouTuber seringkali menyampaikan materi mereka dalam durasi yang singkat, namun tetap kaya informasi, sehingga mudah dipahami oleh penonton. Hal ini memungkinkan pengguna untuk tidak hanya mempelajari teori tetapi juga menerapkan konsep-konsep tersebut secara praktis. Dengan demikian, mereka dapat menilai kualitas konten dan memanfaatkannya sebagai referensi dan sumber belajar tambahan dalam pembelajaran pemrograman [4].

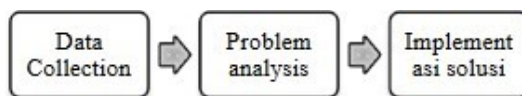
Dalam konteks ini, penelitian ini bertujuan untuk menganalisis bagaimana platform seperti YouTube dapat meningkatkan proses pembelajaran pemrograman, khususnya melalui penggunaan algoritma klasifikasi Naive Bayes. Algoritma Naive Bayes, yang terkenal karena pendekatannya yang berbasis probabilitas sederhana dan asumsi independensi antar variabel, menawarkan sejumlah keuntungan. Ini termasuk tingkat error yang rendah dan akurasi yang lebih tinggi dalam dataset besar, menjadikannya alat yang sangat efektif untuk pembelajaran pemrograman [5]. Penelitian ini bertujuan untuk mengeksplorasi bagaimana algoritma ini dapat dimanfaatkan dalam konteks edukasi, khususnya dalam memfasilitasi pembelajaran yang lebih mudah diakses dan dipahami.

Tujuan akhir dari penelitian ini adalah untuk memberikan wawasan tentang bagaimana pembelajaran pemrograman melalui media sosial, khususnya YouTube, dapat disederhanakan dan dibuat lebih mudah diakses di masa depan. Dengan memanfaatkan keunggulan algoritma Naive Bayes dalam analisis dan pengklasifikasian data, penelitian ini akan mengkaji bagaimana metode ini dapat diterapkan untuk memperkaya pengalaman pembelajaran online. Hal ini

diharapkan akan memberikan gambaran yang lebih jelas tentang bagaimana algoritma pembelajaran dapat disesuaikan untuk meningkatkan kualitas dan keterjangkauan pendidikan, khususnya dalam bidang pemrograman. Dengan demikian, hasil penelitian ini akan berkontribusi pada pengembangan metode pembelajaran yang lebih efektif dan inklusif untuk masa depan.

2. METODE

Tahapan penelitian ini terdapat tiga tahapan dapat di lihat pada Gambar 1.



Gambar 1: Tahapan proses penelitian.

Proses Data Collection pada penelitian ini terdapat beberapa langkah yang pertama di lakukan adalah literatur review lalu dilanjutkan dengan pengamatan atau observasi secara langsung seperti yang diperlihatkan Gambar 2.

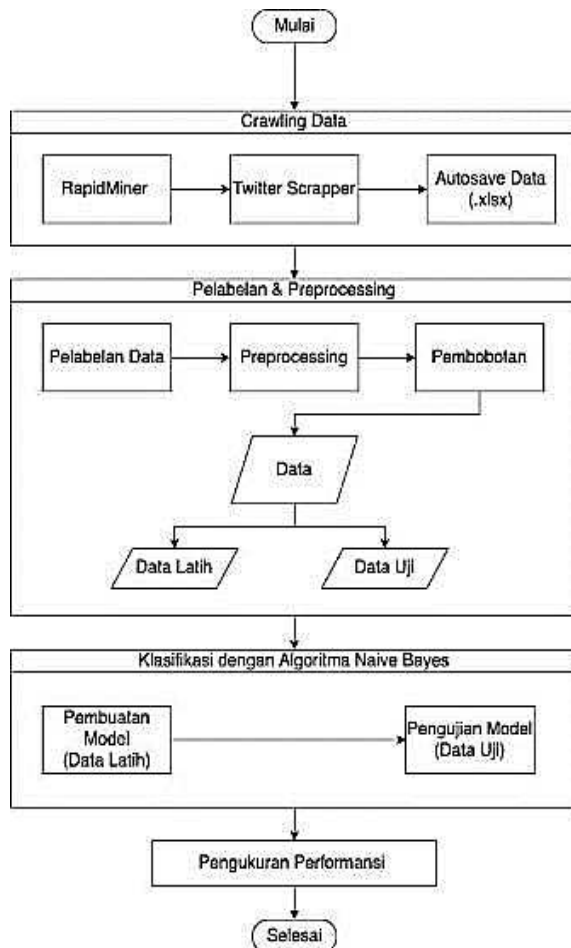


Gambar 2: Tahapan Proses Data Collection

Literatur review membantu proses penelitian ini antara lain jurnal yang berhubungan dengan data mining dan machine learning proses klasifikasi file atau data dokumen. Sedangkan proses pengamatan atau observasi menggunakan teknik observasi langsung. Pengamatan atau Observasi dilakukan dengan mencermati video-video yang diunggah pada YouTube secara cermat dan teliti di pahami mendalam. Pada tahapan analisis masalah.

Masalahan yang di dapatkan melalui analisis dokumen dan observasi langsung, kemudian diusulkan solusi dengan melihat perkembangan pada teknologi dan ilmu pengetahuan. Masalah yang terdeteksi dapat menimbulkan kurang pahaman penonton dalam melihat video tutorial. Penerapan solusi adalah langkah solusi akhir dalam fase penelitian, ada tiga kategorgi algoritma Naïve Bayes pada alat Rapid Miner yang di libatakan pada penelitian ini pengklasifikasian atau analisis data sentimen[5].

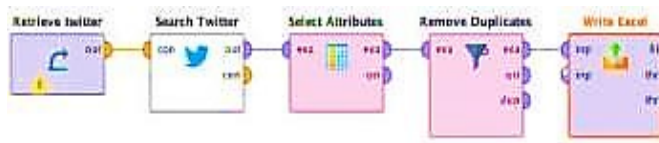
Pada analisis sentimen, beberapa proses tahapan harus dilakukan agar mencapai hasil pengujian yang maksimal. Gambar 3 menunjukkan proses tahapan dilakukan, mulai dari data mining, pemberian label, proses pengolahan dan pemberian label pada prediksi [6].



Gambar 3. Flowchart System Klasifikasi Analisis Sentimen

1) Crawling Data/ Data mining

Adalah langkah pengambilan data yang dijadikan referensi berupa para pengguna YouTube lainnya. Data diunggah dengan cara otomatis dan kata kunci "ketua" pada komentar, menggunakan fasilitas penghapusan komentar di RapidMiner, diikuti dengan penghapusan konten duplikat yang menghapus konten yang sama dan secara otomatis disimpan dalam file Excel[7]. Rantai proses data mining pada Gambar 4. Di bawah ini



Gambar 4. Proses Crawling Data

2) Pelabelan

Adalah langkah yang menggunakan pengelompokan data dengan cara pemberian label kelas. berikut contoh dataset diberi label seperti gambar 5.

	text
0	Anis pembohong pinter dodol jamu Prabowo top
1	Yo prabowo Prabowo pokok e
2	Pokoknya prabowo
3	Saya yakin prabowo orang yg tegas,,semoga bel...
4	seorang jendral akan sil siap mempertahankan k...
5	Percaya... gagasan pak Prabowo ok, wakilnya ga...
6	bakalan stroke ke.3 karena kalah
7	GILAAAA bapak keren untuk 10mnt awal, saya sam...
8	Smoga pk prabowo menjagi pemimpin negara ini d...
9	Saya ingin melihat sebelum saya meninggal, Ind...
10	Ya Allah semoga pak prabowo bisa menjadi pemim...

Gambar 5. Pemberian Label Kelas

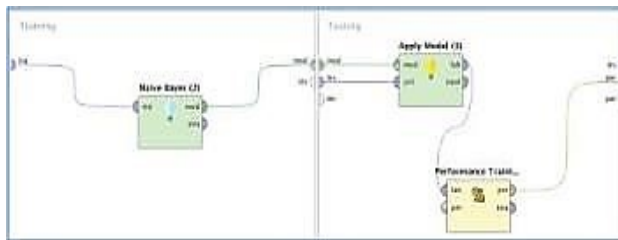
3) Preprocessing

Selama tahap prapemrosesan, beberapa proses akan dilakukan agar data asli siap untuk diproses. Pemrosesan awal data terjadi dalam beberapa langkah, termasuk pembersihan, pelipatan kotak, tokenisasi, normalisasi, pemfilteran, dan rooting[8]. Cleanup Pada tahap ini dilakukan pembersihan atribut username, hashtag, URL, tanda baca, dan simbol.

- Case Folding merupakan Tahapan normalisasi isi tweet menjadi huruf non kapital.
- Tokenisasi merupakan Tahapan pemisahan setiap kata dalam sebuah kalimat menjadi satuan kata.
- Normalize, tahapan ini memiliki fungsi mengurangi kata atau mengubah kata yang baku menjadi tidak baku.
- Filtering, tahapan mengubah atau menghilangkan kata yang tampil namun tidak berdampak besar atau tidak bermakna.
- Stemming adalah Tahapan merubah kata yang memiliki imbuhan menjadi kata dasar.
- Pembobotan, tahapan pembobotan ini, data diolah terlebih dahulu kemudian dibobotkan menggunakan *TF-IDF*. data uji dan data eksperimen merupakan tempat penyimpanan hasil pembobotan ini.

4) Klasifikasi

Klasifikasi Adalah tahapan pengklasifikasian data yang sudah melalui tahapan sebelumnya. Langkah ini termasuk memastikan penerapan *algoritme Naïve Bayes*. Gambar 6. merupakan isi operator “*Cross Validation*” Di aplikasi RapidMiner.



Gambar 6. Cross Validation Operator

3. HASIL DAN PEMBAHASAN

3.1 *Crawling Data*

Crawling data Mengumpulkan dataset yang relevan dengan studi kasus, seperti review tentang pembahasan belajar programming menggunakan netlytic.

3.2 *Pre-Processing*

Pre-Processing Pra-pemrosesan data (atau pre processing data) adalah tahap awal dalam analisis data di mana data yang telah dikumpulkan dipersiapkan untuk analisis lebih lanjut[9]. Berikut tahapannya, Pembersihan Data (*Data Cleaning*), Transformasi Data (*Data Transformation*), Encoding Variabel Kategorikal (*Categorical Variable Encoding*), Penghapusan Noise (*Noise Removal*), Pengaturan Fitur (*Feature Scaling*), Ekstraksi Fitur (*Feature Extraction*). Data dibersihkan menggunakan *software google colab*.

3.3 TF-IDF

Menurut [10], metode Term Frequency Invers Document Frequency (TF-IDF) Adalah metod yang digunakan untuk menentukan seberapa jauh keterhubungan kata (term) terhadap dokumendengan memberikan bobot setiap kata pada saat menghitung bobot digunakan TF-IDF, yang pertama dihitung nilai TF per kata pad bobot masing-masing Sedangkan nilai IDF diformulasikan pada Persamaan 1.

$$IDF(word) = \log \frac{td}{df} \quad (1)$$

The screenshot displays a Dataiku pipeline with the following tasks and connections:

- Read Excel**: A light blue task box with an 'inp' port on the left, 'fil' and 'out' ports on top, and a green checkmark icon. It connects to the 'Sort' task.
- Sort**: A pink task box with 'exA' and 'ex.B' ports on top, an 'on' port on the right, and a green checkmark with a yellow warning icon. It connects to the 'Filter Example Range' task.
- Filter Example Range**: A pink task box with 'exA' and 'ex.B' ports on top, an 'on' port on the right, and a green checkmark with a yellow warning icon. It connects to the 'WordList to Data' task.
- WordList to Data**: A tan task box with 'wor' and 'ex.B' ports on top, and a green checkmark icon. It connects to the 'Hominal to Text' task.
- Hominal to Text**: A pink task box with 'exA' and 'ex.B' ports on top, an 'on' port on the right, and a green checkmark icon. It connects to the 'Process Documents...' task.
- Process Documents...**: A tan task box with 'wor' and 'ex.B' ports on top, and a green checkmark icon. It has an 'out' port on the right that connects to a 'res' output port.

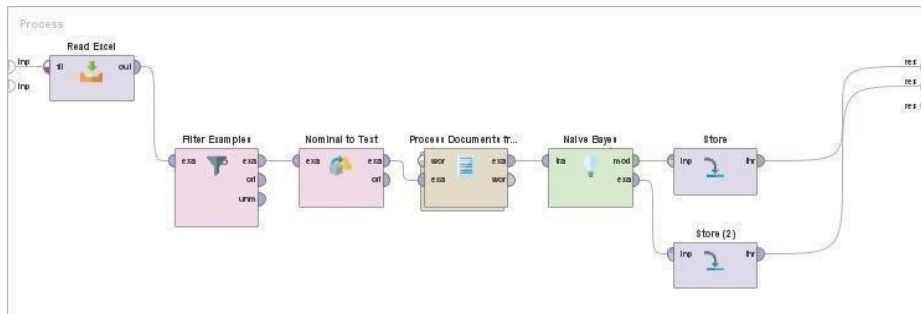
3.4 WordCloud

The image displays two word clouds generated from the same text. The left word cloud is a standard horizontal layout, while the right word cloud is a circular arrangement. Both feature the words 'belajar' (learn) and 'bang' (wow) as the most prominent terms.

260 | Analisis Sentimen Belajar Programming pada Media Sosial

3.5 Naive Bayes Classifier

Adalah metode pembelajaran mesin yang menggunakan perhitungan probabilitas dan statistik [3]. Menurut thomas bayes ilmuwan asal ingris, naïve bayes classifier merupakan probabilitas masa depan di ambil dari pengalaman masa sebelumnya. Gambar 8, 9, 10, 11, dan 12 adalah proses naïve bayes classifier yang dilakukan.



Gambar 8 Tokenize proses dokumen to data

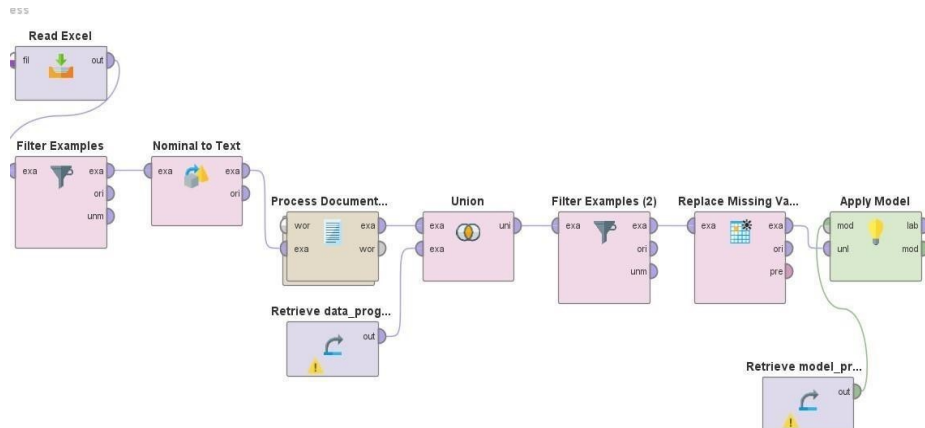


Gambar 9 Hasil data training

Open in  Turbo Prep  Auto Model Filter (1,499 / 1,499 examples):

Row No.	sentimen	text	aalamin	aamin	abad	abang	abangnya	abdurrahman	abis	acak
1	positif	sudah cantik p...	0	0	0	0	0	0	0	0
2	negatif		0	0	0	0	0	0	0	0
3	negatif	bu tolong solu...	0	0	0	0	0	0	0	0
4	positif	kupon yg dapa...	0	0	0	0	0	0	0	0
5	negatif	hi business lea...	0	0	0	0	0	0	0	0
6	positif	coach ada kah...	0	0	0	0	0	0	0	0
7	positif	penjelasannya ...	0	0	0	0	0	0	0	0
8	negatif	works bantuan...	0	0	0	0	0	0	0	0
9	positif	keren	0	0	0	0	0	0	0	0
10	negatif	thank atad inspir...	0	0	0	0	0	0	0	0
11	negatif	assstrima kasi...	0	0	0	0	0	0	0	0
12	positif	terimakasih cici ...	0	0	0	0	0	0	0	0
13	positif	bersyukur pad...	0	0	0	0	0	0	0	0
14	positif	terima kasih il...	0	0	0	0	0	0	0	0
15	positif	jgn di tawar ta...	0	0	0	0	0	0	0	0
16	positif	hemat ya jalan ...	0	0	0	0	0	0	0	0
17	positif	terima kasih p...	0	0	0	0	0	0	0	0

Gambar 10 Data analisi sentiment prediksi



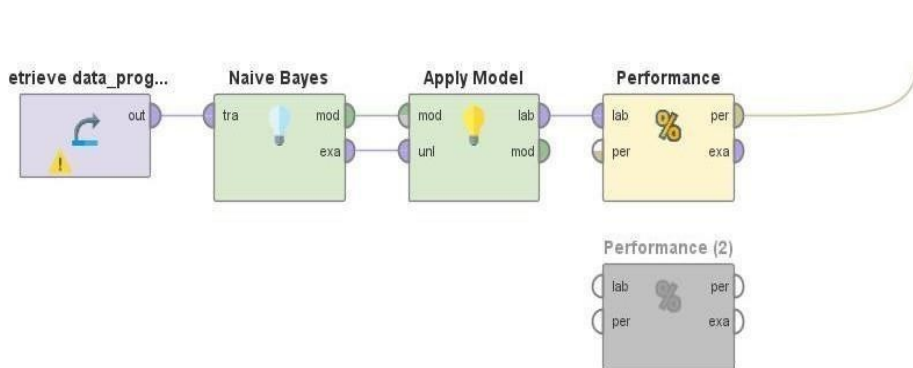
Gambar 11 Proses Pembuatan Model

Row No.	sentimen	prediction[s...	confidence[...	confidence[...	confidence[...	text	aamlin	abang	about	absen
1	?	positif	1	0	0	kuli restoran ti...	0	0	0	0
2	?	positif	1	0	0	programming t...	0	0	0	0
3	?	negatif	0	1	0	keren	0	0	0	0
4	?	positif	1	0	0	bang tolong m...	0	0	0	0
5	?	positif	1	0	0	permisi bang p...	0	0	0	0
6	?	negatif	0	1	0	nuhun kang	0	0	0	0
7	?	positif	1	0	0	heran mata m...	0	0	0	0
8	?	negatif	0	1	0	kehaling kaca...	0	0	0	0
9	?	negatif	0	1	0	chanel benera...	0	0	0	0
10	?	positif	1	0	0	terimakasih m...	0	0	0	0
11	?	negatif	0	1	0	belajar coding ...	0	0	0	0
12	?	negatif	0	1	0	salfok candy b...	0	0	0	0
13	?	negatif	0	1	0	chanel kaya gi...	0	0	0	0
14	?	negatif	0	1	0	temuan erro m...	0	0	0	0
15	?	positif	1	0	0	tujuan disuruh ...	0	0	0	0
16	?	positif	1	0	0	terimakasih tip...	0	0	0	0
17	?	positif	1	0	0	software coco...	0	0	0	0

Gambar 12. Hasil Model

3.6 Performance

Kinerja (performance) dalam data mining adalah ukuran sejauh mana suatu sistem atau algoritma data mining mampu menjalankan tugasnya dengan efisien dan efektif. Kinerja ini dapat diukur dengan berbagai metrik yang sesuai dengan tujuan analisis data yang dilakukan [12]. Beberapa faktor yang dapat memengaruhi kinerja dalam data mining meliputi akurasi hasil, kecepatan eksekusi, kemampuan untuk mengekstrak pola yang berarti, serta kemampuan untuk menangani data yang besar dan kompleks.



Gambar 13 Naïve Bayes Operator

accuracy: 80.99%

	true positif	true negatif	true negatis	class precision
pred. positif	650	0	0	100.00%
pred. negatif	285	563	0	66.39%
pred. negatis	0	0	1	100.00%
class recall	69.52%	100.00%	100.00%	

Gambar 14 Akurasi Naïve Bayes

4. KESIMPULAN

Hasil dari akurasi Naïve Bayes di atas adalah sentiment positif memiliki prediksi 100% sedangkan sentiment Negatis sebanyak 66,39%. Penulisan tersebut bertujuan untuk menganalisis sentimen belajar pemrograman yang dibagikan dalam komentar-komentar di platform media sosial YouTube. Metode yang digunakan dalam analisis adalah algoritma *klasifikasi Naïve Bayes*, sebuah teknik dalam pembelajaran mesin yang dapat mengklasifikasikan sentimen-sentimen tersebut ke dalam kategori positif, negatif, atau netral. Hasil penelitian menunjukkan kemampuan algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen-sentimen pada komentar-komentar belajar programming di YouTube dengan tingkat akurasi yang signifikan. Analisis ini dapat memberikan wawasan tentang pandangan, evaluasi, dan respon pengguna terhadap konten pembelajaran programming di platform YouTube, yang dapat digunakan sebagai dasar untuk perbaikan atau pengembangan lebih lanjut dalam konten edukasi tersebut.

REFERENSI

- [1] D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF

- Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021, doi: 10.25126/justsi.v2i1.24.
- [2] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [3] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
- [4] A. Abdullah, A. Achmad, and S. Sahibu, "Media Pembelajaran Interaktif Mata Kuliah Pemograman Web Berbasis Android," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 11, no. 1, p. 45, 2021, doi: 10.35585/inspir.v11i1.2626.
- [5] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017, doi: 10.33369/pseudocode.4.2.120-128.
- [6] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 296, 2022, doi: 10.30865/json.v4i2.5381.
- [7] A. Rachmat C and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.
- [8] B. W. I. Riski, K. Dwi, R. Nooraeni, T. Siahaan, and Y. Dhea, "Analisis Text Mining dari Cuitan Twitter Mengenai Infrastruktur di Indonesia dengan Metode Klasifikasi Naïve Bayes," *Eig. Math. J.*, vol. 2, no. 2, pp. 92–101, 2019, doi: 10.29303/emj.v1i2.36.
- [9] M. N. Akbar and Nirwana Samrin, "Analisis Sentimen Komentar Pengguna Aplikasi Threads Pada Google Playstore Menggunakan Algoritma Multinomial Naive Bayes Classifier," *AGENTS J. Artif. Intell. Data Sci.*, vol. 3, no. 2, pp. 21–29, 2023, doi: 10.24252/jagti.v3i2.67.
- [10] R. Putra NH, "Penerapan Teknik Web Scraping & Metode Term Frequency-Invers Document Frequency (Tf-Idf) Pada Pencarian Lowongan Kerja," 2022.
- [11] E. Purnaningrum and I. Ariqoh2, "Google Trends Analytics Dalam Bidang Pariwisata," *Maj. Ekon.*, vol. 24, no. 2, pp. 232–243, 2019, doi: 10.36456/majeko.vol24.no2.a2069.
- [12] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013, doi: 10.1038/hdy.2009.180.