



Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Tjong A Fie Mansion Menggunakan Metode Naïve Bayes Classifier

Sentiment Analysis of Visitor Comments on Tjong A Fie Mansion Tourist Attraction Using the Naïve Bayes Classifier Method

Erlina Siregar & Andre Hasudungan Lubis*

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Medan Area, Indonesia

Diterima: 21 April 2025; Direview: 22 April 2025; Disetujui: 24 April 2025

*Corresponding Email: andrehasudunganlubis@uma.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen komentar pengunjung terhadap objek wisata Tjong A Fie Mansion di Kota Medan dengan menggunakan metode Naïve Bayes Classifier. Data diperoleh secara manual dari Google Maps sebanyak 100 komentar, yang kemudian melalui tahapan preprocessing meliputi case folding, tokenisasi, penghapusan stopword, dan stemming. Selanjutnya, dilakukan ekstraksi fitur menggunakan metode TF-IDF serta proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Evaluasi kinerja model dilakukan dengan menggunakan confusion matrix. Hasil pengujian menunjukkan bahwa pembagian data pelatihan sebesar 80% dan data pengujian sebesar 20% menghasilkan akurasi tertinggi, yaitu sebesar 80% serta hasil sentimen 100% positif. Temuan ini menunjukkan bahwa metode Naïve Bayes mampu mengklasifikasikan komentar berbasis teks secara efektif dan efisien. Hasil analisis sentimen ini diharapkan dapat memberikan masukan bagi pengelola objek wisata dalam meningkatkan kualitas pelayanan, serta menjadi referensi dalam pengembangan sistem pendukung keputusan berbasis opini pengguna.

Kata Kunci: Analisis Sentimen; Naïve Bayes Classifier; Tempat Wisata; TF-IDF; Klasifikasi Teks.

Abstract

This study aims to analyze the sentiment of visitor comments on the Tjong A Fie Mansion tourist attraction in Medan City using the Naïve Bayes Classifier method. A total of 100 comments were manually collected from Google Maps and underwent preprocessing stages, including case folding, tokenization, stopword removal, and stemming. Feature extraction was then performed using the TF-IDF method, followed by classification using the Multinomial Naïve Bayes algorithm. Model performance was evaluated using a confusion matrix. The test results showed that a data split of 80% for training and 20% for testing yielded the highest accuracy, reaching 80%, with a sentiment classification result of 100% positive. These findings indicate that the Naïve Bayes method can effectively and efficiently classify text-based comments. The sentiment analysis results are expected to provide input for tourism managers to improve service quality and serve as a reference for the development of user opinion-based decision support systems.

Keywords: Sentiment Analysis; Naïve Bayes Classifier; Tourist Attraction; TF-IDF; Text Classification.

PENDAHULUAN

Dalam era digital saat ini, perkembangan teknologi informasi telah mengubah cara masyarakat mengekspresikan opini dan berbagi pengalaman, termasuk dalam sektor pariwisata [1], [2]. Media sosial dan platform ulasan daring menjadi sarana utama bagi wisatawan untuk menyampaikan penilaian mereka terhadap suatu destinasi. Data yang dihasilkan dari interaksi ini menjadi sumber informasi berharga yang dapat dianalisis untuk memahami persepsi publik terhadap layanan wisata [3], [4]. Salah satu destinasi budaya yang populer di Kota Medan adalah Tjong A Fie Mansion, rumah bersejarah peninggalan saudagar Tionghoa terkenal yang telah dibuka untuk umum sejak tahun 2009 [5].

Tjong A Fie Mansion mencatat tingkat kunjungan yang tinggi, mencapai lebih dari seribu pengunjung per bulan. Keunikan arsitektur, nilai sejarah, dan lokasinya di kawasan perdagangan Kesawan menjadikan rumah ini objek wisata unggulan [6], [7]. Namun, pandemi COVID-19 sempat menghentikan operasionalnya untuk membatasi kerumunan, mencerminkan pentingnya pemahaman terhadap dinamika kunjungan wisatawan dan persepsi mereka. Dalam konteks inilah, analisis terhadap komentar pengunjung menjadi krusial untuk mengevaluasi dan meningkatkan kualitas layanan pariwisata [6].

Latar belakang ini mendorong perlunya penelitian yang memanfaatkan teknik analisis sentimen terhadap komentar pengunjung, yang banyak tersedia secara daring, khususnya melalui Google Maps [8]. Komentar-komentar ini mengandung opini yang mencerminkan pengalaman subjektif wisatawan, yang bila diolah secara sistematis dapat memberikan wawasan bagi pengelola destinasi. Untuk mengklasifikasikan sentimen dalam komentar tersebut, metode Naïve Bayes Classifier dipilih karena kemampuannya dalam menangani data teks dan menghasilkan klasifikasi yang efisien [9].

Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen pengunjung terhadap tempat wisata Tjong A Fie Mansion ke dalam kategori positif atau negatif. Penelitian ini menggunakan pendekatan Multinomial Naïve Bayes dengan serangkaian tahap seperti preprocessing data, ekstraksi fitur menggunakan TF-IDF, dan evaluasi model dengan confusion matrix [10].

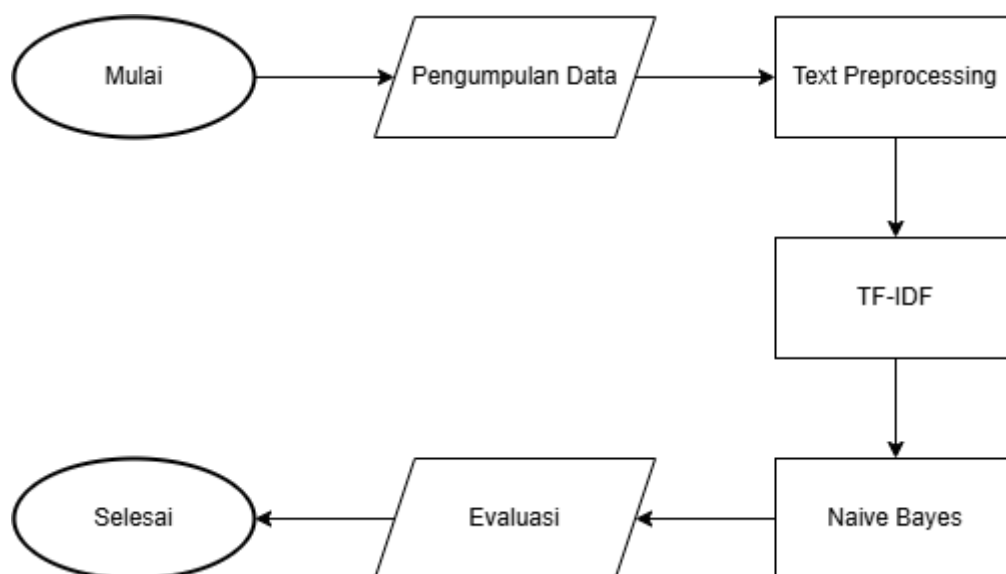
Secara umum, studi ini ingin mengisi celah dalam penelitian terdahulu yang masih terbatas dalam pemanfaatan data ulasan daring untuk mengevaluasi objek wisata

berbasis warisan budaya. Banyak studi analisis sentimen berfokus pada produk atau jasa digital, namun kajian dalam konteks pariwisata lokal, khususnya pada objek bersejarah seperti Tjong A Fie Mansion, masih relatif jarang dilakukan.

Penelitian ini memiliki signifikansi teoretis dan praktis. Secara teoretis, ia memberikan kontribusi dalam pengembangan model analisis sentimen berbasis teks di ranah pariwisata. Secara praktis, hasil penelitian ini dapat menjadi dasar pertimbangan bagi pengelola wisata dalam meningkatkan layanan, menyusun strategi promosi yang lebih efektif, serta mendukung pengambilan keputusan berbasis opini publik.

METODE PENELITIAN

Metode penelitian dalam studi ini disajikan dalam bentuk diagram alur yang menggambarkan tahapan-tahapan yang dilalui sekaligus memberikan gambaran menyeluruh tentang keseluruhan proses penelitian.



Gambar 1 Alur Penelitian

Berikut merupakan penjelasan mengenai alur penelitian pada Gambar 1:

- Pengumpulan data pada penelitian ini berdasarkan berdasarkan komentar dari google maps. Data didapatkan secara manual sebanyak 100 *record* data dengan variabel nama pengguna dan komentar. Selanjutnya data disimpan di *Microsoft Excel* untuk dilakukannya tahap *text preprocessing*.
- Text Preprocessing* merupakan serangkaian tahapan yang dilakukan untuk mempersiapkan data teks mentah sebelum dianalisis atau digunakan dalam model

pemrosesan bahasa alami (*Natural Language Processing/NLP*) [11]. Tujuan utama dari *text preprocessing* adalah untuk membersihkan, menyederhanakan, dan menstandarkan teks agar dapat diproses secara lebih efisien dan akurat oleh algoritma komputer. Beberapa tahapan umum dalam *text preprocessing* meliputi: *Case Folding* – Mengubah seluruh huruf menjadi huruf kecil (*lowercase*) untuk menghindari duplikasi kata berdasarkan kapitalisasi [12]. Tokenisasi – Memecah teks menjadi unit-unit yang lebih kecil, seperti kata atau kalimat [13]. *Stopword Removal* – Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "dan", "di", "yang", dll [14]. Proses ini bertujuan untuk menghapus data yang tidak utuh, data yang mengandung kesalahan, dan kata kata yang tidak relevan [15]. *Stemming* – Mengubah kata ke bentuk dasarnya, misalnya "berlari", "lari-lari", menjadi "lari" [16].

- c. TF-IDF merupakan proses ekstraksi fitur dilakukan dengan memberikan bobot pada setiap kata yang terdapat dalam suatu dokumen, atau dengan kata lain mengonversi kata-kata menjadi representasi numerik. Langkah ini bertujuan untuk menilai sejauh mana pentingnya suatu kalimat dalam dokumen tersebut [17]. Adapun rumus TF-IDF adalah sebagai berikut [18]:

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

Keterangan:

df = banyak dokumen yang mengandung *term*

D = banyak seluruh dokumen

$$W_{ij} = tf \times IDF \quad (2)$$

$$W_{ij} = tf \times \log \frac{D}{df} \quad (3)$$

Keterangan:

W_{ij} = bobot kata dalam tiap-tiap dokumen

tf = banyak kemunculan kata dalam dokumen

D = banyak seluruh dokumen

- d. *Naïve Bayes*

Dengan asumsi dasar bahwa setiap peristiwa atau kondisi bersifat saling independen, algoritma *Naive Bayes* menjadi salah satu metode yang banyak digunakan

dalam proses *text mining*. Asumsi ini memungkinkan *Naive Bayes* untuk melakukan klasifikasi data teks secara efisien dan efektif, terutama ketika data memiliki banyak fitur atau kategori [19]. Dalam penelitian ini, digunakan metode *Multinomial Naive Bayes* (MNB) yang difokuskan pada klasifikasi data berbasis teks. Keunggulan MNB terletak pada sifat independensi hasil klasifikasinya, dimana setiap dokumen diklasifikasikan secara individual tanpa mempertimbangkan keterkaitan antar dokumen. Dengan demikian, MNB mampu menghasilkan klasifikasi yang sepenuhnya berdasarkan isi dari masing-masing dokumen, tanpa dipengaruhi oleh konteks atau dokumen lainnya. Persamaan 1 disajikan sebagai berikut [20]:

$$P = (c|d) \propto P(c) \prod_{i=1}^{n_d} P(W_1|c) \quad (4)$$

Penjelasan rumus Persamaan (4) di atas dimana $P(c|d)$ merupakan probabilitas dari suatu kelas c pada suatu *text* atau dokumen d , untuk $P(c)$ merupakan probabilitas prior c , dan $P(w_1|c)$ merupakan prioritas dari suatu kata pada kelas c .

e. Evaluasi

Confusion matrix adalah sebuah tabel yang digunakan untuk menampilkan jumlah data uji yang diklasifikasikan dengan benar dan salah, sehingga memudahkan dalam mengevaluasi akurasi suatu sistem klasifikasi. Dengan menggunakan *confusion matrix*, kita dapat melihat secara detail kinerja suatu sistem klasifikasi dan mengidentifikasi di mana terjadi kesalahan klasifikasi [9]. Tujuan utama dari *confusion matrix* adalah untuk menilai performa atau akurasi dari suatu sistem klasifikasi dalam mengklasifikasikan data uji [21]. Penelitian ini menggunakan *confusion matrix* sebagai evaluasi untuk menilai performa model klasifikasi dengan membandingkan hasil prediksi model dan label sebenarnya. *Confusion matrix* dapat dilihat pada gambar dibawah ini.

TP (True Positive)	FP (False Positive)
FN (False Negative)	TN (True Negative)

Gambar 2 Confusion Matrix

Keterangan pada Gambar 2:

- *True Positive* merupakan data positif yang diprediksi benar.
- *True Negative* merupakan data negatif yang diprediksi benar.
- *False Positive* merupakan data negatif namun diprediksi sebagai data positif.
- *False Negative* merupakan data positif namun diprediksi sebagai data negatif.

Adapun rumus untuk menghitung akurasi, presisi, *recall*, dan *f1 score*:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$presisi = \frac{TP}{TP + FP} \quad (6)$$

$$recall = \frac{TP}{TP + FN} \quad (7)$$

$$f1\ score = 2 \times \frac{recall \times presisi}{recall + presisi} \quad (8)$$

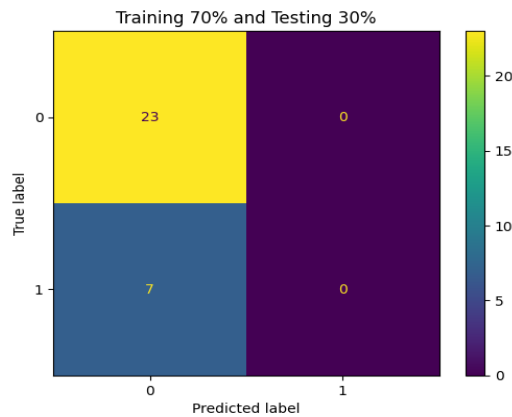
HASIL DAN PEMBAHASAN

Adapun hasil pada penelitian ini di antaranya sebagai berikut:

Tabel 1 Hasil Pengujian

Pengujian	Data Training	Data Testing	Akurasi	Presisi	Recall	F1 Score
1	70 %	30 %	76,67 %	0 %	0 %	0 %
2	80 %	20 %	80 %	0 %	0 %	0 %
3	90 %	10 %	70 %	0 %	0 %	0 %

Hasil pengujian ke-1 menggunakan *data training* 70% dan *testing* 30% mendapatkan hasil akurasi 76%, presisi 0%, *recall* 0%, dan *f1 score* 0%. Selanjutnya hasil pengujian ke-2 menggunakan *data training* 80% dan *testing* 20% mendapatkan hasil akurasi 80%, presisi 0%, *recall* 0%, dan *f1 score* 0%. Terakhir untuk pengujian ke-3 menggunakan *data training* 90% dan *testing* 10% mendapatkan hasil akurasi 70%, presisi 0%, *recall* 0%, dan *f1 score* 0%.



Gambar 3 Confusion Matrix Pengujian 1

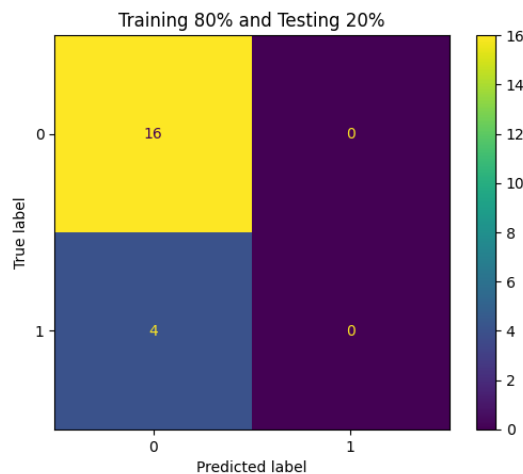
Pada Gambar 3 merupakan hasil *confusion matrix* dari pengujian 1 dengan jumlah *data training* 70% dan *testing* 30%. Berikut adalah untuk menghitung akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 23}{0 + 23 + 0 + 7} = \frac{23}{30} = 0.7667 \text{ atau } 76.67\%$$

$$Presisi = \frac{TP}{TP + FP} = \frac{0}{0 + 0} = 0 \text{ atau } 0\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 7} = 0 \text{ atau } 0\%$$

$$F1 \text{ Score} = 2 \times \frac{(Recal \times Presisi)}{(Recall + Presisi)} = 2 \times \frac{0 \times 0}{0 + 0} = 0 \text{ atau } 0\%$$



Gambar 4 Confusion Matrix Pengujian 2

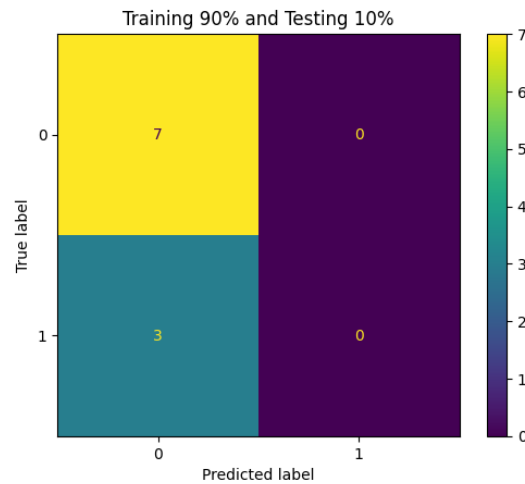
Pada Gambar 4 merupakan hasil *confusion matrix* dari pengujian 2 dengan jumlah *data training* 80% dan *testing* 20%. Berikut adalah untuk menghitung akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 16}{0 + 16 + 0 + 4} = \frac{16}{20} = 0.80 \text{ atau } 80\%$$

$$Presisi = \frac{TP}{TP + FP} = \frac{0}{0 + 0} = 0 \text{ atau } 0\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 4} = 0 \text{ atau } 0\%$$

$$F1 \text{ Score} = 2 \times \frac{(Recal \times Presisi)}{(Recall + Presisi)} = 2 \times \frac{0 \times 0}{0 + 0} = 0 \text{ atau } 0\%$$



Gambar 5 *Confusion Matrix* Pengujian 3

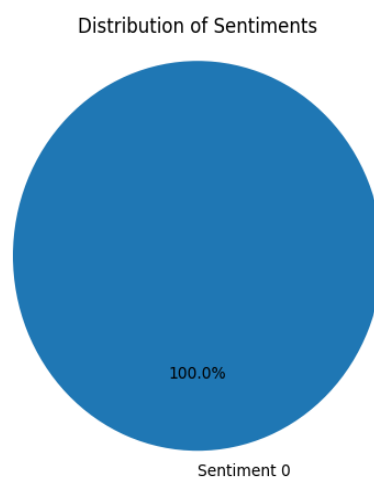
Pada Gambar 5 merupakan hasil *confusion matrix* dari pengujian 3 dengan jumlah *data training* 90% dan *testing* 10%. Berikut adalah untuk menghitung akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 7}{0 + 7 + 0 + 3} = \frac{7}{10} = 0.70 \text{ atau } 70\%$$

$$Presisi = \frac{TP}{TP + FP} = \frac{0}{0 + 0} = 0 \text{ atau } 0\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 3} = 0 \text{ atau } 0\%$$

$$F1 \text{ Score} = 2 \times \frac{(Recal \times Presisi)}{(Recall + Presisi)} = 2 \times \frac{0 \times 0}{0 + 0} = 0 \text{ atau } 0\%$$



Gambar 6 Hasil Persentase dari Sentimen

Pada Gambar 6 dapat dilihat bahwa sentimen positif (0) memiliki persentase sebesar 100%. Hasil ini menggunakan data proporsi data *training* 80% dan *testing* 20% dengan akurasi sebesar 80%.

SIMPULAN

Berdasarkan hasil penelitian, metode *Multinomial Naïve Bayes* terbukti efektif dalam mengklasifikasikan komentar pengunjung terhadap objek wisata Tjong A Fie Mansion ke dalam kategori positif dan negatif. Pengujian dilakukan dalam tiga skenario pembagian data, dengan hasil akurasi tertinggi sebesar 80% pada skenario pelatihan 80% dan pengujian 20%, dan hasil sentimen positif sebesar 100%. Proses *preprocessing data* serta penerapan metode TF-IDF dalam ekstraksi fitur memberikan kontribusi signifikan terhadap performa klasifikasi. Dengan demikian, pendekatan analisis sentimen ini dapat dimanfaatkan sebagai alat bantu evaluasi dalam meningkatkan kualitas layanan wisata, serta memiliki potensi untuk diterapkan pada berbagai studi serupa dalam bidang pariwisata dan pelayanan publik.

DAFTAR PUSTAKA

- [1] Listiyana Syafitri Daulay, Rahmanita Ginting, and Arifin Saleh, "Komunikasi Pariwisata Pihak Pemerintah, Pengelola, Dan Masyarakat Dalam Mengembangkan Pariwisata Berbasis Kearifan Lokal di Kota Padangsidempuan," *Talent. Conf. Ser. Local Wisdom, Soc. Arts*, vol. 3, no. 1, 2020, doi: 10.32734/lwsa.v3i1.809.
- [2] C. S. Akasse and R. Ramansyah, "Strategi Promosi Pariwisata melalui Media Sosial dalam Meningkatkan Pengunjung di Desa Wisata," *J. Socius J. Sociol. Res. Educ.*, vol. 10, no. 1, pp. 52–60, 2023, doi: 10.24036/scs.v10i1.457.
- [3] D. E. Pangkey, L. E. Tulung, and L. J. H. Lotulung, "Pemanfaatan Instagram Sebagai Media Promosi Desa Wisata di Provinsi Bali," *J. Ilmu Pemerintahan, Adm. Publik, dan Ilmu Komun.*, vol. 5, no. 1, pp. 110–120, 2023, doi: 10.31289/jipikom.v5i1.1584.
- [4] V. V. Sinuhaji, N. Siti, S. Siregar, and B. Jamil, "Aktivitas Komunikasi Pemasaran Dinas Pariwisata dan Kebudayaan Kabupaten Karo Dalam Meningkatkan Kunjungan Wisatawan (Studi Deskriptif Kualitatif Wisata Bukit Gundaling Berastagi) Marketing Communication Activities of the Karo District Tourism and Cultu," *J. Ilmu Pemerintahan, Adm. Publik dan Ilmu Komun.*, vol. 1, no. 2, pp. 105–118, 2019.
- [5] R. Rudiansyah and T. S. Sijabat, "Pengaruh Budaya Tionghoa Terhadap Kuliner Di Kota Medan," *J. Cakrawala Mandarin*, vol. 6, no. 2, pp. 486–501, 2022.
- [6] D. Ramadhani, *Pariwisata Tjong A Fie Mansion di Kota Medan Pada Era Pandemi COVID-19 (Studi Etnografi Virtual)*. Lhokseumawe: Universitas Malikussaleh, 2023.
- [7] I. Erwani, "Philology Studies at the Tjong A Fie Mansion Museum," in *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, 2020, pp. 83–89.
- [8] M. R. Darmansyah, A. H. Lubis, and M. I. Hisham, *The Sentiment Analysis Utilization for Indonesian SMEs*. 2024.
- [9] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, pp. 697–711, 2021.
- [10] F. Muharram, "Analisis Sentimen Pengguna Twitter Terhadap Kinerja Walikota Medan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf. dan Ilmu Komput.*, pp. 01–12, 2023.

- [11] M. Hafid, N. Firmansyah, and S. N. Budiman, "Sistem Rekomendasi Konversi Mata Kuliah Berbasis NLP Menggunakan Metode BERT Di Universitas Islam Balitar," vol. 4, pp. 674–682, 2025.
- [12] A. Rozaqi, A. Triayudi, and R. T. Aldisa, "Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 184, 2022, doi: 10.30865/json.v4i1.4907.
- [13] A. Riyani, M. Zidny Nafan, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," *Jlk*, vol. 2, no. 1, pp. 23–27, 2019.
- [14] Y. Nurdiansyah, F. Rahman, and P. Pandunata, "Analisis Sentimen Opini Publik Terhadap Undang-Undang Cipta Kerja pada Twitter Menggunakan Metode Naive Bayes Classifier," *Pros. Semin. Nas. Sains Teknol. dan Inov. Indones.*, vol. 3, no. November, pp. 201–212, 2021, doi: 10.54706/senastindo.v3.2021.158.
- [15] F. F. Irfani, M. Triyanto, A. D. Hartanto, and Kusnawi, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [16] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [17] F. Novianti and K. R. N. Wardani, "Analisis Sentimen Masyarakat Terhadap Data Tweet Traveloka Selama Rapid Test Antigen Menggunakan Algoritma Naïve Bayes," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 3, pp. 922–933, 2023, doi: 10.29100/jipi.v8i3.3973.
- [18] F. H. Pasaribu, N. Khairina, D. Noviantri, S. Susilawati, and R. Syah, "Analysis of The Multilayer Perceptron Algorithm on Twitter User's Sentiment Towards The COVID-19 Vaccine," *J. Informatics Telecommun. Eng.*, vol. 7, no. 1, pp. 155–163, 2023, doi: 10.31289/jite.v7i1.9664.
- [19] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *J. Inf. Syst. Res.*, pp. 1136–1144, 2023.
- [20] E. D. P. Daulay and I. Asror, "Sentimen Analisis Pada Ulasan Google Play Store Menggunakan Metode Naive Bayes," *E-Proceeding Eng.*, p. 8400, 2020.
- [21] R. Syafaat Amardita and M. Dwifebri Purbolaksono, "Analisis Sentimen Terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *J. Ris, Komput.*, pp. 2407–389, 2022.

