



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 Siregar¹⁾ *, Andre Hasudungan Lubis²⁾

1)Prodi Teknik Informatika, Fakultas Teknik, Universitas Medan Area, Indonesia

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

Tjong A Fie Mansion merupakan rumah berlantai dua mulai dibuka untuk dikunjungi umum sejak tahun 2009 lalu. Tjong A Fie Mansion mempunyai pengunjung yang jumlahnya dapat mencatat hingga lebih dari 1000 orang pengunjung setiap bulannya bahkan bisa mencapai 2000 pengunjung tiap bulannya, paling efektif 10 pengunjung situs dikunjungi pada hari libur (*weekend*) dan paling efektif 2-5 pengunjung pada hari biasa (*weekdays*). Beberapa cabang olahraga terpaksa dihentikan untuk mengurangi massa yang menumpuk, maka Museum Tjong A Fie Mansion ditutup selama 4 bulan karena mengurangi mobilitas wisatawan (Ramadhani, 2023).

Tjong A Fie adalah seorang China perantauan yang memiliki banyak harta di Medan, Jakarta, serta Singapura. Sebagai bukti lain adalah terdapat sebuah rumah peninggalan seorang saudagar kenamaan Tjong A Fie (1860-1921). Rumah peninggalan beliau hingga kini masih kokoh berdiri dan berada di jalan Kesawan, kawasan yang hingga kini juga dikenal sebagai salah satu sentra perdagangan terbesar di Medan (Sijabat & Rudiansyah, 2022).

Perkembangan teknologi akan memberikan kemudahan bagi masyarakat dalam menggunakan media sosial sebagai salah satu sarana untuk mengekspresikan pendapat ataupun sentimen. Sentimen akan diklasifikasikan menjadi label positif dan negatif dengan algoritma *Naïve Bayes*. Tahapan dalam melakukan analisis sentimen pada penelitian ini adalah *preprocessing data*, pengolahan data, klasifikasi, dan evaluasi (Muharram, 2023).

Menurut Dwi Normawati dan Surya Allit Prayogi, sistem analisis sentimen untuk mengetahui pendapat atau opini masyarakat dapat diuji sebagai studi kasus untuk kemudahan dalam pemahaman implementasi dan perhitungan secara terstruktur dengan metode NBC pada penelitian ini menghasilkan klasifikasi yang baik dengan nilai akurasi yang tinggi sehingga cocok untuk memprediksi tingkat sentimen analisis (Normawati & Prayogi, 2021). Analisis sentimen merupakan proses pemrosesan bahasa alami yang digunakan untuk mengidentifikasi dan mengelompokkan opini publik berdasarkan polaritasnya, seperti positif, negatif, atau netral, yang umumnya diterapkan pada data dari media sosial atau ulasan (Darmansyah dkk., 2024).

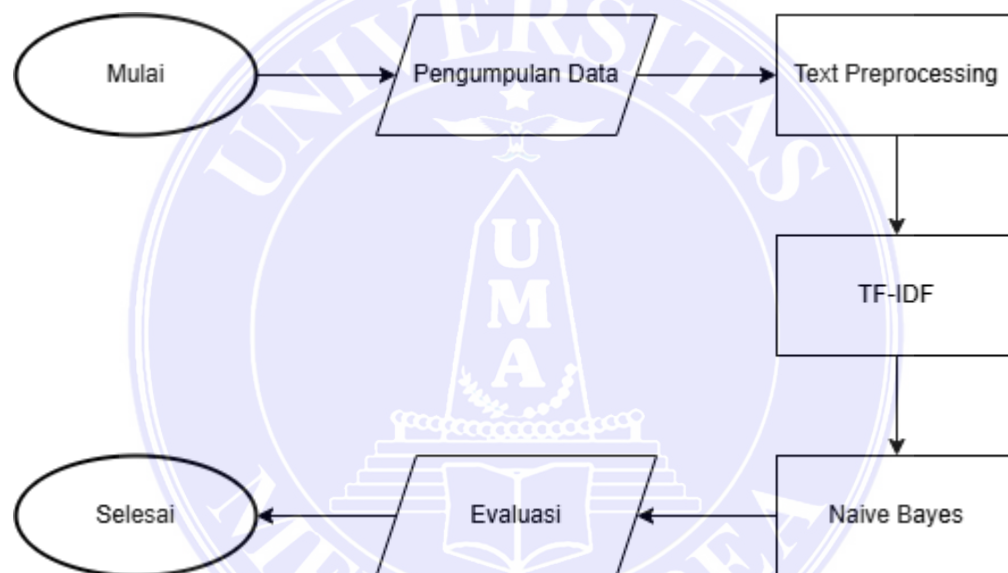
Berdasarkan deskripsi diatas maka penelitian ini mengangkat sebuah judul “Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Tjong A Fie Mansion



Menggunakan Metode *Naïve Bayes Classifier*". Pembahasan penelitian ini agar analisis sentimen pengunjung terhadap tempat wisata dapat bermanfaat bagi masyarakat. Penting analisis sentimen pada tempat wisata Tjong A Fie Mansion yaitu memahami persepsi pengujung, meningkatkan pengambilan keputusan, meningkatkan daya saing wisata, dan mendukung strategi promosi digital.

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*) (Hafid dkk., 2025). 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 (Rozaqi dkk., 2022). *Tokenisasi* – Memecah teks menjadi unit-unit yang lebih kecil, seperti kata atau kalimat (Riyani dkk., 2019). *Stopword Removal* – Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "dan", "di", "yang", dll (Nurdiansyah dkk., 2021). Proses ini bertujuan untuk menghapus data yang tidak utuh, data yang mengandung kesalahan, dan kata kata yang tidak relevan (Irfani dkk., 2020). *Stemming* – Mengubah kata ke bentuk dasarnya, misalnya "berlari", "lari-lari", menjadi "lari" (Herlinawati dkk., 2020).

- 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 (Novianti & Wardani, 2023). Adapun rumus TF-IDF adalah sebagai berikut (Pasaribu dkk., 2023):

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

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 (Rizkina & Hasan, 2023). 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 (Daulay & Asror, 2020):

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

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 $P(w_i|c)$ 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 (Normawati & Prayogi, 2021). Tujuan utama dari



confusion matrix adalah untuk menilai performa atau akurasi dari suatu sistem klasifikasi dalam mengklasifikasikan data uji (Syafaat Amardita & Dwifebri Purbolaksono, 2022). 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*:

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

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

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

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



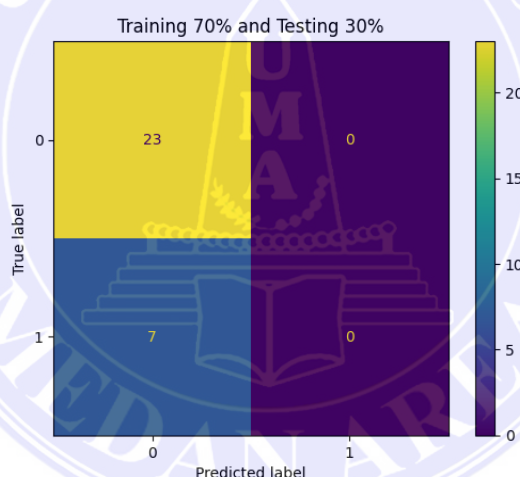
HASIL DAN PEMBAHASAN

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

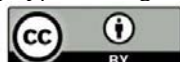
Pada Gambar 3 merupakan hasil *confusion matrix* dari pengujian 1 dengan jumlah *data training* 70% dan *testing* 30%. Berikut adalah untuk menghitung akurasi:

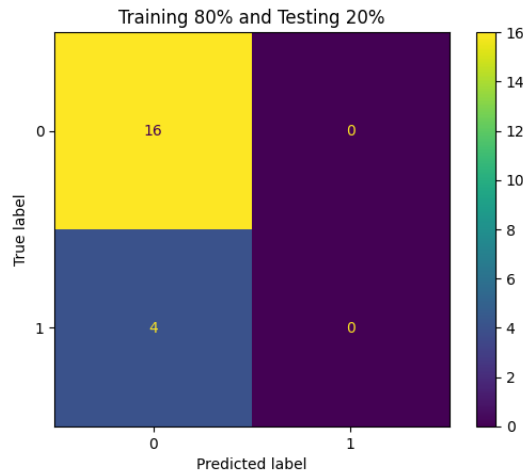
$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 23}{0 + 23 + 0 + 7} = \frac{23}{30} = 0.7667 \text{ atau } 76.67\%$$

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

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

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





Gambar 4 *Confusion Matrix* Pengujian 2

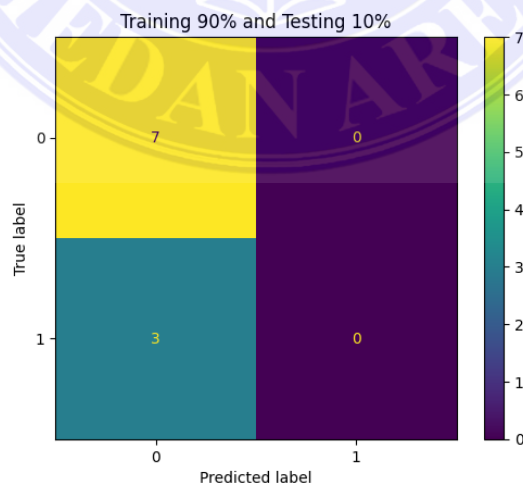
Pada Gambar 4 merupakan hasil *confusion matrix* dari pengujian 2 dengan jumlah *data training* 80% dan *testing* 20%. Berikut adalah untuk menghitung akurasi:

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

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

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

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



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:

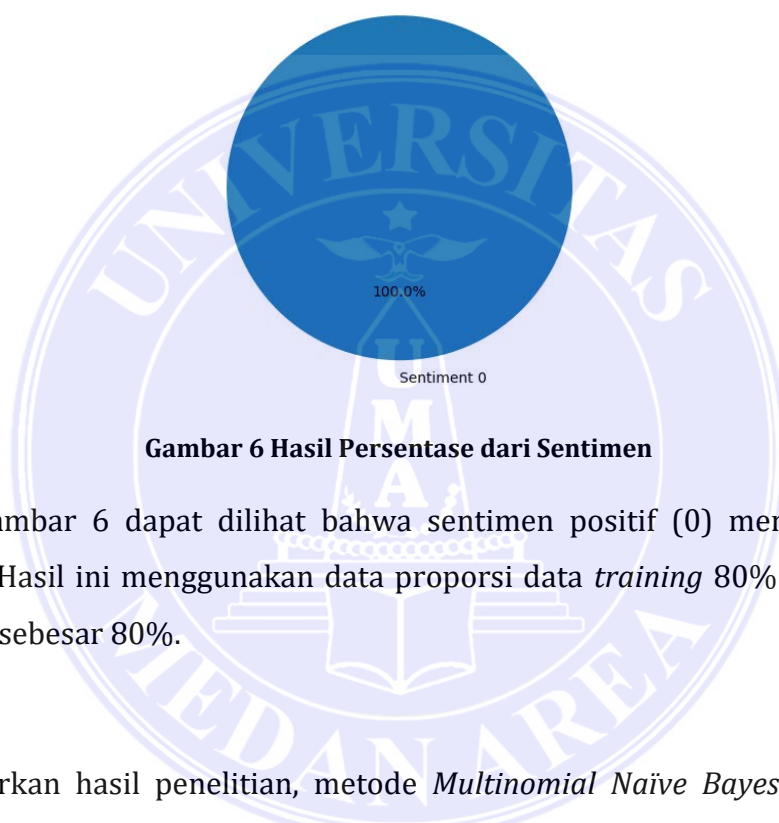
$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 7}{0 + 7 + 0 + 3} = \frac{7}{10} = 0.70 \text{ atau } 70\%$$

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

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

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

Distribution of Sentiments



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

- Darmansyah, M. R., Lubis, A. H., & Hisham, M. I. (2024). *The Sentiment Analysis Utilization for Indonesian SMEs*.
- Daulay, E. D. P., & Asror, I. (2020). Sentimen Analisis Pada Ulasan Google Play Store Menggunakan Metode Naive Bayes. *E-Proceeding of Engineering*, 8400.
- Hafid, M., Firmansyah, N., & Budiman, S. N. (2025). *Sistem Rekomendasi Konversi Mata Kuliah Berbasis NLP Menggunakan Metode BERT Di Universitas Islam Balitar*. 4, 674–682.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>
- Irfani, F. F., Triyanto, M., Hartanto, A. D., & Kusnawi. (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine. *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, 16(3), 258–266. <https://doi.org/10.26487/jbmi.v16i3.8607>
- Muharram, F. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kinerja Walikota Medan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi dan Ilmu Komputer*, 01–12.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika*, 697–711.
- Novianti, F., & Wardani, K. R. N. (2023). Analisis Sentimen Masyarakat Terhadap Data Tweet Traveloka Selama Rapid Test Antigen Menggunakan Algoritma Naïve Bayes. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(3), 922–933. <https://doi.org/10.29100/jupi.v8i3.3973>
- Nurdiansyah, Y., Rahman, F., & Pandunata, P. (2021). Analisis Sentimen Opini Publik Terhadap Undang-Undang Cipta Kerja pada Twitter Menggunakan Metode Naive Bayes Classifier. *Prosiding Seminar Nasional Sains Teknologi dan Inovasi Indonesia (SENASTINDO)*, 3(November), 201–212. <https://doi.org/10.54706/senastindo.v3.2021.158>
- Pasaribu, F. H., Khairina, N., Noviandri, D., Susilawati., & Syah, R. (2023). Analysis of The Multilayer Perceptron Algorithm on Twitter User's Sentiment Towards The COVID-19 Vaccine. *Journal of Informatics and Telecommunication Engineering*, 155–163.
- Ramadhani, D. (2023). *Pariwisata Tjong A Fie Mansion di Kota Medan Pada Era Pandemi COVID-19 (Studi Etnografi Virtual)*. Universitas Malikussaleh.



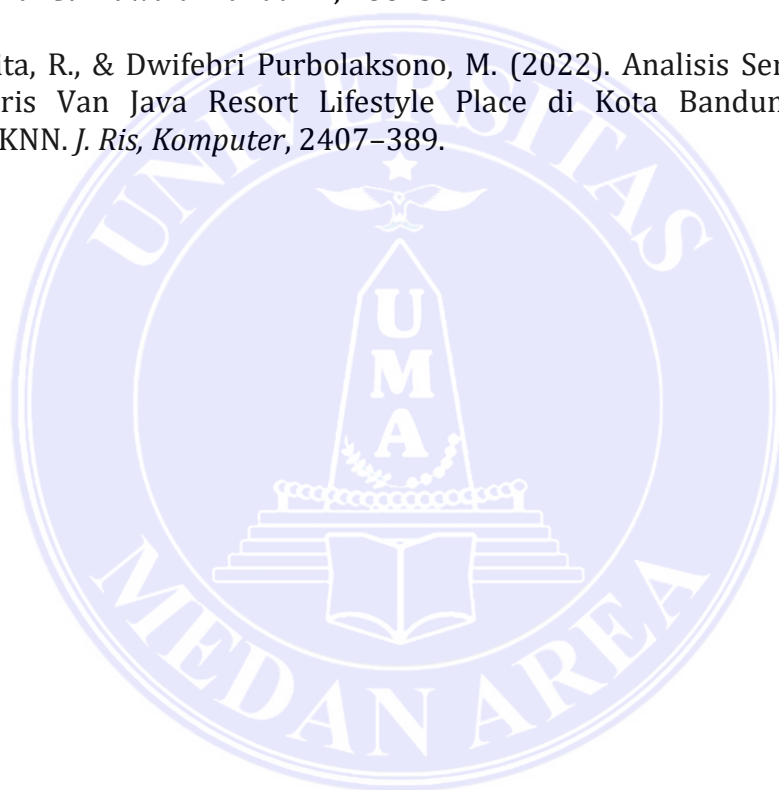
Riyani, A., Zidny Naf'an, M., & Burhanuddin, A. (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen. *Jlk*, 2(1), 23–27.

Rizkina, N. Q., & Hasan, F. N. (2023). Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes. *Jurnal of Information System Research (JOSH)*, 1136–1144.

Rozaqi, A., Triayudi, A., & Aldisa, R. T. (2022). Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(1), 184. <https://doi.org/10.30865/json.v4i1.4907>

Sijabat, T. S., & Rudiansyah. (2022). Pengaruh Budaya Tionghoa Terhadap Kuliner di Kota Medan. *Jurnal Cakrawala Mandarin*, 486–501.

Syafaat Amardita, R., & Dwifabri Purbolaksono, M. (2022). Analisis Sentimen Terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN. *J. Ris, Komputer*, 2407–389.



Date: April 21st, 2025

LETTER OF ACCEPTANCE

Paper Number #848

Dear, Erlina Siregar & Andre Hasudungan Lubisay

This is to inform you that the manuscript entitled: **“Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Tjong A Fie Mansion Menggunakan Metode Naïve Bayes Classifier”**, which was sent on **March 19th, 2025**, is **ACCEPTED**.

We keep to ensuring a high standard of articles published in the **INCODING: Journal of Informatics and Computer Science Engineering**, and the manuscript that is being sent to you has been submitted after a first selection process based on the agreement of the **Associate Editors**. In general, the standard of manuscripts forwarded to me after the vetting is **good**.

This paper is well organized and follows the manuscript guidelines of the journal to a large extent. The introduction section is good and shows the importance of the study. The literature review is adequate. The outcomes of the study are consistent with the findings. The approach used is praiseworthy. In my opinion, it should be published without **revision again**.

Based on the review results, this manuscript is **ACCEPTED**, and **PUBLISHED** in **Mei 2025** for **Volume 3, No. 2, 2025**.

Thank you very much for your contribution. Congratulations on a wonderful job.

Warmest Regards,
Editor In Chief

INCODING
2776-432X (Online - Elektronik)

Agung Suharyanto, S.Sn, M.Si.

Editorial Office:
Mahesa Research Center

INCODING: Journal of Informatics and Computer Science Engineering

UNIVERSITAS MEDAN AREA
Perumahan Griya Nafisa 2 Blok A No 10, Jalan Benteng Hilir
Titi Sewa, RT. 06, Dusun XVI Flamboyan,
Kecamatan Percut Sei Tuan, Deli Serdang, 20371
Sumatera Utara, Indonesia
Phone: 08126493527
Email: incodingjournal@gmail.com

© Hak Cipta © Incoding Journal Online

1. Dilarang Menyalin atau mendistribusikan dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area



Document Accepted 23/5/25