

# JITE (Journal of Informatics and Telecommunication Engineering)

Available online <http://ojs.uma.ac.id/index.php/jite> DOI: 10.31289/jite.vxix.xxx



Received: dd-mm-yyyy

Accepted: dd-mm-yyyy

Published: dd-mm-yyyy

## Klasifikasi Wisatawan Berdasarkan Perilaku Konsumen Dengan Menggunakan Algoritma XGBoost

Jenius zalukhu<sup>1</sup>), Andre Hasudungan Lubis<sup>2</sup>)  
Fakultas Teknik, Universitas Medan Area, Indonesia

[Jeniuszal014@gmail.com](mailto:Jeniuszal014@gmail.com), [andrelubis2201@gmail.com](mailto:andrelubis2201@gmail.com)

Abstrak [Font: Cambria, size, 9, Bold]

Penelitian ini bertujuan pembahasan penerapan Algoritma XGBosst terhadap Wisatawan berdasarkan perilaku konsumen. Tujuan penelitian ini adalah untuk memprediksi atau menganalisis data ulasan wisatawan, serta membantu menyediakan dan memahami kebutuhan sehingga meningkatkan kualitas layanan yang ditawarkan. Bangsa Indonesia memiliki potensi pariwisata yang besar berkat keindahan alam dan keragaman budayanya, Sektor ini berperan penting dalam perekonomian nasional dengan menciptakan lapangan kerja dan mendorong industri kreatif serta perhotelan. Kehadiran para wisatawan meningkatkan pendapatan daerah melalui pajak dan pengeluaran di sektor-sektor seperti hotel, restoran, dan toko oleh-oleh, serta menciptakan lapangan kerja baru. Selain wisatawan dapat meningkatkan pendapatan adanya perlu pemahaman dari setiap perilaku wisatawan penting untuk pengembangan pariwisata yang adaptif dan berkelanjutan.

**Kata Kunci:** Extreme Gradient Boosting ( XGBosst ), Prediksi Wisatawan, Klasifikasi, Machine Learning

**Abstract:** This study discusses the application of the XGBosst Algorithm to Tourists based on consumer behavior. The purpose of this study is to predict or analyze tourist review data, and to help provide and understand needs so as to improve the quality of services offered. Indonesia has great tourism potential thanks to its natural beauty and cultural diversity. This sector plays an important role in the national economy by creating jobs and encouraging the creative industry and hospitality. The presence of tourists increases regional income through taxes and spending in sectors such as hotels, restaurants, and souvenir shops, as well as creating new jobs. In addition to tourists being able to increase income, there is a need for an understanding of each tourist behavior that is important for the development of adaptive and sustainable tourism..

**Keywords:** Extreme Gradient Boosting (XGBosst), Tourist Prediction, Classification, Machine Learning

**How to Cite:** Pertama, N.P. Pertama, P. & Ketiga, P. (2021). Judul Hendaknya Ringkas dan Informatif Tidak Lebih dari 15 Kata. *JITE (Journal Of Informatics And Telecommunication Engineering)*. 4 (2): 1-10

## I. PENDAHULUAN

Indonesia merupakan salah satu negara dengan potensi pariwisata yang sangat besar, didukung oleh keindahan alam dan keanekaragaman budaya (Damayanti & Puspitasari, 2024). Sektor pariwisata menjadi salah satu kontributor utama bagi perekonomian nasional, dengan memberikan lapangan pekerjaan dan mendorong pertumbuhan industri kreatif serta perhotelan (Hasibuan et al., 2023). Dengan meningkatnya kebutuhan akan aksesibilitas dan kenyamanan bagi wisatawan, sektor pariwisata mendorong pemerintah untuk mengembangkan infrastruktur, seperti jalan, bandara, dan fasilitas publik, yang juga bermanfaat bagi masyarakat lokal (Pratami et al., 2021). Besarnya potensi pariwisata di Indonesia dapat memikat para

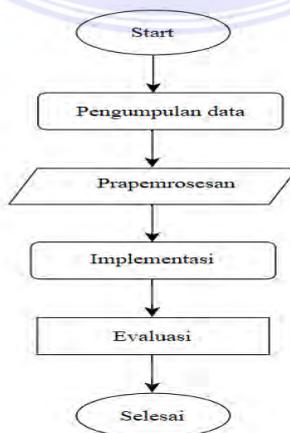
wisatawan yang memiliki peran yang sangat penting dalam memajukan dan mengembangkan industri pariwisata dan ekonomi yang ada diberbagai daerah daerah (Supriaji et al., 2024).

Kehadiran wisatawan dapat membawa dampak langsung terhadap pendapatan daerah melalui pajak dan pengeluaran di berbagai sektor, seperti hotel, restoran, dan toko oleh-oleh, Uang yang dibelanjakan oleh wisatawan tidak hanya meningkatkan pendapatan pelaku usaha lokal, tetapi juga menciptakan lapangan kerja baru (Sugiarto et al., 2023). Dengan demikian, industri pariwisata berkontribusi besar terhadap pertumbuhan ekonomi masyarakat setempat. Selain itu, wisatawan juga berfungsi sebagai duta bagi produk dan budaya lokal, memperbaiki fasilitas dan infrastruktur pariwisata, serta peningkatan perkembangan industri pariwisata (Fitriyah et al., 2025; Hasibuan et al., 2023). Perilaku wisatawan yang berkunjung pada suatu destinasi dapat berpengaruh pada industri pariwisata. Sehingga, penting untuk ditelusuri guna mendapatkan pariwisata yang berkembang lebih adaptif dan berkelanjutan. Dalam menghadapi tantangan dan peluang di sektor pariwisata, pemahaman mendalam tentang perilaku wisatawan menjadi kunci untuk pengembangan yang sukses (Permatasari, 2022).Wisatawan dapat diklasifikasikan yang bertujuan untuk mengelompokkan orang-orang yang melakukan perjalanan berdasarkan karakteristik atau perilaku tertentu (Latif et al., 2024). Dengan memahami perbedaan dalam kebiasaan dan preferensi wisatawan, pihak-pihak terkait, seperti pengusaha, agen perjalanan, dan penyedia layanan, dapat lebih mudah mengetahui kebutuhan spesifik dari setiap segmen. Proses ini memungkinkan mereka untuk merancang sebuah produk dan berbagai layanan yang lebih tepat sasaran, sehingga pengalaman yang ditawarkan menjadi lebih memuaskan bagi pengunjung. Selain itu, pemahaman mendalam tentang beragam kategori wisatawan juga berkontribusi pada pengembangan kebijakan dan destinasi yang lebih baik, yang pada akhirnya akan mendukung pertumbuhan industri pariwisata secara keseluruhan (Annisafiya et al., 2023). Untuk mengklasifikasi wisatawan secara efektif, diperlukan penerapan model algoritma XGBoost. Algoritma XGBoost adalah algoritma machine learning yang dikenal karena kemampuannya dalam menghasilkan model prediktif yang akurat dan efisien (Kurniawanda & Tobing, 2022). Dalam konteks pariwisata, XGBoost dapat digunakan untuk menganalisis data ulasan wisatawan, seperti yang terdapat pada platform TripAdvisor, guna mengidentifikasi sentimen positif atau negatif dalam ulasan tersebut. Penelitian menunjukkan bahwa XGBoost memiliki tingkat akurasi yang tinggi dalam klasifikasi sentimen, mencapai 99%, dibandingkan dengan algoritma lain seperti Decision Tree dan Support Vector Machine (Adnan, 2024). Dengan demikian, penerapan XGBoost dalam analisis sentimen ulasan wisatawan dapat membantu penyedia layanan pariwisata memahami kebutuhan dan preferensi wisatawan, sehingga dapat meningkatkan kualitas layanan yang ditawarkan.

## II. METODE PENELITIAN

### A. Tahapan Penelitian

Penelitian ini mempunyai beberapa tahapan yang harus dilalui dimulai pertama yaitu pengumpulan data sampai dengan evaluasi hasil. Berikut ini adalah Diagram alir penelitian yang menunjukkan keseluruhan proses yang akan dilakukan.



Gambar 1. Flowchart rancangan penelitian metodolog

## B. Pengumpulan Data

Dalam penelitian ini, proses pengumpulan data dilakukan melalui beberapa tahapan yang sistematis untuk memastikan kualitas dan relevansi data yang digunakan dalam klasifikasi wisatawan berdasarkan perilaku konsumen. Data dari penelitian ini diambil dari tautan berikut <https://www.kaggle.com/datasets/ddosad/customer-behaviour-tourism-portal>. Data tersebut berisi informasi tentang wisatawan berdasarkan berbagai atribut seperti perangkat yang digunakan, lokasi perjalanan, dan aktivitas interaksi mereka dengan platform perjalanan. Atribut utama mencakup User ID, jenis produk yang diambil (*Taken Product*), jumlah tampilan pada halaman perjalanan (*View on Travel*), perangkat yang digunakan (*Preferred Device*), serta jumlah komentar dan check-in di berbagai lokasi wisata. Adapun jumlah baris data sebanyak 11.760 baris dengan rangkumannya sebagai berikut.

**Tabel 1. Rangkuman Data**

User ID	Taken Product	View on Travel	Preverred device	...	Travelling Page
1000001	Yes	307	IOS And Android	...	8
1000002	No	367	IOS	...	10
1000003	Yes	277	IOS And Android	...	7
...	...	...	...	...	...
1011757	No	214	Tab	...	12
1011758	No	382	Laptop	...	20
1011759	No	270	Tab	...	14

## C. Prapemrosesan

Tahapan prapemrosesan meliputi normalisasi data, pengecekan nilai null, dan pemilihan atribut. Dalam tahapan prapemrosesan data, normalisasi dapat dilakukan menggunakan metode Min-Max Scaling untuk memastikan bahwa seluruh atribut numerik berada dalam rentang nilai yang seragam, biasanya antara 0 hingga 1. Normalisasi ini penting agar model tidak memberikan bobot berlebihan pada atribut dengan skala yang lebih besar, sehingga meningkatkan stabilitas dan akurasi dalam proses klasifikasi. Dalam penelitian ini, setiap atribut numerik dinormalisasi menggunakan rumus sebagai berikut.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Selain normalisasi, langkah penting lainnya dalam prapemrosesan adalah penanganan *missing values* atau nilai yang hilang. Data yang tidak lengkap dapat mengganggu proses analisis, sehingga dalam penelitian ini dilakukan identifikasi terhadap atribut yang memiliki nilai kosong. Jika jumlah data yang hilang kecil, metode imputasi digunakan dengan menggantinya menggunakan nilai median atau modus dari atribut terkait. Namun, jika jumlah nilai yang hilang signifikan, maka atribut tersebut dapat dihapus untuk menghindari distorsi pada hasil analisis. Setelah tahap ini, dilakukan pemilihan atribut berdasarkan relevansi terhadap klasifikasi perilaku wisatawan. Atribut yang dapat digunakan dalam penelitian ini dapat dilihat di tabel dibawah ini.

**Tabel 2. Atribut Data**

Atribut	Penggunaan
User ID	√
Taken Product	√
Yearly_avg_view_on_travel_page	√
Yearly_avg_comment_on_travel_page	√
Preferred_device	X
Total_likes_on_outstation_checkin_given	√
Yearly_avg_Outstation_checkins	X
Member_in_family	X
Preferred_location_type	X
Yearly_avg_comment_on_travel_page	√
Total_likes_on_outofstation_checkin_received	√
Week_since_last_outstation_checkin	√
Following_company_page	√
Travelling_network_rating	√
Working_flag	√
Ault_flag	√
Daily_Avg_mins_spend_on_traveling_page	√

Dari total 16 atribut yang tersedia dalam dataset, sebanyak 12 atribut digunakan dalam analisis karena dianggap relevan dalam klasifikasi wisatawan, seperti jumlah tampilan halaman perjalanan, jumlah komentar, dan rata-rata waktu yang dihabiskan pada platform perjalanan. Sementara itu, 4 atribut lainnya tidak digunakan karena dinilai kurang berkontribusi secara signifikan terhadap proses klasifikasi.

**D. Implementasi Algoritma XGBosst**

XGBosst adalah sebuah algortima yang ditingkatkan dari gradient boosting decision tree yang efisien dalam membangun boosted trees, XGBosst merupakan metode pemebelajaran mesin yang digunakan untuk menyelesaikan permasalahan regresi dan klasifikasi dengan menggunakan Gradient Boosting Decision Tree. Model Algoritma XGBoost Dapat Diimplementasikan sebagai berikut

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \tag{2}$$

di mana :

- ^i adalah prediksi akhir untuk sampel ke-iii.
- K adalah jumlah pohon keputusan dalam model.
- f<sub>k</sub>(x<sub>i</sub>) adalah pohon keputusan ke-kkk yang memprediksi nilai untuk sampel x<sub>i</sub>

### E. Confusion Matrix

Confusion matrix Merupakan sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam machine learning, Tabel tersebut dapat membandingkan hasil prediksi model dengan nilai aktual, sehingga kita dapat memahami seberapa baik model tersebut dalam mengklasifikasikan data, Metrik-metrik ini berperan membantu dalam mengidentifikasi kekuatan dan kelemahan model, serta memberikan wawasan tentang bagaimana model dapat ditingkatkan untuk performa yang lebih baik.

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{F1 Score} = \frac{2 \times \text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \tag{6}$$

## III. HASIL DAN PEMBAHASAN

### A. Hasil Prapemrosesan

#### 1. Transformasi Data

Dalam tahap prapemrosesan, transformasi data dilakukan untuk mengubah atribut kategorikal menjadi format numerik agar dapat diolah oleh algoritma machine learning. Proses ini mencakup konversi nilai biner, seperti Taken Product (Ya/Tidak), menjadi representasi numerik (1/0), serta penggunaan teknik one-hot encoding untuk atribut dengan lebih dari dua kategori, sehingga seluruh data dapat dianalisis secara lebih efisien dan akurat oleh model. Rangkuman dari hasilnya sebagai berikut

Tabel 3. Transformasi Data

User ID	Taken Product	View on Travel	...	Travelling Page
1000001	1	307	...	8
1000002	0	367	...	10
1000003	1	277	...	7
...	...	...	...	...
1011757	0	214	...	12
1011758	0	382	...	20
1011759	0	270	...	14

#### 2. Hasil Normalisasi Data

Data tersebut merupakan Pernormalisasian dari berbagai atribut, untuk Setiap baris mewakili satu pengguna dengan berbagai variabel yang telah dinormalisasi agar memiliki skala yang seragam, sehingga memudahkan analisis menggunakan algoritma machine learning. Kolom "*Taken Product*" menunjukkan apakah pengguna telah membeli suatu produk wisata, di mana nilai 1 berarti membeli dan 0 berarti tidak. Sementara itu, kolom "*View on Travel*" mencerminkan jumlah tampilan halaman perjalanan yang dinormalisasi dalam rentang tertentu. Kolom "*Travelling Page*" menunjukkan intensitas interaksi pengguna dengan halaman perjalanan yang juga telah disesuaikan skalanya. Dengan proses normalisasi ini, model dapat menganalisis data secara lebih akurat tanpa adanya bias akibat perbedaan skala antar atribut

Tabel 4. Normalisasi Data

User ID	Taken Product	View on Travel	...	Travelling Page
1000001	1	0.634033	...	0.029630
1000002	0	0.773893	...	0.037037
1000003	1	0.564103	...	0.025926
...	...	...	...	...
1011757	0	0.417249	...	0.044444
1011758	0	0.808858	...	0.074074
1011759	0	0.547786	...	0.051852

### 3. Akurasi Hasil Data

Data tersebut menunjukkan kinerja model dalam klasifikasi menggunakan metrik **Akurasi**, **Precision**, **Recall**, dan **F1-Score**. Precision dapat mengukur ketepatan prediksi positif, Recall mengukur seberapa banyak data positif yang terdeteksi, dan F1-Score menyeimbangkan keduanya. Untuk kelas pertama, model sangat akurat dengan Precision 0,99 dan Recall 1,00, artinya hampir semua prediksi positif benar dan tidak ada yang terlewat. Kelas kedua memiliki Precision 1,00 dan Recall 0,95, yang berarti semua prediksi positif benar, tetapi ada sedikit data positif yang tidak terdeteksi. Secara keseluruhan, model memiliki kinerja sangat baik dengan keseimbangan antara ketepatan dan kemampuan mendeteksi data positif.

Tabel 5. Akurasi Hasil Data

Akurasi	Precision	Recall	F1- Score
99.1%	99%	100%	99%

Confusion Matrix menunjukkan bahwa sebuah model klasifikasi memiliki kinerja yang sangat baik, dengan 1.957 prediksi benar untuk kelas 0 dan 374 prediksi benar untuk kelas 1. Namun, terdapat 20 kesalahan klasifikasi di mana kelas 1 diklasifikasikan sebagai kelas 0 dan 1 kesalahan di mana kelas 0 diklasifikasikan sebagai kelas 1.



Gambar 2. Confusion Matrix

## F. KESIMPULAN

Industri pariwisata di Indonesia memiliki peran penting dalam mendorong pertumbuhan ekonomi melalui penciptaan lapangan kerja, peningkatan sektor perhotelan, serta pengembangan infrastruktur yang juga bermanfaat bagi masyarakat lokal. Wisatawan yang berkunjung tidak hanya memberikan dampak ekonomi melalui pengeluaran mereka di berbagai sektor, tetapi juga berkontribusi dalam promosi budaya dan produk lokal. Oleh karena itu, pemahaman mengenai perilaku wisatawan menjadi faktor kunci dalam pengembangan industri pariwisata yang lebih adaptif dan berkelanjutan. Dengan mengelompokkan wisatawan berdasarkan karakteristik dan kebiasaan mereka, para pelaku industri, seperti pengusaha dan penyedia layanan, dapat lebih mudah menyesuaikan strategi pemasaran dan meningkatkan kualitas pengalaman wisatawan. Penelitian ini menggunakan data dari platform perjalanan untuk mengidentifikasi pola perilaku wisatawan. Data yang dikumpulkan mencakup atribut seperti jumlah tampilan halaman perjalanan, jumlah komentar yang diberikan, serta perangkat yang digunakan. Sebelum dianalisis, data melalui tahap prapemrosesan yang mencakup normalisasi dengan metode Min-Max Scaling untuk memastikan skala yang seragam, serta penanganan nilai yang hilang agar tidak mengganggu hasil analisis. Dari total 16 atribut dalam dataset, sebanyak 12 atribut dipilih karena memiliki kontribusi yang lebih signifikan dalam proses klasifikasi wisatawan. Dengan langkah-langkah ini, penelitian memastikan bahwa model yang digunakan memiliki performa yang optimal dalam mengklasifikasikan wisatawan berdasarkan pola perilaku mereka.

Implementasi algoritma XGBoost dalam penelitian ini menunjukkan hasil yang sangat baik dalam mengklasifikasikan wisatawan dengan tingkat akurasi tinggi. Model ini mampu menganalisis ulasan wisatawan secara efektif dan memberikan hasil klasifikasi yang lebih akurat dibandingkan dengan algoritma lain, seperti Decision Tree dan Support Vector Machine. Evaluasi model menggunakan confusion matrix menunjukkan bahwa XGBoost memiliki presisi dan recall yang hampir sempurna, yang berarti model dapat mendeteksi dan mengelompokkan wisatawan secara akurat. Dengan penerapan metode ini, penyedia layanan pariwisata dapat lebih memahami preferensi wisatawan, meningkatkan kualitas layanan, serta merancang strategi yang lebih efektif dalam mengembangkan destinasi wisata yang kompetitif dan berkelanjutan.

## DAFTAR PUSTAKA

- Adnan, S. A. R. P. (2024). *Analisis Sentimen Terhadap Objek Wisata di Kabupaten Ngada Menggunakan Metode XGBoost*. Universitas Muhammadiyah Malang.
- Annisafiya, N., Kusuma, D. A., & Ruchjana, B. N. (2023). Penerapan Analisis Diskriminan untuk Klasifikasi Pengaruh Data Warisan Budaya Takbenda terhadap Banyaknya Wisatawan Domestik. *Jurnal Matematika Integratif*, 19(2), 149-161.
- Damayanti, R. A., & Puspitasari, A. Y. (2024). Kajian potensi daya tarik wisata heritage di Indonesia. *Jurnal Kajian Ruang*, 4(1), 13-34.
- Fitriyah, S. N. L., Sari, D. K., & Hariasih, M. (2025). Peran Digital Marketing, Daya Tarik, dan EWOM Dalam Meningkatkan Minat Berkunjung ke Taman Ghanjaran Desa Wisata Ketapanrame Trawas. *Journal of Economic, Management, Accounting and Technology*, 8(1), 169-186.
- Hasibuan, I. M., Mutthaqin, S., Erianto, R., & Harahap, I. (2023). Kontribusi sektor pariwisata terhadap perekonomian nasional. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, 8(2).
- Kurniawanda, M. R., & Tobing, F. A. T. (2022). Analysis sentiment cyberbullying in instagram comments with xgboost method. *IJNMT (International Journal of New Media Technology)*, 9(1), 28-34.
- Latif, N., Tempola, F., Lutfi, S., Kapita, S. N., & Rosihan, R. (2024). Klasifikasi Destinasi Wisata Terbaik Berdasarkan Pilihan Pengunjung Menggunakan Metode Naive Bayes. *Jurnal Dialektika Informatika (Detika)*, 5(1), 45-49.
- Permatasari, I. (2022). Peran model pengembangan pariwisata berbasis masyarakat (Community based tourism) dalam mewujudkan pariwisata berkelanjutan (Sustainable tourism) di Bali. *Kertha Wicaksana*, 16(2), 164-171.
- Pratami, M., Gunari, B. F., & Rilansari, V. (2021). Strategi Pengoptimalan Infrastruktur Pendukung Pariwisata Berkelanjutan. *Jurnal Destinasi Pariwisata*, 9(2), 329-336.
- Sugiarto, A., Manalu, S. P. R., & Pakpahan, E. (2023). Pengaruh jumlah kunjungan wisatawan dan pajak restoran terhadap pertumbuhan ekonomi kabupaten tapanuli utara dengan PAD Sebagai Variabel Intervening. *Jesya (Jurnal Ekonomi Dan Ekonomi Syariah)*, 6(1), 221-232.

- Supriaji, U., Dirga, D. M., Yakin, A., & others. (2024). *HALAL TOURISM (KONSEP PARIWISATA HALAL DI INDONESIA) Menjawab Kebutuhan Wisata yang Nyaman dan Terjamin Kehalalannya bagi Wisatawan Muslim Dunia yang Berkunjung di Indonesia*. PT. Penerbit Qriset Indonesia.
- World Tourism Organization, International Tourism: A Global Perspective. World Tourism Organization (UNWTO), 1997.
- Dennis L. Foster, Sales and Marketing For The Travel Profesional. Jakarta: Raja Grafindo Persada, 2002.
- Reza, V. (2020). Pariwisata Halal Dalam Pengembangan Ekonomi Indonesia. *Jurnal An-Nahl*, 7
- Cherif, I. L., & Kortebi, A. (2019). On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification. *IFIP Wireless*  
<https://doi.org/10.1109/WD.2019.8734193> Day
- Delima, R., Hosianna, M., Pebrianty, D., Amalia, J., & Analysis, A. C. R. (2023). Credit Risk Analysis dengan Algoritma Extreme Gradient Boosting dan Adaptive Boosting.
- Herni Yulianti, S. E., Oni Soesanto, & Yuana Sukmawaty. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics Theory and Application*, 4(1), 21–26. <https://doi.org/10.31605/jomta.v4i1.1792>

