

## Penerapan Algoritma XGBoost Dalam Menganalisa Keberlanjutan Pelanggan Tour dan Travel

Ratih Yulia Hayuningtyas<sup>1</sup>, Wina Yusnaeni<sup>2</sup>, Ida Darwati<sup>3</sup>, Syifa Tania<sup>4</sup>, Harko Abditama<sup>5</sup>

<sup>1,2,3,4,5</sup> Sistem Informasi, Universitas Bina Sarana Informatika  
Indonesia

\* Corresponding Author. E-mail: [ratih.ryl@bsi.ac.id](mailto:ratih.ryl@bsi.ac.id)

### Abstrak

Pelanggan churn salah satu istilah yang digunakan untuk menyebut kehilangan pelanggan dalam dunia bisnis. Pelanggan churn merupakan tantangan terbesar dalam dunia bisnis, karena memiliki dampak bagi setiap perusahaan. Salah satunya adalah usaha tour dan travel, dampak yang dihasilkan dari pelanggan churn pada tour dan travel dapat berupa menurunnya laba, biaya operasional meningkat karena mencari pelanggan tour dan travel membutuhkan biaya besar dibandingkan dengan mempertahankan pelanggan. Setiap Perusahaan memiliki strategi untuk mempertahankan pelanggan salah satunya dengan cara menerapkan *machine learning*. Pada penelitian ini menggunakan data publik untuk menentukan pelanggan churn dengan menggunakan algoritma XGBoost. *Extreme Gradient Boost* (XGBoost) yang bekerja dengan cara membangun model secara bertahap untuk meningkatkan akurasi dari prediksi. Dalam penelitian ini model XGBoost bekerja melalui beberapa tahapan yaitu pemrosesan data, pembagian dataset, secara pengujian dengan algoritma yang nantinya akan menghasilkan akurasi dari sebuah model, evaluasi dan kurva ROC dan AUC. Hasil dari penelitian ini dengan model XGBoost menghasilkan *accuracy* sebesar 87,7%, *precision* 74,4%, *recall* 74,4% , *F1-Score* 74,4%, dan nilai AUC 0,95 selain itu dalam penelitian ini juga menghasilkan sebuah aplikasi untuk memprediksi pelanggan yang belum memiliki sebuah label.

### Abstract

*Customer churn is a term used to describe customer loss in the business world. Customer churn is a major challenge in the business world, impacting every company. One example is the tour and travel industry, the impact of customer churn in tour and travel businesses can include decreased profits and increased operational costs because acquiring tour and travel customers is more expensive than retaining them. Every company has a strategy for customer retention, one of which is implementing machine learning. This study uses public data to determine customer churn using the XGBoost algorithm. Extreme Gradient Boost (XGBoost) works by gradually building a model to improve prediction accuracy. In this study, the XGBoost model works through several stages: data processing, dataset division, algorithm testing, which ultimately results in model accuracy, evaluation, and ROC and AUC curves. The results of this study with the XGBoost model produced an accuracy of 87.7%, precision of 74.4%, recall of 74.4%, F1-Score of 74.4%, and an AUC value of 0.95. In addition, this study also produced an application to predict customers who do not yet have a label.*

**Keywords:** XGBoost; Prediksi; Algoritma; Prediksi Churn; Pelanggan

### 1. Introduction

Pelanggan merupakan asset terpenting dalam sebuah usaha tour dan travel, untuk mempertahankan pelanggan dalam menggunakan tour dan travel merupakan

salah satu tantangan terbesar demi keberlanjutan perusahaan. Bertahan dengan persaingan (Budiarno et al., 2022) antar penyedia jasa tour dan travel yang lain merupakan tantangan yang harus dilewati

karena semakin maraknya perusahaan yang bergerak dalam bidang yang sama (Prasetyo et al., 2023). Fasilitas yang baik, nyaman serta memadai juga dapat menjadi salah satu alasan seorang pelanggan terhadap perusahaan sehingga memutuskan untuk kembali menggunakan jasanya.

Masalah prediksi churn atau menganalisa keberlanjutan pelanggan telah muncul sebagai subjek penting dalam penambangan data, karena dapat memberikan hasil pelanggan yang berisiko churn bagi Perusahaan (B. A. Maulana & Hidayati, 2025). Dampak dari pelanggan churn dalam dunia bisnis bisa mempengaruhi laba dari perusahaan dan biaya operasional, selain itu mendapatkan pelanggan baru membutuhkan biaya lebih besar dibandingkan mempertahankan pelanggan yang sudah ada (Rizki Kurniawan et al., 2024). Setiap Perusahaan memiliki strategi untuk mempertahankan pelanggan dan setiap perusahaan harus mengetahui potensi pelanggan yang memiliki kecenderungan churn.

Dalam penelitian ini menerapkan *machine learning* untuk membantu melihat pelanggan yang memiliki kecenderungan churn pada tour dan travel menggunakan algoritma XGBoost. *eXtreme Gradient Boost* (XGBoost) dapat menangani masalah *machine learning* dengan skala besar, karena memiliki beberapa fitur yang berguna untuk mempercepat perhitungan, selain itu XGBoost juga dapat mengatasi masalah klasifikasi karena memiliki kemampuan yang fleksibilitas (Yulianti et al., 2022).

Penelitian (Yulianti et al., 2022) menggunakan algoritma XGBoost untuk membuat pengelompokan nasabah kartu kredit yang macet, dihasilkan model penelitian yang cukup baik dengan akurasi sebesar 80,02%, presisi 85,32%, recall

94,86% dan masuk kedalam *good classification*.

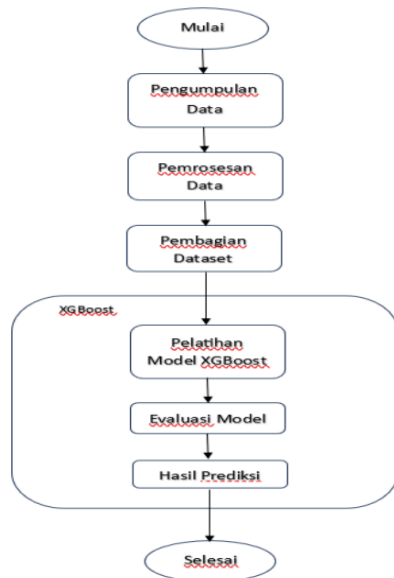
Tahun 2022, penelitian oleh Govari, (Givari et al., 2022) membanding algoritma SVM, Random Forest dan XGBoost dalam menentukan persetujuan kartu kredit, hasil penelitian menunjukkan bahwa algoritma XGBoost pada penelitian ini mendapatkan nilai akurasi, recall dan presisi dengan hasil tertinggi.

Penelitian terkait tahun 2023 (M. D. Maulana et al., 2024) menerapkan XGBoost dalam mengklasifikasikan kualitas air minum, dengan membagi data latih 80% dan data uji 20% mendapatkan hasil akurasi 82,29%, presisi 78,62%, recall 85,90%.

Berdasarkan latarbelakang ini, maka tujuan dari penelitian ini menerapkan algoritma XGBoost pada data tour dan travel untuk membangun sebuah model untuk menganalisa keberlanjutan pelanggan atau prediksi churn dengan *machine learning*. Sehingga setiap usaha dapat meningkatkan kembali kualitas dan mengurangi kecenderungan churn. Selanjutnya mengetahui seberapa besar akurasi dari proses prediksi churn menggunakan algoritma XGBoost.

## 2. Materials and Methods

Pada penelitian ini menggambarkan tahapan dari penelitian yang digunakan untuk mengetahui langkah-langkah penelitian yang dilakukan. Berikut Adalah gambar tahapan penelitian.



Gambar 1. Diagram Alir Penelitian

a. Pengumpulan Data

Data diambil dari data publik yang di unduh dari <https://www.kaggle.com/datasets/tejashvi14/tour-travels-customer-churn-prediction> (Tejashvi, 2024). Data sebanyak 954 data, dengan 6 atribut dan 1 label.

Tabel 1. Atribut Data

Atribut	Keterangan
<i>Age</i>	Usia Pengguna
<i>FrequentFlyer</i>	Pelanggan sering melakukan penerbangan (Yes, No)
<i>AnnualIncomeClass</i>	Pendapatan tahunan pengguna (Low Income, Middle Income, High Income)
<i>ServicesOpted</i>	Jumlah layanan yang dipilih (1-6)
<i>AccountSynced ToSocialMedia</i>	Akun perusahaan disinkronkan dengan media social (Yes, No)
<i>BookedHotelOrNot</i>	Pelanggan memesan hotel menggunakan layanan perusahaan (Yes, No)
<i>Target</i>	Kelas (1,0) 0-Tidak Churn (Tidak berhenti berlangganan) 1-Churn (Berhenti berlangganan)

b. Pemrosesan Data

Mengubah data menjadi data yang lebih mudah untuk diproses menggunakan metode XGBoost (Jange, 2022). Pada penelitian ini pemrosesan data dengan cara menghilangkan data yang kosong, mengubah data kategorikal menjadi format numerik untuk memudahkan pada saat melatih metode.

c. Pembagian Dataset

Dataset yang akan digunakan untuk pemrosesan di bagi menjadi data latih yaitu 80% dan data uji 20%. Data ini nantinya akan dimasukkan kedalam XGBoost untuk mendapatkan model XGBoost yang akan digunakan untuk melakukan prediksi.

Pembagian dataset ini bertujuan untuk menguji kemampuan sebuah model untuk memprediksi data baru yang tidak digunakan sehingga performa dari sebuah model dapat dievaluasi secara objektif (Brahmandjati et al., 2024)

d. Model XGBoost

XGboost merupakan algoritma yang menggabungkan *weak learner* umumnya berupa *decision tree* yang menghasilkan model prediksi yang kuat (Murdiansyah, 2024). Selain itu XGBoost merupakan algoritma *machine learning* yang unggul dalam akurasi, efisiensi dan pencegahan *overfitting*, algoritma ini sering digunakan untuk tugas-tugas prediksi yang kompleks (Brahmandjati et al., 2024).

Dalam penelitian ini, menggunakan XGBoost akan menghasilkan model dari data tour dan travel. Kemudian dari model ini akan menghasilkan sebuah evaluasi yang berisikan akurasi, presisi, recall dan f-score dari sebuah model. Evaluasi bertujuan untuk mengevaluasi performa dari suatu model dalam memprediksi data baru (Saputra et al., 2024). Selain itu penelitian ini menghasilkan *confusion matriks* dan kurva ROC serta nilai AUC. Luaran dari penelitian ini menghasilkan aplikasi dari model XGBoost dan dapat diproses apabila memasukkan data baru yang tidak memiliki label atau kelas, hasilnya berupa prediksi

terhadap data baru dengan label churn pelanggan atau tidak churn pelanggan.

### 3. Results and Discussion

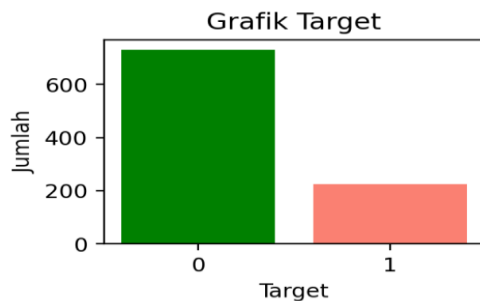
#### a. Data

Penelitian ini menggunakan dataset dari Kaggle. Data ini sudah memiliki label dengan kolom target. Data berjumlah 954 yang terdiri dari 730 pelanggan tidak churn (0) dan 224 pelanggan churn (1). Berikut data yang akan digunakan dalam penelitian ini.

Data Awal

	Age	FrequentFlyer	AnnualIncomeClass	ServicesOpted	AccountSyncedToSocialMedia	BookedHotelOrNot	Target
0	34	No	Middle Income	6	No	Yes	0
1	34	Yes	Low Income	5	Yes	No	1
2	37	No	Middle Income	3	Yes	No	0
3	30	No	Middle Income	2	No	No	0
4	30	No	Low Income	1	No	No	0
5	27	Yes	High Income	1	No	Yes	1
6	34	No	Middle Income	4	Yes	Yes	0
7	34	No	Low Income	2	Yes	No	1
8	30	No	Low Income	3	No	Yes	0
9	36	Yes	High Income	1	No	No	1

Gambar 2. Data Awal



Gambar 3. Grafik label target

#### b. Pemrosesan Data

Pada pemrosesan data dalam penelitian ini pertama menggunakan fungsi *dropna()* untuk menghapus nilai yang kosong pada data (Lumingkewas & Mokodaser, 2025). Selanjutnya pengubahan nilai kolom yang berisi kategorikal diubah menjadi numerik yang mudah dipahami oleh model (Sausan et al., 2024) ada beberapa kolom yang diubah menjadi numerik yaitu *FrequentFlyer*, *AnnualIncomeClass*, *AccountSyncedToSocialMedia*, *BookedHotelOrNot*. Penelitian ini

membutuhkan data yang bersih dan lengkap untuk menghasil hasil prediksi yang akurat.

```
# 3. Pemrosesan
# Hapus baris yang tidak lengkap
df = df.dropna()

#hapus norecord
if "FrequentFlyer" in df.columns:
    df = df[df["FrequentFlyer"].astype(str).str.lower().str.strip() != "no record"]

mapping_yesno = {"Yes": 1, "No": 0}
for col in ["FrequentFlyer", "AccountSyncedToSocialMedia", "BookedHotelOrNot"]:
    if col in df.columns:
        df[col] = df[col].map(mapping_yesno)

if "AnnualIncomeClass" in df.columns:
    df["AnnualIncomeClass"] = df["AnnualIncomeClass"].map({
        "Low Income": 0, "Middle Income": 1, "High Income": 2
    })
```

Gambar 4. Pemrosesan Data

Data Setelah Preprocessing

	Age	FrequentFlyer	AnnualIncomeClass	ServicesOpted	AccountSyncedToSocialMedia	BookedHotelOrNot	Target
0	34	0	1	6	0	0	0
1	34	1	0	5	1	0	1
2	37	0	1	3	1	0	0
3	30	0	1	2	0	0	0
4	30	0	0	1	0	0	0
5	27	1	2	1	0	1	1
6	34	0	1	4	1	1	0
7	34	0	0	2	1	0	1
8	30	0	0	3	0	1	0
9	36	1	2	1	0	0	1

Jumlah data setelah preprocessing: 84 baris, 7 kolom

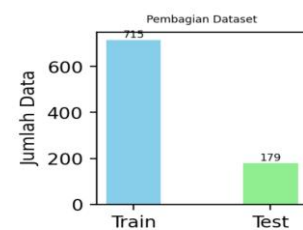
Gambar 5. Hasil Pemrosesan Data

#### c. Pembagian Dataset

Dataset akan dibagi menjadi data uji dan data latih, tujuan pembagian dataset untuk mengukur seberapa baik model dapat membuat prediksi yang akurat terhadap data yang belum pernah ada sebelumnya (Lumingkewas & Mokodaser, 2025).

```
# 5. Split data
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)
```

Gambar 6. Pembagian dataset



Jumlah data Train: 715  
Jumlah data Test: 179

Gambar 7. Grafik pembagian dataset

#### d. Model XGBoost

Evaluasi model, mengukur seberapa baik kinerja model dalam memprediksi nilai dari data uji (Dachi & Sitompul, 2023). Berikut adalah hasil dari model XGBoost.

### Hasil Evaluasi

Accuracy: 0.877

Precision: 0.744

Recall: 0.744

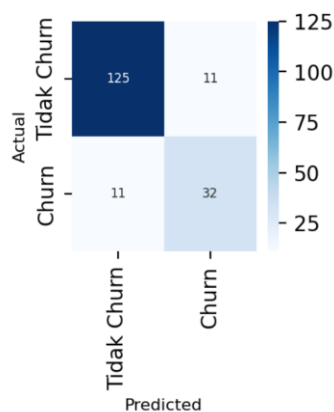
F1-score: 0.744

Gambar 8. Evaluasi XGBoost

Accuracy yang di hasilkan yaitu 0,877, precision 0,744, recall 0,744 dan F1-Score 0,744. Accuracy dengan 0,877 artinya model berhasil memprediksi 87,7% dari total target dengan benar baik itu churn atau tidak churn. Precision sebesar 0,744 artinya model ini berhasil memprediksi 74,4% dari total prediksi churn dengan benar, dalam penelitian ini model berhasil memprediksi 11 target churn dengan benar dari total 43 prediksi churn yang dibuat oleh model.

Recall nilainya sebesar 0,744 atau 74,4% yang artinya model berhasil mengenali 74,4% prediksi churn dengan benar. Model juga melewati 26,6% prediksi churn sebenarnya. Recall 0,744 atau 74,4% menunjukkan keseimbangan antara presisi dan recall.

Confusion Matrix

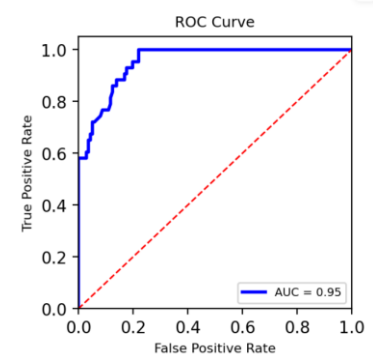


Gambar 9. Confusion Matriks

Confusion matrix merupakan tabel dua dimensi yang menampilkan jumlah prediksi benar dan prediksi salah, selain itu confusion matrix juga memberikan informasi antara kelas sebenarnya dengan kelas diprediksi (Dachi & Sitompul, 2023)

Dari gambar confusion matriks dibagi menjadi empat bagian yaitu TP (true positif), TN (true negative), FP (false positif), dan FN (false negative). Dari gambar di atas bisa diartikan untuk true positif yaitu ada 32 model yang berhasil dengan benar bahwa data tersebut adalah churn dan kenyataannya benar target churn. True negative ada 125 model berhasil menebak dengan benar bahwa foto tersebut tidak churn dan kenyataannya memang benar tidak churn. False positif ada 11 model yang salah memprediksi hasilnya churn ternyata kenyataannya tidak churn. False negative sebanyak 11 model yang berarti pada saat memprediksi bukan tidak churn dan bukan churn padahal kenyataannya hasilnya churn.

Kurva ROC & AUC

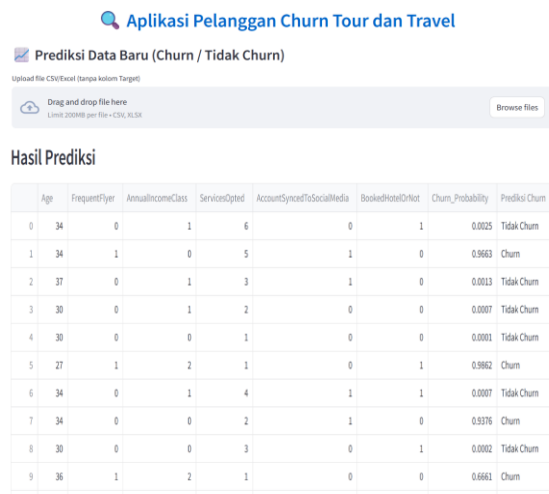


Gambar 10. Kurva ROC dan AUC

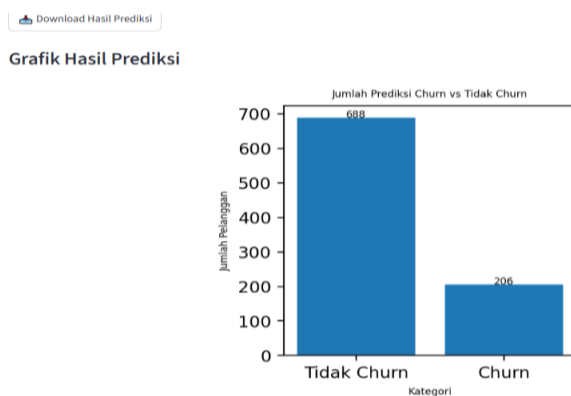
Kurva ROC (receiver operating characteristic) menunjukkan seberapa baik model dapat memisahkan antara label target (Naseer et al., 2024). Dari kurva ROC dapat menghitung nilai AUC (area under curve) yang merupakan ukuran dari performa model, dari penelitian ini didapatkan nilai AUC sebesar 0.95 termasuk ke dalam kriteria sempurna (excellent). Semakin

besar nilai AUC maka semakin baik pula kinerja sebuah model

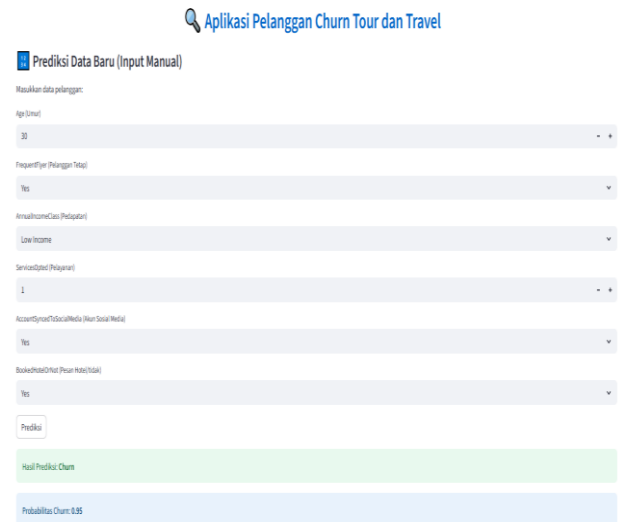
Setelah memiliki model dengan algoritma XGBoost, kemudian membuat sebuah aplikasi yang digunakan untuk memprediksi pelanggan churn atau tidak churn. Cara menggunakan aplikasi ini dengan cara memasukkan data baru tanpa target dapat berupa excel atau mengisi data langsung. Berikut adalah tampilan dari aplikasi.



Gambar 10. Hasil Prediksi



Gambar 11. Grafik Hasil Prediksi



Gambar 12. Form Hasil Prediksi

#### 4. Conclusions

Dari hasil penelitian yang sudah dilakukan, dapat ditarik kesimpulan bahwa model XGBoost dapat digunakan untuk memprediksi target churn atau tidak churn dalam pelanggan tour dan travel dengan akurasi 87,7%, precision 74,4%, recall 74,4% dan F1-Score 74,4%. Dengan melihat hasil akurasi dapat dikatakan model ini baik digunakan untuk prediksi dan dapat diimplementasikan. Selain itu untuk mengukur kinerja model dapat menghitung luas dibawah kurva ROC atau AUC, pada penelitian ini nilai AUC sebesar 0.95 yang termasuk ke dalam kategori sempurna (*excellent*). Manfaat dari penelitian ini, diharapkan dapat membantu pihak tour dan travel untuk melihat prediksi pelanggan menggunakan aplikasi yang telah dihasilkan, sehingga pihak tour dan travel bisa memikirkan cara untuk pelanggan churn supaya tetap menjadi pelanggan yang berkelanjutan. Berdasarkan hasil penelitian ini, disarankan untuk menambahkan model algoritma yang lain untuk memprediksi dan membandingkan hasil akurasi dari algoritma lain yang digunakan untuk mengetahui seberapa baik model ini melakukan prediksi.

#### References

Brahmandjati, A., Rahim, A. M. A., &

- Asharudin, F. (2024). Optimasi Prediksi Diabetes Dengan Algoritma XGBoost Dan Teknik Preprocessing Data. *Jurnal Ilmu Komputer Dan Pendidikan*, 3(1), 116–125. <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>, <https://doi.org/10.62411/tc.v24i2.12633>
- Budiarno, Udayana, I. B. N., & Lukitaningsih, A. (2022). Pengaruh Kualitas Layanan, Kualitas Produk Terhadap Kepuasan Pelanggan Dalam Membentuk Loyalitas Pelanggan. *Equilibrium: Jurnal Penelitian Pendidikan Dan Ekonomi*, 19(02), 226–233. <https://doi.org/10.25134/equi.v19i02.4531>
- Dachi, J. M. A. S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, 2(2), 87–103. <https://doi.org/10.55606/jurrimipa.v2i2.1470>
- Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *Nuansa Informatika*, 16(1), 141–149. <https://doi.org/10.25134/nuansa.v16i1.5406>
- Jange, B. (2022). Prediksi Harga Saham Bank BCA Menggunakan XGBoost. *ARBITRASE: Journal of Economics and Accounting*, 3(2), 231–237. <https://doi.org/10.47065/arbitrase.v3i2.495>
- Lumingkewas, C., & Mokodaser, W. G. (2025). Integrasi XGBoost dan Visualisasi Gradio untuk Memprediksi Pendapatan Pembayar Asuransi: Studi Kasus Rumah Sakit Swasta di Manado. *Techno.Com*, 24(2), 365–377. <https://doi.org/10.62411/tc.v24i2.12633>
- Maulana, B. A., & Hidayati, N. (2025). Churn Prediction in Credit Customers Using Random Forest and XGBoost Methods. 6(1), 81–89.
- Maulana, M. D., Hadiana, A. I., & Umbara, F. R. (2024). Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3251–3256. <https://doi.org/10.36040/jati.v7i5.7308>
- Murdiansyah, D. T. (2024). Prediksi Stroke Menggunakan Extreme Gradient Boosting. *JIKO (Jurnal Informatika Dan Komputer)*, 8(2), 419. <https://doi.org/10.26798/jiko.v8i2.1295>
- Naseer, W. A., Sarwido, & Wahono, B. B. (2024). Gradient Boosting Optimization With Pruning Technique for Prediction of BMT- AL-Hikmah Permata Customer Data. *Jurnal Informatika Teknologi Dan Sains (JINTEKS)*, 6(3), 719–727.
- Prasetyo, G. E., Megawaty, D. A., & Putra, A. D. (2023). Sistem Pelayanan Jasa Tour and Travel Berbasis Web. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 2(2), 85–92. <https://doi.org/10.58602/jima-ilkom.v2i2.21>
- Rizki Kurniawan, M., Nurul Sabrina, P., & Ilyas, R. (2024). Prediksi Customer Churn Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3369–3375. <https://doi.org/10.36040/jati.v7i5.7476>
- Saputra, A. A., Sari, B. N., & Rozikin, C. (2024). Penerapan Algoritma Extreme Gradient Boosting (XGboost) Untuk Analisis Risiko Kredit. *Jurnal Ilmiah*

*Wahana Pendidikan, 10(April), 27–36.*

Sausan, Pratiwi, D. M., & Mufidah, L. (2024). Perbandingan Metode Decision Tree Classifier dan XGBoost Classifier Dalam Memprediksi Penyakit Jantung. *Conference on Electrical Engineering, Informatics Technology and Creative Media 2024, 4*, 991–1000.

Tejashvi. (2024). *Tour & Travels Customer*

*Churn Prediction.*

Yulianti, S. E. H., Soesanto, O., & Sumawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics: Theory and Applications, 4(1)*, 21–26. <https://doi.org/10.31605/jomta.v4i1.1792>