

Penerapan Algoritma XGBoost Dalam Prediksi Harga Sewa Kos Di Kota Samarinda

Amalia Rahman^{1*}, Amelia Yusnita², Hanifah Ekawati³

¹ Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia

² Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia

³ Program Studi Bisnis Digital, STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: ¹2243005@wicida.ac.id, ²amelia@wicida.ac.id, ³hanifah@wicida.ac.id

(* : 2243005@wicida.ac.id)

Abstrak - Pertumbuhan jumlah penduduk dan meningkatnya aktivitas ekonomi di Kota Samarinda menyebabkan kebutuhan terhadap hunian sementara seperti rumah kos semakin tinggi. Namun, penentuan harga sewa kos masih banyak dilakukan berdasarkan intuisi pemilik tanpa mempertimbangkan faktor-faktor objektif seperti fasilitas, spesifikasi kamar, aksesibilitas transportasi, dan kedekatan terhadap fasilitas umum. Penelitian ini bertujuan untuk membangun model prediksi harga sewa kos berbasis algoritma *Extreme Gradient Boosting* (XGBoost) dengan pendekatan *Knowledge Discovery in Database* (KDD). Data penelitian diperoleh melalui proses *web scraping* dari situs Mamikos yang menghasilkan 231 data awal, kemudian dilakukan pembersihan dan penghapusan *outlier* sehingga tersisa 225 data valid. Lima fitur utama hasil *feature engineering* digunakan dalam model, yaitu Fasilitas Skor, Spesifikasi Gabungan Skor, Tempat Terdekat Skor, Transportasi Skor, dan Sistem Sewa Skor. Hasil evaluasi menunjukkan bahwa model XGBoost memiliki nilai *Mean Absolute Error* (MAE) sebesar Rp348,822, *Root Mean Squared Error* (RMSE) sebesar Rp416.139, dan koefisien determinasi (R^2) sebesar 0,612. Nilai tersebut menunjukkan bahwa model mampu menjelaskan 61,2% variasi harga sewa kos dengan performa prediksi yang cukup baik. Analisis *feature importance* menunjukkan bahwa Fasilitas Skor dan Spesifikasi Gabungan Skor merupakan faktor yang paling berpengaruh terhadap harga sewa, sedangkan faktor transportasi dan sistem sewa memiliki kontribusi lebih rendah. Penelitian ini diharapkan dapat menjadi acuan bagi pemilik kos, penyewa, dan pemangku kebijakan dalam menentukan harga sewa yang lebih objektif dan kompetitif berbasis pendekatan data mining.

Kata Kunci: XGBoost, Prediksi Harga Sewa, Data Mining, Rumah Kos, Kota Samarinda

1. PENDAHULUAN

Kota Samarinda merupakan salah satu wilayah dengan pertumbuhan ekonomi tercepat di Provinsi Kalimantan Timur. Berdasarkan data Badan Pusat Statistik (BPS) Kota Samarinda tahun 2024, jumlah penduduk mencapai 858.079 jiwa dengan laju pertumbuhan 0,88% per tahun dan luas wilayah 718,00 km² yang terbagi dalam 10 kecamatan. Kepadatan tertinggi terdapat di Kecamatan Samarinda Ulu (6.327 jiwa/km²), sementara yang terendah berada di Palaran dengan 309 orang per km². Ketidakseimbangan distribusi kepadatan ini secara langsung memengaruhi keperluan tempat tinggal sewaan di lokasi spesifik, khususnya dekat pusat pendidikan dan pusat ekonomi. Menurut laporan BPS (2024), sekitar 32,12% rumah tangga di Samarinda belum menempati rumah milik sendiri dan memilih untuk menyewa atau mengontrak. Sektor jasa berkontribusi paling dominan terhadap perekonomian kota, yaitu sebesar 71,89%, dengan tingkat partisipasi angkatan kerja (TPAK) mencapai 65,49%. Kondisi ini menjadikan Samarinda sebagai destinasi bagi mahasiswa dan pekerja migran dari berbagai wilayah di Kalimantan Timur. Pertumbuhan sektor pendidikan dan mobilitas tenaga kerja mendorong peningkatan permintaan terhadap hunian sementara berupa rumah kos [1].

Meskipun permintaan tinggi, proses penentuan harga sewa kos di Samarinda masih banyak dilakukan secara konvensional, berdasarkan intuisi atau perbandingan sederhana dengan kos lain yang dianggap sejenis. Pendekatan ini sering menghasilkan ketidaksesuaian antara harga dan kondisi riil kos, sehingga muncul variasi harga yang tidak konsisten untuk fasilitas dan lokasi yang relatif sama. Sejumlah penelitian menunjukkan bahwa faktor-faktor seperti kelengkapan fasilitas kamar, luas ruangan, akses transportasi, keamanan lingkungan, serta kedekatan dengan fasilitas umum berpengaruh signifikan terhadap harga sewa. Namun, analisis berbasis data secara sistematis masih jarang diterapkan oleh pemilik kos maupun penyewa, sehingga proses penentuan harga sering bersifat subjektif.

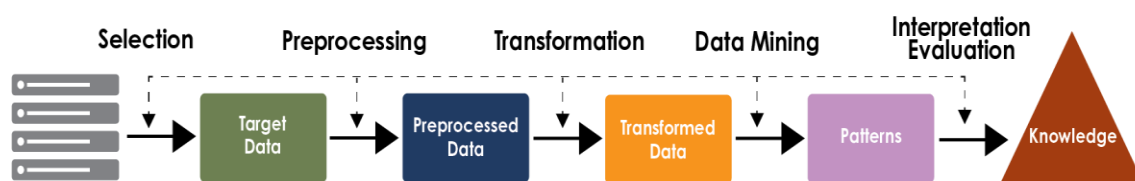
Harga sewa merupakan salah satu faktor penting yang dipertimbangkan calon penyewa dalam menentukan properti yang akan disewanya [2]. Kemajuan perkembangan *machine learning* (ML) telah membuka peluang besar dalam pemodelan harga properti berbasis data historis. Salah satu algoritma yang terbukti unggul untuk kasus regresi non-linear adalah *Extreme Gradient Boosting* (XGBoost). [3] XGBoost merupakan implementasi lanjutan dari *Gradient Boosted Decision Tree* (GBDT) yang memanfaatkan fungsi regularisasi guna menghindari *overfitting* serta meningkatkan efisiensi komputasi. [4] menjelaskan bahwa XGBoost dilengkapi dengan teknik regularisasi, pengaturan *subsampling*, dan optimasi komputasi yang efisien. XGBoost memiliki kemampuan *interpretability* melalui analisis *feature importance*, yang memungkinkan pengguna untuk mengetahui variabel paling berpengaruh terhadap harga. [5] Metode XGBoost adalah algoritma pengembangan dari *gradient tree boosting* yang berbasis algoritma ensemble, secara efektif bisa menanggulangi kasus *machine learning* yang berskala besar. Metode XGBoost dipilih karena memiliki beberapa fitur tambahan yang berguna untuk mempercepat sistem perhitungan dan mencegah *overfitting*. [6] Algoritma XGBoost, yang dikenal efektif dalam menangani data non-linear. Algoritma ini merupakan hasil pengembangan dari metode *gradient boosting decision tree* yang terkenal karena efisiensinya, kecepatan prosesnya, serta kemampuannya dalam memodelkan hubungan non-linear antar variabel dengan tingkat akurasi tinggi.

Penelitian terdahulu menunjukkan keunggulan XGBoost dalam mencapai tingkat akurasi tinggi pada berbagai studi prediksi harga properti. [7] melaporkan bahwa algoritma XGBoost memberikan performa prediksi paling akurat dibandingkan *Random Forest* dan *Gradient Boosting*, dengan nilai R^2 sebesar 0,7746 pada data harga rumah di Yogyakarta, menunjukkan tingkat generalisasi yang lebih baik terhadap variasi data properti. [8] membuktikan bahwa model XGBoost unggul dalam memprediksi harga sewa ruko di kota besar dengan tingkat akurasi mencapai R^2 tertinggi sebesar 0.6123. [9] juga menunjukkan bahwa kombinasi XGBoost dengan metode *ensemble learning* dapat meningkatkan nilai R^2 hingga 0,9076 pada prediksi harga rumah di Bandung. [10] Algoritma XGBoost terbukti efektif dalam regresi dan klasifikasi untuk data berdimensi tinggi, serta mampu menangani variabel yang kompleks. [11] menggunakan XGBoost dalam prediksi angka harapan hidup dan menunjukkan bahwa algoritma ini mampu menangkap hubungan non-linear antarvariabel dengan akurasi tinggi mencapai 96,8%. [12] menunjukkan bahwa penentu utama harga sewa kos bukan semata faktor lokasi, melainkan kelengkapan fasilitas kamar. Melalui analisis *feature importance* pada model XGBoost, fasilitas seperti AC, kloset duduk, dan kamar mandi dalam tercatat memiliki pengaruh paling besar terhadap variasi harga sewa, bahkan lebih tinggi dibandingkan kontribusi variabel lokasi. [13] menunjukkan bahwa algoritma XGBoost mampu menghasilkan performa prediksi yang sangat baik pada data kompleks. Model XGBoost yang dikembangkan dalam konteks prediksi harga mobil bekas di pasar India mencapai nilai R^2 sebesar 75,18% dengan MAE 87.071 rupee, yang menegaskan efektivitas algoritma ini dalam memodelkan hubungan non-linear dan menangani data beragam. [14] tentang preferensi mahasiswa ITB terhadap pemilihan hunian sewa menunjukkan bahwa kelengkapan fasilitas menjadi faktor prioritas utama, diikuti oleh harga dan keamanan, dengan mahasiswa yang menyewa hunian kategori harga tertinggi menjadikan faktor jarak dan fasilitas kamar mandi dalam sebagai prioritas utama. [15] dalam penelitian pemilihan kos harian di daerah Sleman mengidentifikasi delapan faktor prioritas yaitu lokasi, fasilitas kamar, fasilitas umum, harga dan kebijakan, keamanan, fleksibilitas reservasi, ulasan dan reputasi, serta kenyamanan dan kebersihan, dengan ulasan dan reputasi memiliki bobot prioritas tertinggi (0,22). Berdasarkan temuan-temuan tersebut, XGBoost terbukti memiliki keunggulan dalam hal akurasi, efisiensi, dan kemampuan interpretasi variabel, menjadikannya algoritma yang ideal untuk penelitian prediksi harga sewa kos di Samarinda. Namun, sebagian besar penelitian sebelumnya masih berfokus pada properti berskala besar seperti rumah dan apartemen, belum pada konteks mikro seperti kos. Dalam konteks Kota Samarinda, kajian tentang prediksi harga sewa kos berbasis *machine learning* masih terbatas, terutama dengan pemanfaatan data *real-time* hasil *web scraping*. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma XGBoost dalam memprediksi harga sewa kos di Kota Samarinda serta menganalisis faktor-faktor yang paling berpengaruh terhadap harga tersebut, meliputi fasilitas kos, spesifikasi kamar, kedekatan lokasi strategis, dan akses transportasi. Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) yang mencakup tahapan *data selection*, *data preprocessing*, *data transformation*, *data mining*, serta *evaluatin/interpretation*. Hasil penelitian ini diharapkan dapat menjadi referensi bagi pemilik kos, calon penyewa, pengembang aplikasi kos, maupun pengambil kebijakan dalam menetapkan harga sewa yang lebih objektif dan kompetitif berbasis pendekatan data mining.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) sebagai kerangka kerja utama dalam proses pengolahan data hingga penemuan pola pengetahuan. Pendekatan KDD dipilih karena mampu menggambarkan alur penelitian berbasis data mining secara sistematis dan terukur, mulai dari pengumpulan data mentah hingga diperolehnya pengetahuan yang dapat diinterpretasikan. Secara umum, tahapan KDD yang diterapkan dalam penelitian ini meliputi lima tahap utama, yaitu (1) *Data Selection*, (2) *Data Preprocessing*, (3) *Data Transformation*, (4) *Data Mining*, dan (5) *Interpretation/Evaluation*, [16], [17]. Adapun tahapan penelitian ditunjukkan secara konseptual pada Gambar 2.1.



Gambar 1. Tahap Penelitian

Tahapan penelitian yang dilakukan terdiri atas beberapa langkah utama sebagai berikut:

A. *Data Selection* (Pemilihan Data)

Tahap ini bertujuan mengumpulkan data yang relevan dengan tujuan penelitian. Data diperoleh melalui *web scraping* dari situs Mamikos menggunakan ekstensi DataMiner, yang menghasilkan 231 entri kos dengan 12 atribut utama. Informasi yang dikumpulkan mencakup nama kos, harga sewa, sistem pembayaran sewa, jenis kos, spesifikasi kamar, fasilitas kamar dan kamar mandi, fasilitas umum, fasilitas parkir, kecamatan lokasi kos, daftar tempat strategis terdekat, serta akses transportasi publik. Data tersebut mencerminkan kos di beberapa kecamatan di Kota Samarinda seperti Samarinda Ulu, Samarinda Ilir, Samarinda Kota, Samarinda Seberang, Samarinda Utara, Loa Janan Ilir, Sungai Pinang, dan Sungai Kunjang.

B. *Pra-Pemrosesan Data (Data Preprocessing)*

Tahap ini dilakukan untuk meningkatkan kualitas data sebelum masuk ke proses pemodelan. Proses yang dilakukan meliputi penghapusan data duplikat, pembersihan format harga menjadi numerik, normalisasi teks, dan penghapusan *outlier* menggunakan metode *Interquartile Range* (IQR). Selain itu, beberapa kolom teks dinormalisasi untuk memastikan konsistensi antarentri, seperti penyeragaman penulisan kecamatan. Hasil dari tahap ini adalah 225 data valid yang siap digunakan untuk tahap transformasi.

C. *Data Transformation* (Transformasi Data / *Feature Engineering*)

Transformasi data dilakukan untuk mengubah informasi mentah menjadi fitur numerik yang dapat diproses oleh algoritma XGBoost. Transformasi dilakukan melalui *feature engineering* pada lima atribut utama: Langkah-langkah yang dilakukan yaitu:

1) Fasilitas Skor

Menghitung skor berdasarkan jumlah fasilitas yang ditemukan pada empat kolom fasilitas (kamar, kamar mandi, umum, dan parkir). Semakin banyak fasilitas, semakin tinggi skor.

2) Tempat Terdekat Skor:

Menghitung skor berdasarkan rata-rata jarak (km) ke fasilitas umum seperti kampus, rumah sakit, pusat perbelanjaan, dan tempat makan menggunakan formula:

$$\text{Skor} = 1 - \frac{\text{Jarak (km)}}{\text{Jarak Maksimum (km)}}$$

Semakin dekat jarak lokasi kos ke fasilitas umum, semakin tinggi skor yang diperoleh.

3) Transportasi Skor:

Mengubah deskripsi akses transportasi menjadi jarak estimasi, kemudian dikonversi menggunakan formula yang sama dengan tempat terdekat.

4) Spesifikasi Gabungan Skor:

Hasil penggabungan ukuran kamar (m²) yang diekstraksi dari teks dan status listrik. Ukuran kamar diberi bobot 70% dan listrik 30%.

5) Sistem Sewa Skor:

Mengonversi jenis pembayaran sewa (harian, mingguan, bulanan, hingga tahunan) menjadi nilai skala 0–1 sesuai durasi sewanya.

Hasil tahap ini berupa dataset numerik dengan lima variabel prediktor utama.

D. *Data Mining* (Penerapan Algoritma XGBoost)

Tahap ini merupakan inti penelitian, yaitu penerapan algoritma XGBoost untuk memprediksi harga sewa kos. Data dibagi menjadi 80% data latih dan 20% data uji menggunakan *train-test split* dengan *random_state* 42 untuk menjamin reproduktibilitas. Parameter XGBoost disetel dengan *learning rate* 0.03, *max depth* 4, dan 1000 *estimators*, serta *early stopping* sebanyak 50 iterasi untuk mencegah *overfitting*.

E. *Interpretation/Evaluation* (Interpretasi Hasil/Evaluasi)

Evaluasi model dilakukan dengan tiga metrik utama, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R²). Ketiga metrik ini digunakan untuk mengukur tingkat kesalahan prediksi, stabilitas model, serta kemampuan model dalam menjelaskan variasi harga sewa kos. Tahap ini dilanjutkan dengan analisis *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh terhadap hasil prediksi.

2.2 Algoritma XGBoost

Extreme Gradient Boosting (XGBoost) merupakan algoritma *ensemble boosting* yang menggabungkan sejumlah pohon keputusan untuk memperbaiki kesalahan model sebelumnya. Secara matematis, fungsi objektif XGBoost dirumuskan sebagai [18]:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

dengan:

$l(y_i, \hat{y}_i)$: fungsi loss antara nilai aktual dan prediksi,

$\Omega(f_k)$: fungsi regularisasi untuk mengontrol kompleksitas model,

K: jumlah pohon keputusan

γ, λ : parameter regularisasi.

Regularisasi pada XGBoost menjadikan model lebih efisien dalam mencegah *overfitting* dibanding metode *boosting* konvensional. Model ini dioptimalkan menggunakan *gradient descent* terhadap turunan kedua (*second-order gradient*), mempercepat konvergensi dan meningkatkan stabilitas hasil [19].

2.3 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini diperoleh melalui proses *web scraping* dari situs Mamikos menggunakan ekstensi DataMiner. Proses scraping dilakukan untuk mengumpulkan data publik mengenai informasi harga sewa dan fasilitas kos yang tersedia di Kota Samarinda. Hasil pengumpulan data mencakup 231 entri kos dengan 12 atribut utama yang menggambarkan kondisi dan karakteristik masing-masing properti. Data mencakup kos yang tersebar di beberapa kecamatan, yaitu Samarinda Ulu, Samarinda Ilir, Samarinda Kota, Samarinda Seberang, Samarinda Utara, Loa Janan Ilir, Sungai Pinang, dan Sungai Kunjang. Namun demikian, dataset ini belum mencakup seluruh kecamatan di Kota Samarinda karena bergantung pada ketersediaan data publik dari situs Mamikos.

2.3.1 Variabel Deskriptif (Identifikasi / Kontekstual)

Variabel Deskriptif merupakan variabel yang memberikan informasi umum tentang data kos, tidak digunakan langsung dalam model.

Tabel 1 Variabel Deskriptif (Identifikasi / Kontekstual)

Nama Variabel	Deskripsi	Tipe Data	Keterangan
Nama Kos	Nama properti kos yang terdaftar pada situs Mamikos	String	Identitas unik setiap entri
Kecamatan	Lokasi administratif kos di wilayah Kota Samarinda	Categorical	Untuk analisis distribusi wilayah
Jenis Kos	Jenis penghuni kos (putra, putri, campur)	Categorical	Untuk segmentasi penghuni kos

2.3.2 Variabel Independen (Fitur Model)

Variabel Independen merupakan variabel (x) faktor mentah yang memengaruhi harga sewa.

Tabel 2. Variabel Independen (Fitur Model)

Nama Variabel	Deskripsi Variabel	Tipe Data
Sistem Sewa	Satuan waktu pembayaran sewa (harian, mingguan, bulanan, 3 bulanan, 6 bulanan, tahunan)	Categorical
Spesifikasi Tipe Kamar	Ukuran dan kondisi kamar (termasuk listrik atau tidak)	String
Fasilitas Kamar	Fasilitas yang tersedia di kamar seperti AC, kasur, meja, kursi, dan Wi-Fi	Text
Fasilitas Kamar Mandi	Jenis kamar mandi (dalam/luar) dan perlengkapannya	Text
Fasilitas Umum	Fasilitas bersama seperti dapur, laundry, ruang tamu, dan CCTV	Text
Fasilitas Parkir	Jenis area parkir yang disediakan (motor, mobil, atau keduanya)	Categorical
Tempat Terdekat	Daftar fasilitas umum di sekitar kos (kampus, rumah sakit, pusat belanja, tempat makan)	Text
Transportasi	Akses dan jarak ke sarana transportasi umum (halte, terminal, jalan utama)	Text

2.3.3 Intermediate Variables (Hasil Transformasi Fitur)

Intermediate Variables merupakan variabel hasil konversi fitur teks dan kategorikal menjadi numerik agar bisa dibaca oleh model.

Tabel 2.3 *Intermediate Variables* (Hasil Transformasi Fitur)

Nama Variabel	Deskripsi Transformasi	Rumus / Metode	Rentang Nilai
Fasilitas Skor	Jumlah fasilitas dari keempat atribut fasilitas (kamar, kamar mandi, umum, parkir). Semakin lengkap fasilitas skor tinggi.	Jumlah kata kunci fasilitas ÷ total kata kunci	0–1
Tempat Terdekat Skor	Skor berdasarkan rata-rata jarak (km) ke fasilitas umum. Semakin dekat skor tinggi.	$1 - (\text{jarak}/\text{km_max})$	0–1
Transportasi Skor	Skor kedekatan dengan akses transportasi publik. Semakin dekat skor tinggi.	$1 - (\text{jarak}/\text{km_max})$	0–1
Spesifikasi Gabungan Skor	Kombinasi ukuran kamar (m ²) dan kondisi listrik (termasuk/tidak).	Normalisasi ukuran + biner listrik	0–1
Sistem Sewa Skor	Skor berdasarkan lamanya periode sewa. Harian (0.0), mingguan (0.1), bulanan (0.3), 3 bulanan (0.5), 6 bulanan (0.7), tahunan (1.0).	Normalisasi durasi sewa	0–1

2.3.4 Variabel Dependen (Target / Label Kinerja Model)

Variabel Dependen merupakan variabel (y) nilai harga sewa kos yang menjadi target prediksi XGBoost.

Tabel 2.4 Variabel Dependen (Target / Label Kinerja Model)

Nama Variabel	Deskripsi	Tipe Data	Keterangan
Harga Sewa	Harga sewa kamar kos per bulan dalam satuan rupiah	Numeric	Variabel target (output) yang diprediksi oleh model XGBoost

2.4 Evaluasi Model

Evaluasi model dilakukan menggunakan tiga metrik utama:[18]

1. Mean Absolute Error (MAE):

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

Mengukur rata-rata selisih absolut antara nilai aktual dan prediksi

2. Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}$$

Mengukur besar kesalahan prediksi, dengan penalti lebih besar terhadap kesalahan ekstrem.

3. Koefisien Determinasi (R²):

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (Y - Y_i)^2}$$

Menunjukkan seberapa besar variasi data aktual yang dapat dijelaskan oleh model.

3. HASIL DAN PEMBAHASAN

Tahap hasil dan pembahasan ini menjelaskan proses penerapan algoritma *Extreme Gradient Boosting* (XGBoost) dalam memprediksi harga sewa kos di Kota Samarinda berdasarkan pendekatan *Knowledge Discovery in Database* (KDD). Proses KDD yang digunakan terdiri dari lima tahap utama, yaitu: (1) *Data Selection*, (2) *Data Preprocessing*, (3) *Data Transformation*, (4) *Data Mining*, (5) *Interpretation/Evaluation*. Dataset yang dianalisis terdiri dari 231 entri yang diperoleh melalui proses web scraping dari situs Mamikos, kemudian dibersihkan dan diseleksi hingga tersisa 225 data valid.

3.1 Data Selection (Pemilihan Data)

Jumlah Data Awal: 231
Jumlah Kolom: 12

	Nama Kos	Kecamatan	Jenis Kos	Harga Sewa	Sistem Sewa	Spesifikasi Tipe Kamar	Fasilitas Kamar	Fasilitas Kamar Mandi	Fasilitas Umum	Fasilitas Parkir	Tempat Terdekat	Transportasi
0	Kost 115 Ketapang Samarinda Ilir	Samarinda Ilir	Kos Campur	Rp770.000	bulan	3 x 4 meter Tidak termasuk listrik	TV Kabel	Bak mandi	R. Tami R. Makan Dapur Jemuran R. Keluarga	Parkir Motor	Rumah Makan Pakkadjia 1.5 km Bank Mandiri Samar...	NaN
1	Kost 69 Samarinda Seberang Samarinda	Samarinda Seberang	Kos Putra	Rp600.000	bulan	3 x 4 meter Termasuk listrik	Kasur Meja Lemari / Storage Ventilasi Bantal C...	Kloset Jongkok K. Mandi Luar Ember mandi Bak ...	WiFi R. Tami R. Santai Dispenser Dapur Balcon...	Parkir Motor Parkir Sepeda	Blasteran Cafe 1.3 km Masjid Al Baqa 2.7 km Ba...	Terminal Samarinda 2.6 km
2	Kost Abled Samarinda Utara Samarinda	Samarinda Utara	Kos Campur	Rp699.999	bulan	3 x 3 meter Tidak termasuk listrik	Dapur Pribadi	K. Mandi Dalam Kloset Jongkok	Jemuran	Parkir Motor Parkir Sepeda	Rumah Makan Zam-Zam 3.1 km	NaN
3	Kost Affiah Samarinda Ulu	Samarinda Ulu	Kos Putri	Rp800.000	bulan	2.75 x 2.75 meter Termasuk listrik	Kasur Lemari / Storage Kipas Angin Bantal Guling	Kloset Duduk K. Mandi Luar Ember mandi Bak mandi	R. Jemur Dapur Dispenser K. Mandi Luar	Parkir Motor Parkir Sepeda	Rumah Makan Banjar Manis 2.8 km Masjid Jami Al...	NaN
4	Kost Putri Samarinda Ulu	Samarinda Ulu	Kos Putri	Rp800.000	bulan	3 x 4 meter Termasuk listrik	Kasur Lemari / Storage Kipas Angin Bantal Guling	Kloset Jongkok K. Mandi Luar Ember mandi Bak ...	R. Jemur Dapur Dispenser K. Mandi Luar	Parkir Motor Parkir Sepeda	Rumah Makan Banjar Manis 2.8 km Masjid Jami Al...	NaN

Gambar 2 Tampilan Awal Dataset

Pada Gambar 2 menampilkan potongan lima baris pertama dari dataset hasil *web scraping*. Dataset terdiri atas 12 atribut utama, sebagian besar masih dalam bentuk teks mentah sehingga tidak dapat digunakan langsung untuk pemodelan. Kondisi ini menunjukkan bahwa data mentah dari situs daring memerlukan proses pembersihan dan standarisasi sebelum dapat diolah menggunakan metode *machine learning*.

3.2 Data Preprocessing (Prapemrosesan Data)

```
# DATA CLEANING
# Hapus duplikat
print("\nJumlah Duplikat:", df.duplicated().sum())

Jumlah Duplikat: 0
```

Gambar 3. Pemeriksaan Data Duplikat

Pada Gambar 3. menunjukkan hasil eksekusi `df.duplicated().sum()` yang menghasilkan nilai 0. Tidak adanya duplikasi menandakan bahwa setiap entri merepresentasikan properti kos yang unik sehingga tidak diperlukan langkah penghapusan duplikat. Tahap ini penting untuk menjaga integritas dan objektivitas data.

```
df.drop_duplicates(inplace=True)
print("Jumlah Data Setelah Hapus Duplikat:", len(df))

Jumlah Data Setelah Hapus Duplikat: 231
```

Gambar 4. Jumlah Data Setelah Penghapusan Duplikat

Pada Gambar 4. menampilkan jumlah data yang tetap 231 setelah pemeriksaan duplikasi, menegaskan bahwa seluruh data merupakan entri unik dan siap memasuki proses pembersihan berikutnya.

Jumlah Data Kosong pada Harga: 0

	Nama Kos	Harga Sewa
0	Kost 115 Ketapang Samarinda Ilir	770000.0
1	Kost 69 Samarinda Seberang Samarinda	600000.0
2	Kost Abled Samarinda Utara Samarinda	699999.0
3	Kost Affiah Samarinda Ulu	800000.0
4	Kost Putri Samarinda Ulu	800000.0

Gambar 5. Konversi Harga Sewa Ke Numerik

Pada 5. Gambar memperlihatkan hasil transformasi kolom Harga Sewa dari format teks seperti “Rp770.000” menjadi nilai numerik 770000.0. Selain itu, ditampilkan informasi bahwa tidak ada nilai kosong (NaN) setelah proses konversi, sehingga semua entri harga dapat digunakan untuk analisis. Transformasi ini merupakan langkah penting agar variabel harga dapat diproses oleh model XGBoost yang hanya menerima input numerik.

```
# Hapus Outlier dari Harga
Q1 = df["Harga Sewa Kos"].quantile(0.25)
Q3 = df["Harga Sewa Kos"].quantile(0.75)
IQR = Q3 - Q1
df = df[(df["Harga Sewa Kos"] >= Q1 - 1.5*IQR) &
        (df["Harga Sewa Kos"] <= Q3 + 1.5*IQR)]
print("Jumlah data setelah hapus outlier:", len(df))
```

Jumlah data setelah hapus outlier: 225

Gambar 6. Proses Penghapusan Outlier

Pada Gambar 6 menggambarkan proses eliminasi nilai ekstrem pada variabel harga sewa menggunakan metode Interquartile Range (IQR). Jumlah data berkurang dari 231 menjadi 225 observasi. Penghapusan outlier bertujuan meningkatkan stabilitas model dan mencegah distorsi akibat nilai harga yang terlalu rendah atau terlalu tinggi.

```
df.shape
print("Jumlah Data Akhir:", len(df))
print("Jumlah Kolom:", len(df.columns))
```

Jumlah Data Akhir: 225
Jumlah Kolom: 12

Gambar 7. Pengecekan Ulang Jumlah Dataset

Pada Gambar 7. menampilkan output df.shape setelah tahap pembersihan. Dataset akhir berjumlah 225 baris dan 12 kolom, menandakan bahwa data telah siap digunakan pada proses transformasi.

3.3 Data Transformation (Transformasi Fitur)

Jumlah Kos per Kecamatan

	Kecamatan	Jumlah Kos
0	Samarinda Ulu	110
1	Samarinda Utara	53
2	Sungai Pinang	22
3	Sungai Kunjang	21
4	Samarinda Seberang	8
5	Loa Janan Ilir	7
6	Samarinda Kota	3
7	Samarinda Ilir	1

Gambar 8. Jumlah Kos Berdasarkan Kecamatan

Pada Gambar 8. menunjukkan distribusi kos berdasarkan kecamatan. Samarinda Ulu mendominasi dengan 110 unit, diikuti Samarinda Utara dengan 53 unit, dan Sungai Pinang dengan 22 unit. Pola ini menggambarkan konsentrasi permintaan hunian di wilayah pendidikan dan pusat aktivitas ekonomi.

Jumlah Kos Berdasarkan Jenis Kos

	Jenis Kos	Jumlah Kos
0	Kos Putri	100
1	Kos Campur	84
2	Kos Putra	41

Gambar 9. Jumlah Kos Berdasarkan Jenis Kos

Pada Gambar 9. Kos Putri menjadi kategori terbanyak (100 unit), disusul Kos Campur (84 unit) dan Kos Putra (41 unit). Temuan ini sesuai dengan karakteristik kota pendidikan, di mana permintaan kos putri umumnya lebih tinggi.

	Nama Kos	Sistem Sewa	Kos	Sistem Sewa	Skor
0	Kost 115 Ketapang Samarinda Ilir	bulan			0.08
1	Kost 69 Samarinda Seberang Samarinda	bulan			0.08
2	Kost Abied Samarinda Utara Samarinda	bulan			0.08
3	Kost Afifah Samarinda Ulu	bulan			0.08
4	Kost Afifah Samarinda Ulu	bulan			0.08

Gambar 10 Konversi Sistem Sewa

Pada Gambar 10 memperlihatkan bahwa seluruh entri contoh menggunakan sistem sewa “bulanan” yang dikonversi menjadi skor 0.08 sesuai skala durasi sewa 0-1. Konversi ini dilakukan agar model dapat memaknai durasi sewa secara numerik. Transformasi ini penting untuk membaca hubungan antara periode sewa dan harga, meskipun variasinya rendah akibat dominasi mode sewa bulanan.

	Spesifikasi Tipe Kamar	Luas Kamar	Listrik	Skor	Spesifikasi Skor	Spesifikasi Gabungan Skor
0	3 x 4 meter Tidak termasuk listrik	12.00 m ²	1		0.700000	0.790000
1	3 x 4 meter Termasuk listrik	12.00 m ²	1		0.700000	0.790000
2	3 x 3 meter Tidak termasuk listrik	9.00 m ²	1		0.500000	0.650000
3	2.75 x 2.75 meter Termasuk listrik	7.56 m ²	1		0.500000	0.650000
4	2.75 x 2.75 meter Termasuk listrik	7.56 m ²	1		0.500000	0.650000

Gambar 11. Ekstraksi Spesifikasi Kamar

Pada Gambar 11. menampilkan Kolom Spesifikasi Tipe Kamar diolah untuk menghasilkan ukuran kamar dalam m² serta status listrik. Ukuran kamar diekstraksi dari pola teks seperti “3 x 4 meter”, sedangkan status listrik dikonversi menjadi skor biner. Kedua nilai ini digabung menjadi Spesifikasi Gabungan Skor dengan bobot 70% untuk luas dan 30% untuk listrik.

...	Nama Kos	Skor FKamar	Skor FKMandi	Skor FUmum	Skor FParkir	Fasilitas Skor
0	Kost 115 Ketapang Samarinda Ilir	0.045	0.000	0.091	0.045	0.045
1	Kost 69 Samarinda Seberang Samarinda	0.182	0.000	0.273	0.045	0.125
2	Kost Abied Samarinda Utara Samarinda	0.045	0.000	0.045	0.045	0.034
3	Kost Afifah Samarinda Ulu	0.136	0.045	0.091	0.045	0.080
4	Kost Putri Samarinda Ulu	0.136	0.000	0.091	0.045	0.068

Gambar 12. Skor Fasilitas

Pada Gambar 12. menunjukkan hasil perhitungan Fasilitas Skor berdasarkan jumlah kata kunci fasilitas pada empat kolom fasilitas. Semakin lengkap fasilitas yang tersedia, semakin tinggi skor fasilitas.

	Nama Kos	Tempat Terdekat	Tempat Terdekat	Jarak KM	Tempat Terdekat	Skor
0	Kost 115 Ketapang Samarinda Ilir	Rumah Makan Pakkadja	1.5 km Bank Mandiri Samar...	1.725		0.746324
1	Kost 69 Samarinda Seberang Samarinda	Blasteran Cafe	1.3 km Masjid Al Baqa 2.7 km Ba...	2.000		0.705882
2	Kost Abied Samarinda Utara Samarinda	Rumah Makan Zam-Zam	3.1 km	3.100		0.544118
3	Kost Afifah Samarinda Ulu	Rumah Makan Banjar Manis	2.8 km Masjid Jami Al...	2.275		0.665441
4	Kost Putri Samarinda Ulu	Rumah Makan Banjar Manis	2.8 km Masjid Jami Al...	2.275		0.665441

Gambar 13. Skor Tempat Terdekat

Pada Gambar 13. menampilkan proses ekstraksi jarak ke fasilitas penting seperti kampus, rumah sakit, pusat perbelanjaan, dan tempat makan. Nilai jarak dikonversi ke skor menggunakan formula $1 - (\text{jarak}/\text{km_max})$. Nilai kosong diisi dengan median.

	Nama Kos	Transportasi	Transportasi Jarak KM	Transportasi Skor
0	Kost 115 Ketapang Samarinda Ilir	NaN	NaN	0.800000
1	Kost 69 Samarinda Seberang Samarinda	Terminal Samarinda 2.6 km	2.6	0.832258
2	Kost Abied Samarinda Utara Samarinda	NaN	NaN	0.800000
3	Kost Afifah Samarinda Ulu	NaN	NaN	0.800000
4	Kost Putri Samarinda Ulu	NaN	NaN	0.800000

Gambar 14. Skor Trasportasi

Pada Gambar 14. memperlihatkan proses konversi deskripsi akses transportasi ke skor numerik. Jika tidak ditemukan angka jarak, penilaian dilakukan berdasarkan kata kunci seperti “dekat halte” atau “akses mudah”.

	Nama Kos	Sistem Sewa Skor	Spesifikasi Gabungan Skor	Fasilitas Skor	Tempat Terdekat Skor	Transportasi Skor	Total Skor	Harga Sewa
0	Kost 115 Ketapang Samarinda Ilir	0.08	0.79	0.045455	0.746324	0.800000	2.461778	770000.0
1	Kost 69 Samarinda Seberang Samarinda	0.08	0.79	0.125000	0.705882	0.832258	2.533140	600000.0
2	Kost Abied Samarinda Utara Samarinda	0.08	0.65	0.034091	0.544118	0.800000	2.108209	699999.0
3	Kost Afifah Samarinda Ulu	0.08	0.65	0.079545	0.665441	0.800000	2.274987	800000.0
4	Kost Putri Samarinda Ulu	0.08	0.79	0.068182	0.665441	0.800000	2.403623	800000.0

Gambar 15. Hasil Trasformasi Fitur dan Total Skor

Pada Gambar 15 menampilkan hasil kumulatif lima fitur utama (Sistem Sewa Skor, Spesifikasi Gabungan Skor, Fasilitas Skor, Tempat Terdekat Skor, Transportasi Skor). Total Skor menunjukkan representasi komprehensif dari kualitas kos yang akan digunakan dalam pemodelan.

3.4 Data Mining

```
# SPLIT DATA (TRAIN & TEST)
fitur = [
    "Sistem Sewa Skor",
    "Spesifikasi Gabungan Skor",
    "Fasilitas Skor",
    "Tempat Terdekat Skor",
    "Transportasi Skor"
]
target = "Harga Sewa"

df = df.dropna(subset=fitur + [target])
X = df[fitur]
y = df[target]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print("\nJumlah Data Latih:", len(X_train))
print("Jumlah Data Uji:", len(X_test))

Jumlah Data Latih: 180
Jumlah Data Uji: 45
```

Gambar 16 Pembagian Data Latih dan Uji

Pada Gambar 16 menunjukkan pembagian dataset menjadi 180 data latih (80%) dan 45 data uji (20%) menggunakan `train_test_split` dengan `random_state=42`. Pembagian acak terkontrol ini menjamin reproduktibilitas hasil dan memberikan evaluasi objektif terhadap performa model.

```
XGBRegressor(
    base_score=None, booster=None, callbacks=None,
    colsample_bylevel=None, colsample_bynode=None,
    colsample_bytree=0.8, device=None, early_stopping_rounds=50,
    enable_categorical=False, eval_metric='rmse', feature_types=None,
    feature_weights=None, gamma=0.1, grow_policy=None,
    importance_type=None, interaction_constraints=None,
    learning_rate=0.03, max_bin=None, max_cat_threshold=None,
    max_cat_to_onehot=None, max_delta_step=None, max_depth=4,
    max_leaves=None, min_child_weight=3, missing=nan,
    monotone_constraints=None, multi_strategy=None, n_estimators=1000,
    n_jobs=None, num_parallel_tree=None, ...)

```

Gambar 17 Pelatihan Model XGBoost

Pada Gambar 17. menampilkan konfigurasi parameter model XGBoost yang digunakan, seperti `learning_rate=0.03`, `max_depth=4`, dan `n_estimators=1000`. Parameter tersebut dirancang untuk mencapai keseimbangan antara akurasi dan

generalisasi. Penggunaan *early_stopping_rounds=50* membantu mencegah *overfitting* dengan menghentikan pelatihan saat model tidak lagi membaik pada data uji.

3.5 Interpretation/Evaluation (Interpretasi Hasil/Evaluasi)

Mean Absolute Error (MAE): 348,822
Root Mean Squared Error (RMSE): 416,139
Koefisien Determinasi (R^2): 0.612

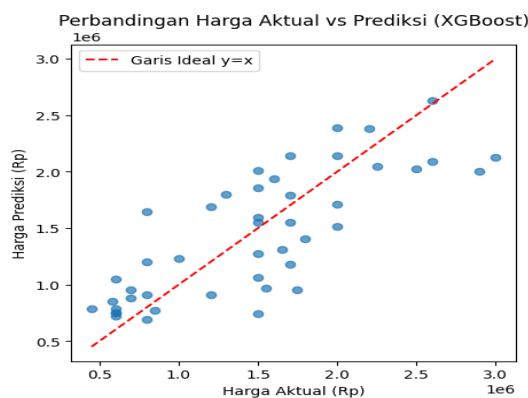
Gambar 18. Hasil Evaluasi Model XGBoost

Pada Gambar 18. menampilkan tiga metrik evaluasi utama yang digunakan untuk menilai kinerja model XGBoost dalam memprediksi harga sewa kos, yaitu *Mean Absolute Error* (MAE) sebesar 348.822, *Root Mean Squared Error* (RMSE) sebesar 416.139, dan koefisien determinasi (R^2) sebesar 0,612. Nilai-nilai tersebut menunjukkan bahwa model mampu menjelaskan sekitar 61,2% variasi harga sewa kos dengan tingkat kesalahan relatif rendah. Ini menandakan performa yang baik untuk model regresi non-linear. Hasil ini mengindikasikan bahwa model sudah mampu mengenali pola dasar hubungan antara atribut kos dan harga sewa.

	Harga Aktual (Rp)	Harga Prediksi (Rp)	Selisih (Rp)
9	Rp 1,700,000	Rp 1,200,000	Rp -500,000
188	Rp 600,000	Rp 700,000	Rp 100,000
122	Rp 1,500,000	Rp 1,600,000	Rp 100,000
213	Rp 600,000	Rp 1,000,000	Rp 400,000
151	Rp 1,500,000	Rp 2,000,000	Rp 500,000

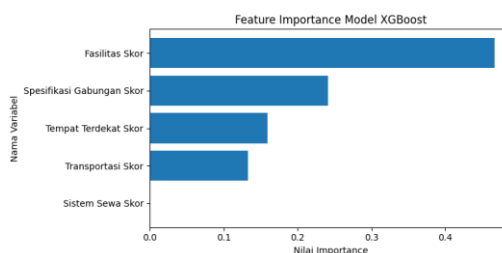
Gambar 19. Perbandingan Harga Aktual dan Prediksi

Pada Gambar 19. menunjukkan nilai aktual dan hasil prediksi harga kos. Mayoritas selisih prediksi berada pada kisaran ratusan ribu rupiah sehingga dapat disimpulkan bahwa model memiliki akurasi prediksi yang cukup baik.



Gambar 20. Scatter Plot Harga Aktual vs Prediksi

Pada Gambar 20. menampilkan visualisasi hubungan antara harga aktual dan prediksi. Titik-titik data tersebar mengikuti garis diagonal merah ($y = x$), yang menunjukkan bahwa model mampu mengikuti pola umum harga sewa. Penyebaran yang tidak terlalu jauh dari garis ideal menandakan kualitas prediksi yang baik dan stabilitas model XGBoost dalam mempelajari hubungan antar fitur.



Gambar 21. Feature Importance Model XGBoost

Pada Gambar 21. menampilkan visualisasi skor kontribusi fitur (*feature importance*) yang dihasilkan melalui perhitungan internal XGBoost. Skor tersebut menunjukkan besarnya kontribusi masing-masing variabel dalam mengurangi nilai kesalahan prediksi (*loss reduction*) pada proses pembentukan pohon keputusan. Variabel dengan kontribusi prediktif lebih besar akan memperoleh skor yang lebih tinggi, sedangkan fitur yang memberikan pengaruh lebih kecil dalam proses pemisahan data akan menghasilkan skor kontribusi yang lebih rendah.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) untuk memprediksi harga sewa kos di Kota Samarinda berdasarkan pendekatan *Knowledge Discovery in Database* (KDD). Melalui tahapan yang sistematis mulai dari *data selection*, *preprocessing*, *transformation*, *data mining*, dan *Interpretation/Evaluation*. Proses web scraping pada situs Mamikos menghasilkan 231 data awal, yang setelah melalui proses pembersihan dan penghapusan outlier tersisa 225 data valid dan siap diolah untuk pemodelan. Lima fitur utama yang dikembangkan melalui *feature engineering* skor fasilitas, skor spesifikasi gabungan, skor tempat terdekat, skor transportasi, dan skor sistem sewa berhasil merepresentasikan karakteristik kos secara kuantitatif dan relevan untuk kebutuhan prediksi. Model XGBoost yang dikembangkan menunjukkan performa prediksi yang cukup baik, dengan nilai *Mean Absolute Error* sebesar Rp348.822, *Root Mean Squared Error* sebesar Rp416.139, dan nilai koefisien determinasi (R^2) sebesar 0,612. menunjukkan bahwa model mampu menjelaskan sekitar 61,2% variasi harga sewa kos dengan tingkat kesalahan relatif rendah, ini menandakan performa yang baik untuk model regresi non-linear. Hasil ini mengindikasikan bahwa model sudah mampu mengenali pola dasar hubungan antara atribut kos dan harga sewa. Hasil Analisis *feature importance* menampilkan bahwa fasilitas dan spesifikasi kamar merupakan faktor paling dominan dalam menentukan harga sewa kos, sementara akses transportasi dan sistem sewa memiliki kontribusi lebih rendah.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma XGBoost efektif dalam memprediksi harga sewa kos berbasis data nyata dengan kompleksitas non-linear. Temuan ini memberikan dasar ilmiah bagi pemilik kos untuk menetapkan harga sewa yang lebih objektif dan kompetitif, sekaligus membantu calon penyewa dalam membandingkan harga berdasarkan parameter yang terukur. Penelitian selanjutnya dapat mengintegrasikan variabel spasial berbasis GIS, kualitas bangunan, atau data historis harga agar model prediksi memiliki akurasi lebih tinggi dan cakupan analisis yang lebih komprehensif.

REFERENCES

1. Badan Pusat Statistik Kota Samarinda. (2024). Statistik Daerah 2024 Kota Samarinda.
2. Wisnuadhi, B., Setiawan, I., & Korespondensi, P. (2021). Rekomendasi Fitur Yang Mempengaruhi Harga Sewa Menggunakan Pendekatan Machine Learning. 8(4), 673–682. <https://doi.org/10.25126/jtiik.202183305>
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
4. Nalluri, M., Pentela, M., & Eluri, N. R. (2020). A Scalable Tree Boosting System: XG Boost. *International Journal of Research Studies in Science, Engineering and Technology*, 7(12), 36–51. <https://doi.org/10.22259/2349-476X.0712005>
5. Yulianti, E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *JOMTA Journal of Mathematics: Theory and Applications*, 4(1).
6. Izzati, P. M., & Fitriyani. (2025). Implementasi Algoritma XGBoost Untuk Prediksi Capaian Bulanan Pendapatan Daerah Kota Bandung. *Jurnal Computer Science and Information Technology (CoSciTech)*, 6(2), 104–111. <https://doi.org/10.37859/coscitech.v6i2.9578>
7. Wulan Sari, B., Prabowo, D., Artikel, S., & Kunci, K. (2025). Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost Informasi Artikel A B S T R A K. *Intellect : Indonesian Journal of Innovation Learning and Technology*, 04(01), 42–51. <https://doi.org/10.57255/intellect.v4i1.1385>
8. Munawar, Y. F., & Arisal, A. (2025). Analisis Prediksi Harga Sewa Ruko Menggunakan Pendekatan Machine Learning. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(3), 2538–2544. <https://doi.org/10.31004/riggs.v4i3.2184>
9. Naufal Hibatulloh, M., Danu Prakoso, G., Dwi Putri Yunus, A., Dwi Putra, T., Amikom Surakarta, S., Veteran, J., Kartasura, K., Sukoharjo, K., & Tengah, J. (2025). Prediksi Harga Rumah di Bandung 2024 Menggunakan Ensemble Learning: Analisis Komparatif dan Interpretabilitas. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(2), 2477–5126. <https://doi.org/10.30591/jpit.v9ix.xxx>
10. Valentino, C., Arida Ayu Rahning Putri, L., Studi Informatika, P., Matematika dan Ilmu Pengetahuan Alam, F., Raya Kampus Udayana, J., Jimbaran, B., & Selatan, K. (2025). Analisis Kinerja XGBoost Menggunakan Bayesian Optimization dalam Prediksi Harga Ethereum. *JNATIA*, 3(4).

11. Kurniawan, W., & Indahyanti, U. (2024). Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost. *Indonesian Journal of Applied Technology*, 1(2), 18. <https://doi.org/10.47134/ijat.v1i2.3045>
12. Syawaludin, & W, A. A. B. S. (2025). Perbandingan Model Ensemble Learning Dalam Memprediksi Harga Sewa Indekos Di Jakarta. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.7737>
13. Adinagoro, R. E., Dwi, I., Premana, R., Pamungkas, R. B., Aryasetya, P., & Joedhistiro, D. (2025). Pengembangan Sistem Prediksi Harga Mobil Bekas Di Pasar India Menggunakan Algoritma XGBoost dan NLP. Seminar Nasional Teknologi Informasi Dan Bisnis (SENATIB).
14. Rahmadyani, H., Cahyaningtyas, A., & Fahri, M. (2022). Preferensi Mahasiswa ITB terhadap Pemilihan Hunian Sewa di Kota Bandung. *Jurnal Arsitektur ALUR*, 5(2).
15. Danianti, D., & Prastowo, D. W. (2023). Penerapan metode Ahp Danmfep Dalam menentukan Pemilihan Kos Harian Untuk wisatawan di Daerah Sleman. *Indonesian Journal of Business Intelligence*, 6(2).
16. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases* (© AAAI) (Vol. 17).
17. Ramang, E., Haryono, K., Lailiyah, S., & Sa'ad, M. I. (2025). Implementation of Data Clustering for Informatics Engineering Study Program Students at STMIK Widya Cipta Dharma Using the K-Means Method Implementasi Clustering Data Mahasiswa Program Studi Teknik Informatika STMIK Widya Cipta Dharma Menggunakan Metode K-Means. *Sebatik*. <https://doi.org/10.46984/sebatik.v29i1.0000>
18. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>