

Klasifikasi Kelayakan Pembelian Rumah Menggunakan Algoritma Decision Tree

Fitri Yanti^{1*}, Jaka Sutresna², Ahmad Musyafa³, Muhammad Fadhilah Akbar⁴

^{1,2,3,4}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: ¹*dosen00848@unpam.ac.id, ²dosen00833@unpam.ac.id, ³dosen00668@unpam.ac.id,
⁴mfadillahakbar877@gmail.com

(*Email Corresponding Author: dosen00848@unpam.ac.id)

Received: 26 Agustus 2026 | Revision: 1 September 2026 | Accepted: 7 September 2026

Abstrak

Harga rumah di setiap daerah berbeda-beda sesuai dengan daerah dan kategorinya masing-masing, khususnya harga rumah yang berada di kota Tangerang Selatan. Karena banyaknya variasi karakteristik properti sehingga masyarakat atau calon pembeli sering mengalami kesulitan untuk menentukan apakah harga rumah tersebut termasuk kedalam kategori murah atau mahal. Dalam penelitian ini memberikan solusi untuk mengelompokkan dan mengkategorikan harga rumah yang ada di Tangerang Selatan dengan menerapkan pendekatan data mining menggunakan algoritma Decision Tree. Algoritma ini dapat membantu untuk mengklasifikasi harga rumah di Tangerang Selatan dengan atribut yang digunakan terdiri dari 8 variabel, meliputi nomor data, nama rumah, harga dari rumah, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan kapasitas mobil dalam garasi. Implementasi model dilakukan menggunakan bantuan software Orange Data Mining serta bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa model Decision Tree yang terbentuk memiliki 39 nodes dan 20 leaves dengan pembagian kelas target berupa "Setuju" dan "tidak". Evaluasi performa model menggunakan metode 5-fold cross-validation menghasilkan Classification Accuracy (CA) sebesar 45% dan AUC sebesar 0.467. Analisis matriks konfusi menunjukkan adanya tingkat misklasifikasi yang tinggi akibat ketimpangan skala antar-fitur data numerik. Melalui klasifikasi ini, diharapkan calon pembeli dapat lebih mudah mengenali potensi penawaran properti secara lebih terstruktur.

Kata Kunci: Klasifikasi Harga Rumah, Decision Tree, Data Mining, Orange, Python.

Abstract

House prices in each region vary according to the region and category, especially house prices in South Tangerang. Due to the wide variety of property characteristics, the public or prospective buyers often have difficulty determining whether the house price is included in the cheap or expensive category. This study provides a solution to group and categorize house prices in South Tangerang by applying a data mining approach using the Decision Tree algorithm. This algorithm can help to classify house prices in South Tangerang with attributes used consisting of 8 variables, including data number, house name, house price, building area, land area, number of bedrooms, number of bathrooms, and car capacity in the garage. The model implementation was carried out using the assistance of Orange Data Mining software and the Python programming language. The results showed that the Decision Tree model formed had 39 nodes and 20 leaves with the target class division being "Agree" and "disagree". Evaluation of model performance using the 5-fold cross-validation method resulted in a Classification Accuracy (CA) of 45% and an AUC of 0.467. Confusion matrix analysis revealed a high level of misclassification due to scale imbalances between numerical data features. This classification is expected to help potential buyers more easily identify potential property offerings in a more structured manner.

Keywords: House Price Classification, Decision Tree, Data Mining, Orange, Python.

1. PENDAHULUAN

Rumah merupakan salah satu kebutuhan primer manusia yang memiliki fungsi sebagai tempat tinggal, tempat berlindung, serta sarana untuk menjalankan aktivitas kehidupan sehari-hari. Seiring dengan meningkatnya jumlah penduduk, urbanisasi, serta berkembangnya kawasan perkotaan, kebutuhan terhadap hunian terus mengalami peningkatan. Kondisi tersebut menyebabkan variasi harga rumah semakin beragam karena dipengaruhi oleh berbagai faktor, seperti luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, kapasitas garasi, lokasi, dan fasilitas pendukung lainnya. Variasi karakteristik tersebut sering kali menyebabkan masyarakat mengalami kesulitan dalam menentukan apakah suatu rumah layak untuk dibeli sesuai dengan harga yang ditawarkan.

Dalam praktiknya, keputusan pembelian rumah masih banyak dilakukan berdasarkan pengalaman pribadi, rekomendasi agen properti, maupun perbandingan harga dengan rumah di sekitar lokasi. Pendekatan tersebut cenderung bersifat subjektif karena belum didukung oleh analisis data secara sistematis. Akibatnya, calon pembeli berpotensi memperoleh rumah dengan harga yang kurang sesuai terhadap karakteristik properti yang dimiliki. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu membantu proses pengambilan keputusan secara lebih objektif, konsisten, dan terukur.

Perkembangan Artificial Intelligence (AI), khususnya pada bidang Machine Learning, telah memberikan berbagai solusi dalam pengolahan data untuk mendukung pengambilan keputusan. Salah satu teknik yang banyak digunakan adalah Data Mining, yaitu proses menggali pola atau pengetahuan dari sekumpulan data untuk menghasilkan informasi yang

bermanfaat[1][2]. Salah satu metode dalam Data Mining adalah teknik klasifikasi (*classification*), yaitu proses pembelajaran menggunakan data yang telah memiliki label (*supervised learning*) sehingga model mampu mengelompokkan data baru berdasarkan pola yang dipelajari [3].

Salah satu algoritma klasifikasi yang banyak digunakan adalah *Decision Tree*. Algoritma ini memiliki keunggulan berupa struktur pohon keputusan yang mudah dipahami (*interpretable*), mampu menangani data numerik maupun kategorikal, serta menghasilkan aturan keputusan (*decision rules*) yang dapat digunakan sebagai dasar pengambilan keputusan. Struktur tersebut memungkinkan pengguna untuk memahami hubungan antara atribut masukan dengan hasil klasifikasi yang dihasilkan oleh model [4]. Oleh karena itu, *Decision Tree* dapat diterapkan pada data properti yang memiliki berbagai karakteristik untuk menghasilkan pola klasifikasi yang lebih mudah diinterpretasikan.

Berbagai penelitian terdahulu telah membuktikan bahwa algoritma berbasis pohon keputusan mampu memberikan hasil yang baik dalam analisis data properti. Penelitian yang dilakukan oleh [5] menerapkan algoritma *Decision Tree* dengan pendekatan CRISP-DM untuk mengklasifikasikan kategori harga rumah ke dalam tiga kelas, yaitu rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa model yang dibangun mampu mencapai tingkat akurasi sebesar 81,74%, sehingga *Decision Tree* dinilai efektif dalam melakukan klasifikasi harga rumah sekaligus menghasilkan model yang mudah diinterpretasikan[6].

Penelitian lainnya dilakukan oleh [7] yang menggunakan *Decision Tree Regressor* untuk menganalisis faktor-faktor yang memengaruhi harga rumah di wilayah Sleman. Hasil penelitian menunjukkan bahwa atribut luas bangunan merupakan faktor yang paling dominan dalam menentukan harga rumah, diikuti oleh luas tanah, jumlah kamar mandi, jumlah kamar tidur, dan kapasitas carport. Penelitian tersebut juga menunjukkan bahwa karakteristik fisik rumah memiliki pengaruh yang signifikan terhadap penentuan harga properti.

R. A. Saputra and A. Pratama [8] juga menerapkan algoritma *Decision Tree* untuk memprediksi harga rumah di wilayah Tebet, Jakarta Selatan. Hasil penelitian menunjukkan bahwa *Decision Tree* mampu menghasilkan model prediksi yang mudah dipahami karena divisualisasikan dalam bentuk pohon keputusan. Selain itu, algoritma tersebut mampu menangani atribut numerik maupun kategorikal sehingga sesuai untuk diterapkan pada data properti yang memiliki karakteristik beragam.

Penelitian yang dilakukan oleh [9] menggunakan algoritma C4.5, yang merupakan pengembangan dari *Decision Tree*, untuk memprediksi harga rumah berdasarkan atribut luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan luas garasi. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu menghasilkan tingkat akurasi sebesar 85,70%, sehingga dinilai efektif dalam membantu proses prediksi harga rumah sekaligus meningkatkan transparansi dalam pengambilan keputusan di sektor properti.

Berdasarkan penelitian terdahulu tersebut, sebagian besar penelitian masih berfokus pada prediksi harga rumah atau klasifikasi kategori harga rumah. Penelitian-penelitian tersebut menunjukkan bahwa algoritma berbasis pohon keputusan dapat diterapkan pada data properti dengan berbagai karakteristik. Namun, penelitian yang secara khusus menggunakan algoritma *Decision Tree* untuk mengklasifikasikan kelayakan pembelian rumah ke dalam keputusan Setuju dan Tidak Setuju masih belum menjadi fokus utama. Padahal, klasifikasi tersebut memiliki orientasi yang lebih dekat dengan kebutuhan pengambilan keputusan calon pembeli karena tidak hanya memberikan informasi mengenai harga, tetapi juga menghasilkan keputusan berdasarkan karakteristik rumah yang tersedia.

Pada penelitian-penelitian sebelumnya lebih banyak menitikberatkan pada prediksi harga rumah maupun klasifikasi kategori harga menggunakan algoritma berbasis pohon keputusan. Penelitian ini menawarkan pendekatan yang berbeda dengan mengklasifikasikan keputusan kelayakan pembelian rumah, sehingga hasil yang diperoleh diharapkan lebih mudah digunakan sebagai dasar pengambilan keputusan oleh calon pembeli maupun pelaku usaha properti.

Berdasarkan *research gap* tersebut, penelitian ini menerapkan algoritma *Decision Tree* untuk mengklasifikasikan kelayakan pembelian rumah ke dalam dua kelas, yaitu **Setuju** dan **Tidak Setuju**, berdasarkan karakteristik properti. Pendekatan ini diharapkan dapat memberikan alternatif analisis berbasis data dalam membantu proses evaluasi kelayakan pembelian rumah secara lebih sistematis.

Penelitian ini bertujuan menerapkan algoritma *Decision Tree* untuk melakukan klasifikasi kelayakan pembelian rumah berdasarkan atribut harga, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan garasi. Tahapan penelitian menggunakan pendekatan *Knowledge Discovery in Databases (KDD)* yang meliputi *data cleaning*, *data selection*, *data transformation*, *data mining*, dan *evaluation*. Model dibangun menggunakan Orange Data Mining dan divalidasi menggunakan bahasa pemrograman Python dengan pustaka *Scikit-learn*, kemudian dievaluasi menggunakan metode 5-Fold Cross Validation dengan metrik *Classification Accuracy (CA)*, *Precision*, *Recall*, *F1-Score*, dan *Area Under the Curve (AUC)*.

Penelitian ini diharapkan dapat menghasilkan model klasifikasi yang mampu membantu masyarakat, calon pembeli, maupun pelaku usaha properti dalam mengevaluasi kelayakan pembelian rumah secara lebih objektif berdasarkan karakteristik properti yang dimiliki. Selain itu, penelitian ini diharapkan dapat menjadi referensi dalam penerapan algoritma *Decision Tree* pada pengembangan sistem pendukung keputusan di bidang properti.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis *Machine Learning*. Algoritma yang digunakan adalah *Decision Tree* untuk mengklasifikasikan kelayakan pembelian rumah berdasarkan karakteristik

properti. Proses penelitian mengadopsi metodologi Knowledge Discovery in Databases (KDD) yang terdiri atas tahapan *data cleaning*, *data selection*, *data transformation*, *data mining*, dan *evaluation*. Metode KDD dipilih karena mampu memberikan alur kerja yang sistematis dalam proses pengolahan data hingga menghasilkan pengetahuan baru yang dapat digunakan sebagai dasar pengambilan keputusan [3].

2.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data rumah yang diperoleh dari platform properti daring dan telah melalui proses seleksi sesuai kebutuhan penelitian. Setiap data merepresentasikan satu unit rumah dengan beberapa atribut yang memengaruhi keputusan pembelian.

Dataset disusun berdasarkan karakteristik properti yang umum digunakan dalam penelitian prediksi maupun klasifikasi harga rumah, seperti harga rumah, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan kapasitas garasi [7], [9].

Atribut yang digunakan dalam penelitian dapat dilihat pada Tabel 2.

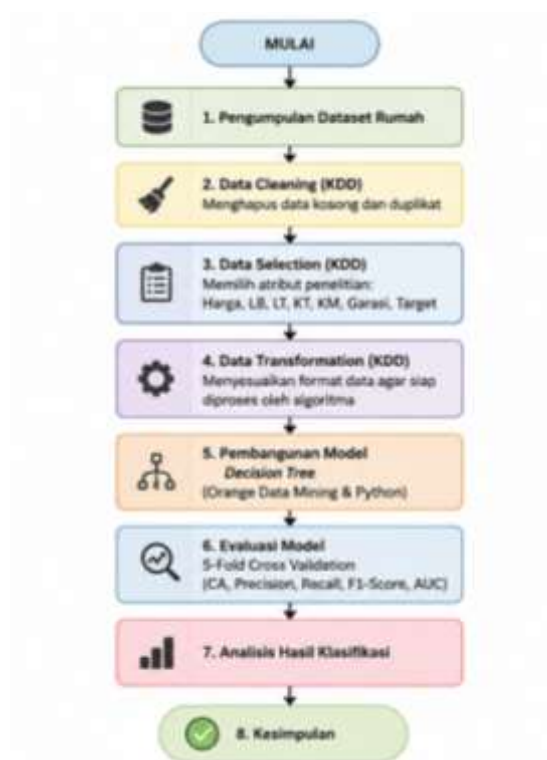
Tabel 2. Variabel Penelitian

No	Variabel	Tipe Data	Keterangan
1	Harga Rumah	Numerik	Harga jual rumah
2	Luas Bangunan	Numerik	Luas bangunan (m ²)
3	Luas Tanah	Numerik	Luas tanah (m ²)
4	Jumlah Kamar Tidur	Numerik	Jumlah kamar tidur
5	Jumlah Kamar Mandi	Numerik	Jumlah kamar mandi
6	Garasi	Numerik	Kapasitas garasi
7	Target	Kategori	Setuju / Tidak Setuju

Atribut-atribut tersebut dipilih karena dianggap memiliki pengaruh terhadap keputusan pembelian rumah berdasarkan karakteristik fisik properti.

2.2 Knowledge Discovery in Databases (KDD)

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) sebagai tahapan utama dalam pengolahan data. KDD merupakan proses sistematis yang digunakan untuk menemukan informasi atau pola yang bermanfaat dari sekumpulan data [3].



Gambar 1. Tahapan Penelitian Menggunakan Metode Knowledge Discovery in Databases (KDD)

Tahapan KDD yang digunakan meliputi:

- a. Data Cleaning
Tahap ini dilakukan untuk menghilangkan data yang kosong (*missing value*), data ganda (*duplicate data*), maupun data yang tidak konsisten sehingga menghasilkan dataset yang siap digunakan dalam proses analisis.
- b. Data Selection
Pada tahap ini dipilih atribut yang digunakan dalam penelitian, yaitu:
 1. Harga Rumah
 2. Luas Bangunan
 3. Luas Tanah
 4. Jumlah Kamar Tidur
 5. Jumlah Kamar Mandi
 6. Garasi
 7. Target (Setuju/Tidak Setuju)
- c. Data Transformation
Data kemudian ditransformasikan ke dalam format numerik maupun kategorikal agar dapat diproses menggunakan algoritma Decision Tree.
- d. Data Mining
Tahap ini merupakan proses pembangunan model klasifikasi menggunakan algoritma Decision Tree pada perangkat lunak Orange Data Mining dan Python.
- e. Evaluation
Tahap evaluasi dilakukan untuk mengetahui performa model menggunakan beberapa metrik evaluasi sehingga dapat diketahui kemampuan model dalam melakukan klasifikasi.

2.3 Metode Perhitungan Matematis Decision Tree (Gini Impurity)

Decision Tree merupakan salah satu algoritma klasifikasi yang membangun model dalam bentuk struktur pohon keputusan (*decision tree*) [3]. Setiap node merepresentasikan atribut, sedangkan setiap cabang menunjukkan hasil keputusan berdasarkan atribut tersebut hingga menghasilkan kelas target pada node akhir (*leaf node*). Algoritma ini banyak digunakan karena mudah diinterpretasikan dan mampu menangani atribut numerik maupun kategorikal [8].

Pada penelitian ini, pemilihan atribut dilakukan menggunakan Gini Index yang ditunjukkan dalam persamaan sebagai berikut.

$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

dengan:

S = himpunan data

p_i = proporsi kelas ke- i

Atribut dengan nilai impurity terkecil dipilih sebagai node utama dalam pembentukan pohon keputusan.

2.4 Implementasi Model

Implementasi model dilakukan menggunakan Orange Data Mining, yaitu perangkat lunak visual untuk proses *machine learning* dan *data mining* yang menyediakan berbagai widget analisis data tanpa memerlukan pemrograman secara penuh [10].

Validasi hasil klasifikasi juga dilakukan menggunakan bahasa pemrograman Python dengan pustaka **Scikit-learn**, yang menyediakan berbagai algoritma dan metrik evaluasi untuk analisis *machine learning* [11]. Sementara itu, proses pembangunan model Decision Tree dilakukan menggunakan perangkat lunak Orange Data Mining

Workflow penelitian terdiri atas beberapa widget utama yaitu:

- a. File
- b. Data Table
- c. Select Columns
- d. Decision Tree
- e. Test & Score
- f. Confusion Matrix
- g. Tree Viewer

Model Decision Tree kemudian dilatih menggunakan dataset yang telah diproses sebelum dilakukan pengujian performa.



Gambar 2. Workflow Implementasi Decision Tree Menggunakan Orange Data Mining

Workflow pada Orange Data Mining digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree*. Widget *File* digunakan untuk membaca *dataset*, *Select Columns* untuk menentukan atribut penelitian, *Decision Tree* untuk membangun model klasifikasi, *Tree Viewer* untuk menampilkan struktur pohon keputusan, sedangkan *Test & Score* dan *Confusion Matrix* digunakan untuk mengevaluasi performa model.

2.5 Evaluasi Model

Evaluasi model dilakukan menggunakan metode *5-Fold Cross Validation*, yaitu teknik validasi silang yang membagi dataset menjadi lima bagian sehingga setiap bagian memperoleh kesempatan menjadi data uji [12]. Teknik ini mampu memberikan estimasi performa model yang lebih stabil dibandingkan pembagian data tunggal, karena seluruh data memperoleh kesempatan yang sama untuk melatih dan menguji model. Dataset dibagi menjadi lima bagian dengan empat bagian digunakan sebagai data pelatihan dan satu bagian sebagai data pengujian, kemudian proses dilakukan secara bergantian sebanyak lima kali [12], [13].

Performa model dievaluasi menggunakan metrik *Classification Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Area Under Curve (AUC)* yang umum digunakan dalam evaluasi model klasifikasi [14]. *Classification Accuracy* mengukur proporsi prediksi yang benar secara keseluruhan. *Precision* mengukur proporsi *true positive* terhadap seluruh prediksi positif. *Recall* atau *sensitivity* mengukur proporsi *true positive* yang berhasil diidentifikasi dari keseluruhan data positif aktual. *F1-Score* merupakan *harmonic mean* dari *precision* dan *recall* sehingga memberikan keseimbangan antara kedua metrik tersebut. *AUC* mengukur kemampuan model dalam membedakan antara kelas positif dan negatif secara keseluruhan [14], [15], [16].

Metrik-metrik tersebut digunakan untuk mengetahui kemampuan model dalam mengklasifikasikan kelayakan pembelian rumah berdasarkan karakteristik properti.

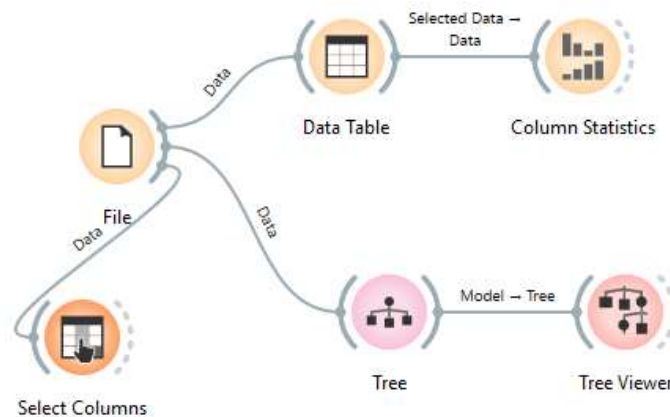
3. HASIL DAN PEMBAHASAN

3.1 Implementasi Model Klasifikasi

Setelah melalui tahapan *Knowledge Discovery in Databases (KDD)* yang meliputi *data cleaning*, *data selection*, *data transformation*, *data mining*, dan *evaluation*, proses selanjutnya adalah membangun model klasifikasi menggunakan algoritma *Decision Tree*. Implementasi model dilakukan menggunakan dataset yang terdiri atas 100 data rumah yang telah melalui proses preprocessing, sedangkan bahasa pemrograman Python digunakan sebagai pendukung dalam proses validasi model.

Orange Data Mining dipilih karena menyediakan lingkungan pengembangan berbasis visual (*visual programming*) yang memudahkan proses analisis data tanpa memerlukan penulisan kode secara keseluruhan. Melalui pendekatan ini, setiap tahapan pemodelan dapat dilakukan secara sistematis, mulai dari pembacaan dataset, pemilihan atribut,

pembangunan model klasifikasi, hingga proses evaluasi performa model. Workflow implementasi yang digunakan pada penelitian ini ditunjukkan pada Gambar 3.



Gambar 3. Workflow Implementasi Algoritma Decision Tree Menggunakan Orange Data Mining

Berdasarkan Gambar 3, proses implementasi diawali dengan penggunaan widget File untuk membaca dataset penelitian yang telah dipersiapkan. Dataset kemudian diteruskan ke widget Select Columns untuk menentukan atribut yang digunakan dalam proses klasifikasi, yaitu harga rumah, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, kapasitas garasi, serta atribut target berupa Setuju dan Tidak Setuju.

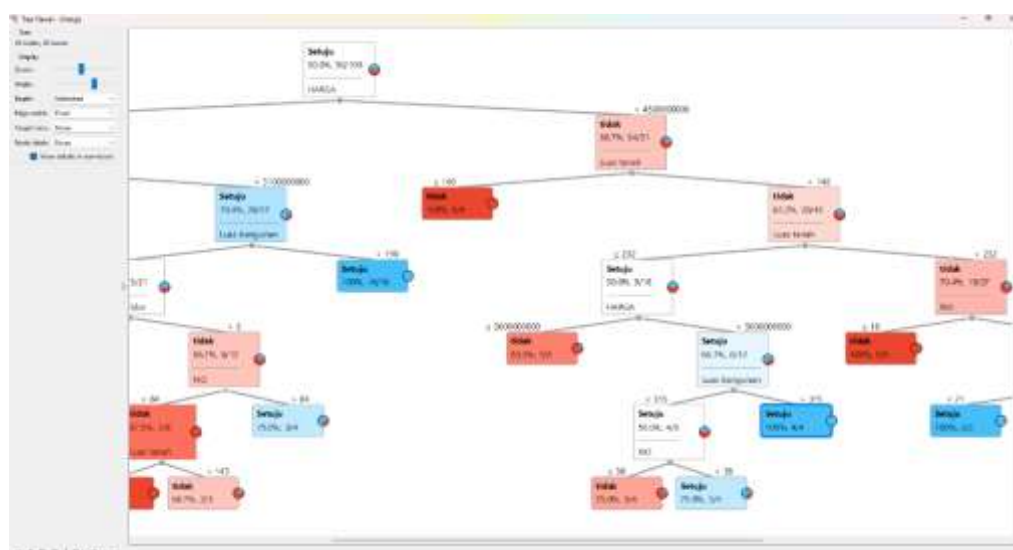
Selanjutnya dataset diproses menggunakan widget Decision Tree untuk membangun model klasifikasi berdasarkan karakteristik data yang tersedia. Model yang dihasilkan kemudian divisualisasikan menggunakan widget Tree Viewer sehingga struktur pohon keputusan dapat diamati secara langsung. Untuk mengetahui performa model, dilakukan proses evaluasi menggunakan widget Test & Score, sedangkan distribusi hasil klasifikasi dianalisis lebih lanjut menggunakan widget Confusion Matrix.

3.2 Pembentukan Model Decision Tree

Model klasifikasi pada penelitian ini dibangun menggunakan algoritma Decision Tree berdasarkan dataset yang telah melalui tahapan *preprocessing*. Algoritma Decision Tree bekerja dengan memilih atribut yang memiliki kemampuan terbaik dalam membedakan kelas target sehingga terbentuk struktur pohon keputusan (*decision tree*). Pada penelitian ini, pembentukan pohon dilakukan menggunakan kriteria Gini Impurity untuk menentukan atribut yang digunakan sebagai *root node* maupun *internal node*.

Kemampuan Decision Tree dalam menghasilkan model yang mudah diinterpretasikan menjadikannya banyak digunakan pada berbagai penelitian klasifikasi karena mampu menyajikan aturan keputusan dalam bentuk struktur pohon [3].

Hasil pembentukan model Decision Tree menggunakan Orange Data Mining ditunjukkan pada Gambar 4.



Gambar 4. Struktur Pohon Keputusan (Tree Viewer)

Berdasarkan Gambar 4, model Decision Tree yang dihasilkan terdiri atas 39 node dan 20 leaf node. Jumlah node menunjukkan tingkat kompleksitas model dalam memisahkan data berdasarkan atribut yang digunakan, sedangkan *leaf node* merepresentasikan keputusan akhir yang dihasilkan oleh model terhadap kelas Setuju maupun Tidak Setuju. Struktur pohon tersebut menggambarkan serangkaian aturan keputusan (*decision rules*) yang dibentuk secara otomatis berdasarkan karakteristik dataset.

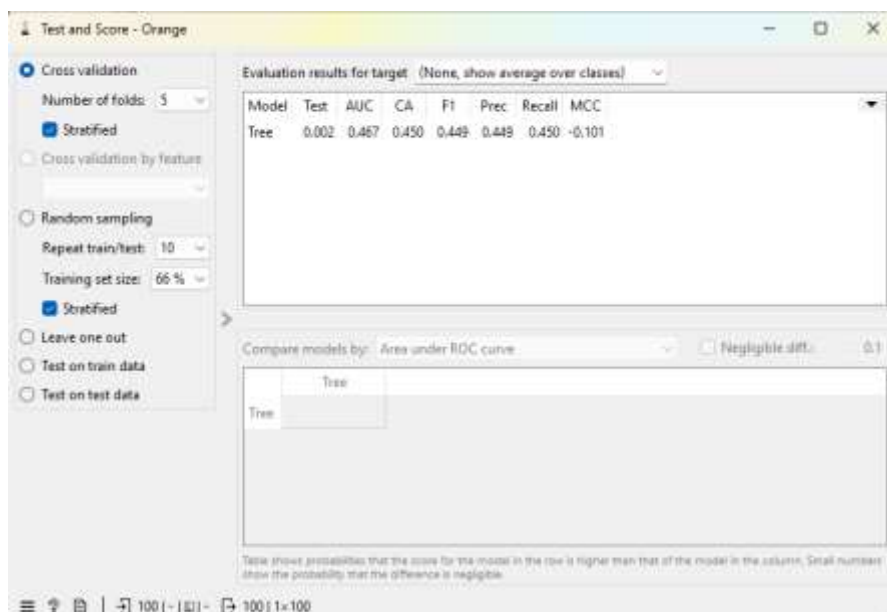
Atribut Harga Rumah terpilih sebagai root node, yang menunjukkan bahwa atribut tersebut memiliki kemampuan paling baik dalam membedakan kelas target dibandingkan atribut lainnya. Hal ini mengindikasikan bahwa harga rumah merupakan faktor utama yang memengaruhi keputusan kelayakan pembelian pada dataset yang digunakan. Setelah pemisahan berdasarkan harga rumah, algoritma melanjutkan proses klasifikasi dengan mempertimbangkan atribut lain, seperti jumlah kamar tidur, sehingga setiap cabang pada pohon menghasilkan aturan keputusan yang berbeda.

Keunggulan model Decision Tree adalah kemampuannya menghasilkan aturan keputusan yang mudah dipahami karena setiap proses klasifikasi dapat ditelusuri melalui struktur pohon yang terbentuk. Dengan demikian, pengguna tidak hanya memperoleh hasil klasifikasi, tetapi juga dapat memahami alasan mengapa suatu rumah dikategorikan sebagai Setuju atau Tidak Setuju berdasarkan kombinasi atribut yang dimiliki.

3.3 Evaluasi Model

Setelah model Decision Tree berhasil dibangun, tahap berikutnya adalah melakukan evaluasi untuk mengetahui kemampuan model dalam mengklasifikasikan kelayakan pembelian rumah. Evaluasi dilakukan menggunakan metode 5-Fold Cross Validation, yaitu teknik validasi silang yang membagi dataset menjadi lima bagian (*fold*). Pada setiap iterasi, empat bagian digunakan sebagai data latih (*training data*) dan satu bagian digunakan sebagai data uji (*testing data*). Proses tersebut dilakukan secara bergantian hingga seluruh data memperoleh kesempatan menjadi data uji. Teknik ini dipilih karena mampu memberikan estimasi performa model yang lebih stabil dibandingkan dengan pembagian data tunggal [3].

Hasil evaluasi model menggunakan widget Test & Score pada Orange Data Mining ditunjukkan pada Gambar 5.



Gambar 5. Hasil Evaluasi Model Menggunakan Test & Score

Untuk memudahkan interpretasi hasil pengujian, nilai evaluasi model disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model Decision Tree

Metrik Evaluasi	Nilai
Classification Accuracy (CA)	0,450
Precision	0,449
Recall	0,450
F1-Score	0,449
Area Under Curve (AUC)	0,467

Berdasarkan Tabel 3, model Decision Tree memperoleh nilai Area Under Curve (AUC) sebesar 0.467, Classification Accuracy (CA) sebesar 0.450, F1-Score sebesar 0.449, Precision sebesar 0.449, dan Recall sebesar 0.450.

Hasil tersebut menunjukkan bahwa model telah mampu melakukan proses klasifikasi terhadap data kelayakan pembelian rumah, namun tingkat performanya masih belum optimal.

Nilai Classification Accuracy (CA) sebesar 0.450 menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan sekitar 45,0% dari keseluruhan data uji. Sementara itu, nilai Precision sebesar 0.449 menunjukkan bahwa ketepatan prediksi model terhadap kelas positif masih relatif rendah. Nilai Recall sebesar 0.450 mengindikasikan bahwa model mampu mengenali sekitar 45,0% data pada kelas yang sebenarnya, sedangkan nilai F1-Score sebesar 0.449 menunjukkan bahwa keseimbangan antara *Precision* dan *Recall* juga masih tergolong rendah.

Selain itu, nilai AUC sebesar 0.467 menunjukkan bahwa kemampuan model dalam membedakan kelas Setuju dan Tidak Setuju masih belum optimal. Nilai tersebut mengindikasikan bahwa model masih mengalami kesulitan dalam mengenali pola yang membedakan kedua kelas target secara konsisten.

Performa model yang relatif rendah diduga dipengaruhi oleh beberapa faktor, antara lain jumlah dataset yang masih terbatas, distribusi data yang belum sepenuhnya merepresentasikan kondisi sebenarnya, serta belum diterapkannya proses optimasi parameter (*hyperparameter tuning*) maupun teknik *pruning* pada algoritma Decision Tree. Selain itu, atribut yang digunakan dalam penelitian masih terbatas pada karakteristik fisik rumah sehingga belum mempertimbangkan faktor eksternal seperti lokasi, aksesibilitas, maupun kondisi lingkungan yang dapat memengaruhi keputusan pembelian.

Meskipun demikian, algoritma Decision Tree tetap berhasil menghasilkan model klasifikasi yang memiliki tingkat interpretabilitas tinggi karena seluruh proses pengambilan keputusan dapat divisualisasikan dalam bentuk pohon keputusan. Keunggulan tersebut menjadikan Decision Tree sesuai untuk digunakan sebagai dasar pengembangan sistem pendukung keputusan pada bidang properti.

3.4 Analisis Confusion Matrix

Analisis ini bertujuan untuk mengetahui distribusi hasil prediksi model terhadap masing-masing kelas target, sehingga dapat diketahui jumlah data yang berhasil maupun gagal diklasifikasikan dengan benar.

Hasil Confusion Matrix menggunakan algoritma Decision Tree ditunjukkan pada Gambar 6.



Gambar 6. Hasil Confusion Matrix Algoritma Decision Tree

Widget Confusion Matrix dari aplikasi Orange Data Mining. Kali ini, evaluasi dilakukan menggunakan data penuh yang berjumlah 100 data.

Sama seperti sebelumnya, opsi yang dipilih di kanan atas adalah "Show: Proportion of actual", yang berarti angka persentase dihitung berdasarkan proporsi total nilai aktual (baris).

Hasil matriks ini memberikan visualisasi mendetail mengapa pada metrik *Test and Score* sebelumnya, akurasi model Tree (1) jeblok di angka 45% dan AUC-nya berada di bawah nilai acak (0.467).

Berikut adalah analisis dan cara membaca matriks tersebut:

a. Struktur Dataset Aktual vs Prediksi

1. Data Aktual (Baris): Dataset memiliki pembagian kelas yang sangat seimbang (*perfectly balanced*), yaitu 50 data dengan label asli "Setuju" dan 50 data dengan label asli "tidak" (Total = 100).
2. Hasil Prediksi Model (Kolom): Model Tree memprediksi 55 data sebagai "Setuju" dan 45 data sebagai "tidak".

b. Membaca Angka Persentase Klasifikasi

Mari kita bedah seberapa akurat tebakan model pada masing-masing kelas aktualnya:

Baris 1: Kelas Aktual "Setuju" (Total 50 data)

1. 50.0% (Benar): Dari 50 data yang aslinya "Setuju", model hanya berhasil menebak dengan benar sebanyak 50%-nya (yaitu 25 data).
2. 50.0% (Salah): Setengahnya lagi (25 data) salah diprediksi oleh model menjadi "tidak".

3. Artinya, untuk kelas "Setuju", performa model benar-benar sama seperti menebak acak dengan koin (50:50).

Baris 2: Kelas Aktual "tidak" (Total 50 data)

1. 60.0% (Salah/False Positive): Dari 50 data yang aslinya "tidak", model salah menebak sebanyak 60%-nya (yaitu 30 data) menjadi "Setuju".

2. 40.0% (Benar): Model hanya berhasil menebak dengan benar sebesar 40%-nya (yaitu 20 data) sebagai "tidak".

3. Artinya, model lebih sering salah menebak (60%) daripada benar (40%) ketika berhadapan dengan data properti yang aslinya masuk kategori "tidak".

c. Korelasi Total Akurasi (Akurasi = 45%)

Jika kita hitung total tebakan benar (kotak berwarna biru diagonal):

Dari total 100 data, model hanya berhasil menebak 45 data dengan benar. Inilah alasan matematis mengapa nilai Classification Accuracy (CA) bernilai 0.450 (45%) pada jendela *Test and Score* sebelumnya.

Model *Decision Tree* saat ini memiliki kecenderungan salah klasifikasi yang tinggi, khususnya pada kelas "tidak" yang sering kali "ditembak" sebagai "Setuju" oleh model (sebesar 60%).

Kondisi ini umumnya terjadi karena model kebingungan akibat skala data (fitur HARGA yang miliaran vs fitur KAMAR yang satuan) yang belum disetarakan, sehingga pohon keputusan membuat aturan batas (*split*) yang acak-acakan.

3.5 Analisis Pembentukan Pohon Keputusan

Untuk memahami proses pembentukan model *Decision Tree*, dilakukan analisis terhadap perhitungan Gini Impurity sebagai dasar dalam menentukan atribut yang digunakan sebagai *root node*. Gini Impurity digunakan untuk mengukur tingkat ketidakhomogenan (*impurity*) suatu himpunan data. Semakin kecil nilai Gini, maka semakin baik atribut tersebut dalam memisahkan kelas target.

Berdasarkan dataset penelitian yang terdiri atas 100 data, dengan distribusi 50 data kelas Setuju dan 50 data kelas Tidak Setuju, diperoleh nilai Gini awal sebagai berikut.

$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

$$Gini(S) = 1 - \left(\frac{50}{100}\right)^2 - \left(\frac{50}{100}\right)^2$$

$$Gini(S) = 1 - (0,5)^2 - (0,5)^2$$

$$Gini(S) = 1 - (0,25 + 0,25)$$

$$Gini(S) = 0,50$$

Nilai Gini sebesar 0,50 menunjukkan bahwa distribusi kelas pada node awal masih berada pada kondisi yang seimbang sehingga diperlukan proses pemisahan menggunakan atribut yang mampu menghasilkan tingkat impurity lebih kecil. Berdasarkan proses pembentukan pohon keputusan, atribut Harga Rumah dipilih sebagai *root node* karena memiliki kemampuan terbaik dalam memisahkan kelas Setuju dan Tidak Setuju.

Selanjutnya, algoritma melakukan proses pemisahan secara bertahap menggunakan atribut lain yang dianggap paling informatif. Setiap percabangan menghasilkan aturan keputusan (*decision rules*) yang digunakan untuk menentukan kelas target berdasarkan kombinasi karakteristik rumah.

Berikut ringkasan lengkap perhitungan manualnya:

Data: 100 rumah, 50 Setuju (layak beli) dan 50 tidak.

Langkah 1 — Gini Impurity, $Gini = 1 - (50/100)^2 - (50/100)^2 = 0.5000$ (tidak murni sempurna, kelas seimbang)

Langkah 2 — Split terbaik: $Harga \leq 4.550.000.000$ CART mencoba semua threshold semua fitur. Pemenangnya adalah split harga ini dengan:

a. Node kiri (n=49): $Gini = 0.4398$

b. Node kanan (n=51): $Gini = 0.4444$

c. $Gini\ Weighted = 0.4422 \rightarrow Gini\ Gain = 0.0578$

Langkah 3 — Node kiri $\rightarrow$ split lagi di $Harga \leq 3.149.500.000$

a. Kiri-kiri (n=12): Setuju=4, tidak=8 $\rightarrow Gini = 0.4444 \rightarrow$ prediksi: tidak

b. Kiri-kanan (n=37): Setuju=29, tidak=8 $\rightarrow Gini = 0.3389 \rightarrow$ prediksi: Setuju

c. $Gini\ Gain = 0.0750$ (terbaik di seluruh pohon)

Langkah 4 — Node kanan $\rightarrow$ split di Kamar Tidur ≤ 4

a. Kanan-kiri (n=11): Setuju=1, tidak=10 $\rightarrow Gini = 0.1653 \rightarrow$ hampir murni! $\rightarrow$ prediksi: tidak

b. Kanan-kanan (n=40): Setuju=16, tidak=24 $\rightarrow Gini = 0.4800 \rightarrow$ prediksi: tidak

Aturan keputusan yang dihasilkan:

a. $Harga \leq 3.15M \rightarrow$ tidak (rumah murah tapi lebih banyak yang tidak layak beli di dataset)

b. $3.15M < Harga \leq 4.55M \rightarrow$ Setuju (harga menengah, 78% layak beli)

c. $Harga > 4.55M$ dan Kamar Tidur $\leq 4 \rightarrow$ tidak (mahal tapi kamar sedikit, 91% tidak layak)

d. $Harga > 4.55M$ dan Kamar Tidur $> 4 \rightarrow$ tidak (mahal besar, mayoritas tidak layak beli)

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Decision Tree berhasil diterapkan untuk mengklasifikasikan kelayakan pembelian rumah berdasarkan atribut luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, garasi, dan harga rumah. Model yang dihasilkan membentuk struktur pohon keputusan dengan Harga Rumah sebagai *root node*, yang menunjukkan bahwa atribut tersebut menjadi faktor dominan dalam proses pembentukan klasifikasi pada dataset penelitian. Struktur pohon juga menghasilkan aturan keputusan berdasarkan batas nilai harga rumah dan jumlah kamar tidur, sehingga proses klasifikasi dapat diinterpretasikan dengan lebih mudah. Hasil evaluasi menggunakan metode 5-Fold Cross Validation memperoleh nilai Classification Accuracy sebesar 0,450, Precision sebesar 0,449, Recall sebesar 0,450, F1-Score sebesar 0,449, dan AUC sebesar 0,467. Hasil tersebut menunjukkan bahwa model telah mampu melakukan proses klasifikasi terhadap kelas Setuju dan Tidak Setuju, meskipun performanya masih belum optimal. Nilai evaluasi tersebut juga menunjukkan bahwa masih terdapat kesalahan dalam proses klasifikasi terhadap sebagian data rumah yang digunakan. Meskipun demikian, penerapan Decision Tree memberikan gambaran mengenai pola hubungan antara karakteristik rumah dengan keputusan pembelian serta menghasilkan aturan keputusan yang dapat dipahami. Dengan demikian, penelitian ini menunjukkan bahwa algoritma Decision Tree dapat digunakan sebagai pendekatan untuk klasifikasi kelayakan pembelian rumah dan dapat menjadi dasar bagi pengembangan model klasifikasi yang lebih optimal pada penelitian selanjutnya.

REFERENCES

- [1] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, pp. 75–82, 2020, doi: <https://doi.org/10.31294/ijcit.v5i1.7951>.
- [2] T. Herawan, R. Ghazali, N. M. Nawati, and M. M. Deris, Eds., *Recent Advances on Soft Computing and Data Mining*, vol. 549, in *Advances in Intelligent Systems and Computing*, vol. 549. Cham: Springer International Publishing, 2017. doi: <https://doi.org/10.1007/978-3-319-51281-5>.
- [3] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques (4th ed.)*. Morgan Kaufmann., vol. 1, no. 1. 2022.
- [4] S. Tan, A. Taeihagh, and K. Baxter, "The Risks of Machine Learning Systems," Apr. 2022, [Online]. Available: <http://arxiv.org/abs/2204.09852>
- [5] B. A. Ghaisani, G. A. Damayanti, R. M. A. Rani, N. Fitri, and W. S. Dharmawan, "Program Studi Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika," *JURNAL INFORMATIKA UPGRI*, vol. 11, no. 2, Dec. 2025, doi: <https://doi.org/10.26877/jiu.v11i2.25421>.
- [6] X. Zhao and X. Nie, "Splitting choice and computational complexity analysis of decision trees," *Entropy*, vol. 23, no. 10, Oct. 2021, doi: <https://doi.org/10.3390/e23101241>.
- [7] S. M. R. Aridho, H. A. K. Aswaha, T. D. Wahyuni, and A. A. Sari, "Analisis Faktor yang Mempengaruhi Harga Rumah Menggunakan Decision Tree," *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, p. 2025, Jul. 2025.
- [8] R. A. Saputra and A. Pratama, "IMPLEMENTASI DECISION TREE UNTUK PREDIKSI HARGA RUMAH DI DAERAH TEBET," *Journal of Information System Management (JOISM) e-ISSN*, vol. 6, no. 2, pp. 2715–3088, Jan. 2025.
- [9] I. A. Subekti, C. B. Andriano, D. Nurdiansyah, and R. Hidayat, "PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI HARGA RUMAH," *Just IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 15, no. 2, pp. 350–355, Jan. 2025.
- [10] J. Demšar *et al.*, "Orange: Data Mining Toolbox in Python Tomaž Curk Matija Polajnar Lañ Zagar," 2013.
- [11] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. doi: <https://doi.org/10.1109/tpami.2009.187>.
- [12] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010, doi: <https://doi.org/10.1109/TPAMI.2009.187>.
- [13] C. Schaffer, "Selecting a Classification Method by Cross-Validation," 1993.
- [14] O. Rainio, J. Teuhio, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: <https://doi.org/10.1038/s41598-024-56706-x>.
- [15] D. Chicco, N. Tötsch, and G. Jurman, "The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Min.*, vol. 14, pp. 1–22, 2021, doi: <https://doi.org/10.1186/s13040-021-00244-z>.
- [16] M. Al-Batah, M. Salem Alzboon, M. Alqaraleh, and F. Ahmad Alzaghou, "Comparative Analysis of Advanced Data Mining Methods for Enhancing Medical Diagnosis and Prognosis," *Data and Metadata*, vol. 3, Jan. 2024, doi: <https://doi.org/10.56294/DM2024.465>.