

ANALISIS ALGORITMA GRADIENT BOOSTING DALAM PENGARUH MASYARAKAT MEMILIH RUMAH SEWA

Rizka Dahlia¹, Lady Agustin Fitriana², Syarah Seimahuira³

^{1,2}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

³Program Studi Sains Data, Universitas Nusa Mandiri

Jln. Kramat Raya, Jakarta Pusat, DKI Jakarta, 10450

¹Rizka.rzl@bsi.ac.id , ²Lady.lag@bsi.ac.id , ³Syarah.syr@nusamandiri.ac.id

Abstract

The high population growth that is not in line with the availability of housing land has created complex challenges, especially related to the increase in land prices which leads to the increase in house prices. Rental housing is becoming an increasingly important alternative for people to get a decent place to live. In addition, rental houses also provide profitable business opportunities for property owners and rental service providers. To help people in choosing a rental house that suits their needs, the application of the Gradient Boosting algorithm can be an effective solution. This algorithm works by sequentially building a series of predictive models, where each model tries to correct the error generated by the previous model. As one of the methods in machine learning, Gradient Boosting belongs to the ensemble learning category, which builds a strong model from the combination of several relatively weaker models. This research utilized data from the Kaggle repository and found that the Gradient Boosting algorithm showed very promising performance in predicting house rental prices, with an accuracy rate of 84.38%. Further evaluation using the Area Under the Curve (AUC) metric showed a result of 92.65%, indicating the algorithm's ability to predict both positive and negative data with high accuracy. The results of this study show that Gradient Boosting has great potential as a tool in accurately predicting house rental prices, as well as helping to address the challenge of affordable housing availability.

Keywords : House Rent, Data Mining, Gradient Boosting Algorithm

Abstrak

Tingginya pertumbuhan populasi yang tidak sejalan dengan ketersediaan lahan perumahan telah menimbulkan tantangan yang kompleks, terutama terkait dengan peningkatan harga tanah yang berujung pada naiknya harga rumah. Rumah sewa hadir menjadi alternatif yang semakin penting bagi masyarakat untuk mendapatkan tempat tinggal yang layak. Selain itu, rumah sewa juga memberikan peluang bisnis yang menguntungkan bagi pemilik properti dan penyedia jasa sewa. Untuk membantu masyarakat dalam memilih rumah sewa yang sesuai dengan kebutuhan mereka, penerapan algoritma Gradient Boosting dapat menjadi solusi yang efektif. Algoritma ini bekerja dengan cara membangun serangkaian model prediktif secara berurutan, di mana setiap model mencoba memperbaiki kesalahan yang dihasilkan oleh model sebelumnya. Sebagai salah satu metode dalam machine learning, Gradient Boosting termasuk dalam kategori ensemble learning, yang membangun model yang kuat dari gabungan beberapa model yang relatif lebih lemah. Penelitian ini menggunakan data dari repository Kaggle dan menemukan bahwa algoritma Gradient Boosting menunjukkan kinerja yang sangat menjanjikan dalam memprediksi harga sewa rumah, dengan tingkat akurasi mencapai 84,38%. Evaluasi lebih lanjut menggunakan metrik Area Under the Curve (AUC) menunjukkan hasil sebesar 92,65%, yang mengindikasikan kemampuan algoritma ini dalam memprediksi baik data positif maupun negatif dengan akurasi yang tinggi. Hasil penelitian ini menunjukkan bahwa Gradient Boosting memiliki potensi besar sebagai alat bantu dalam memprediksi



harga sewa rumah secara akurat, serta membantu mengatasi tantangan ketersediaan perumahan yang terjangkau.

Kata kunci : Rumah Sewa, Data Mining, Algoritma Gradient Boosting

1. PENDAHULUAN

Peningkatan jumlah penduduk yang tidak sejalan dengan lahan rumah memberikan masalah dan tantangan baru. Pertambahan jumlah penduduk tanpa disertai dengan pertumbuhan perumahan yang memadai menyebabkan krisis perumahan ditambah dengan keterbatasan lahan yang mengakibatkan melonjaknya harga properti rumah serta tersedianya perumahan yang buruk. Apalagi rumah menjadi kebutuhan dasar masyarakat, tak hanya sekedar menjadi tempat tinggal, rumah juga menjadi tempat perlindungan dari lingkungan eksternal [1]. Tingginya harga tanah dapat menjadi salah satu faktor utama penyebab harga rumah menjadi mahal. Ketika lahan yang tersedia terbatas sementara permintaan terus meningkat menjadikan harga tanah semakin naik apalagi jika lahan tersebut terletak di lokasi strategis seperti pusat kota yang dekat dengan fasilitas umum, seringkali dengan mahalnya tanah tersebut mahal juga rumah yang akan dibangun. Sehingga bagi masyarakat yang memiliki ekonomi rendah akan menjadi lebih sulit memiliki sebuah rumah dan memilih untuk mengambil rumah sewa atau kontrakan [2].

Kehadiran rumah sewa tidak hanya membantu masyarakat memiliki rumah untuk sementara namun juga menciptakan peluang bisnis yang signifikan bagi para pemilik properti dan penyedia jasa sewa. Di balik manfaat kemudahan dalam kepemilikan rumah sementara, masyarakat juga harus memahami regulasi sewa, pemeliharaan properti hingga hubungan antara pemilik dan penyewa, sehingga masyarakat harus lebih menyadari baik buruknya dalam pemilihan rumah sewa [3]. Kemampuan untuk memprediksi harga sewa rumah secara akurat menjadi sangat penting. Model prediktif yang efektif dapat membantu masyarakat dalam membuat keputusan yang lebih baik terkait pilihan rumah sewa, serta membantu pemilik properti dalam menentukan harga sewa yang wajar. Salah satu metode yang menjanjikan dalam bidang ini adalah algoritma

Gradient Boosting, yang dikenal efektif dalam menangani masalah prediksi dan klasifikasi [4].

Algoritma Gradient Boosting bekerja dengan membangun serangkaian model prediktif secara berurutan, di mana setiap model berusaha memperbaiki kesalahan prediksi model sebelumnya. Algoritma ini termasuk dalam kategori ensemble learning, di mana model yang kuat dibangun dari beberapa model yang relatif lemah [5]. Menurut Iqbal H Sarker, algoritma gradient boosting merupakan teknik ensemble yang digunakan dalam machine learning untuk tugas regresi dan klasifikasi. Teknik ini bekerja dengan membangun model secara berurutan, di mana setiap model baru berusaha untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya [6]. Dalam penelitian terdahulu perbandingan Gradient Boosting dan Light Gradient Boosting untuk melakukan klasifikasi rumah sewa menghasilkan bahwa algoritma Gradient Boosting menjadi model yang paling cocok digunakan dengan hasil accuracy 84.38%, precision 83.33 %, dan recall 87.53%, perbandingan dari confusion matrix, Gradient Boosting mempunyai jumlah hasil prediksi data yang lebih besar ketimbang Light Gradient Boosting [7].

Penerapan Algoritma Gradient Boosting dalam penelitian yang dilakukan oleh taufik Zulhaq Jasman dkk, berjudul Analisis Algoritma Gradient Boosting, AdaBoost dan CatBoost dalam Klasifikasi Kualitas Air. Penelitian yang bertujuan untuk mendapatkan akurasi tertinggi dari ketiga algoritma tersebut yaitu untuk menangani data tidak seimbang menggunakan metode SMOTE (Synthetic Minority Over Sampling Technique). Adapun kinerja dari ketiganya menghasilkan bahwa algoritma CatBoost memiliki akurasi dan kinerja paling tinggi sebesar 68%, sedangkan Gradient Boosting memiliki akurasi sebesar 60% dan AdaBoost sebesar 58%. Disamping itu, performas nilai AUC (Area Under the Curve) CatBoost sebesar 0,678. Namun penelitian menampilkan hasil akurasi dan performa yang masih kurang [8]. Sementara itu, Gradient



Boosting juga diterapkan dalam penelitian Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression method dalam Prediksi Harga Rumah. Penelitian ini melakukan pengujian harga rumah, luas tanah, luas bangunan, jumlah kamar mandi, jumlah kamar tidur hingga garasi, dengan hasil prediksi lebih baik menerapkan random Forest Regression dengan akurasi tertinggi sebesar 81,5 % dibandingkan dengan metode Linear Regression dan Gradient Boosted Trees regression .

Model Algoritma Boosted Gradient Trees juga diterapkan dalam penentuan kelayakan pemberian kredit koperasi dengan hasil penelitian bahwa Gradient Boosting ternyata mampu meningkatkan akurasi dalam menganalisa kelayakan kredit yang diajukan oleh calon debitur, menurutnya semakin kaya informasi atau pengetahuan yang dikandung oleh data training maka akurasi Boosted Gradient Trees akan semakin meningkat [9]. Penelitian ini mengusulkan metode diagnosis preemtif untuk diabetes mellitus menggunakan Light Gradient Boosting Machine (LightGBM) untuk membantu pengenalan dini penyakit di negara-negara dengan keterbatasan ahli medis, seperti Ethiopia. Data dikumpulkan dari Zewditu Memorial Hospital di Addis Ababa. LightGBM dipilih karena kompleksitas komputasinya yang rendah dan kinerjanya yang unggul, dengan akurasi 98.1%, AUC 98.1%, sensitivitas 99.9%, dan spesifisitas 96.3%, melebihi model KNN, SVM, NB, Bagging, RF, dan XGBoost. Hasil ini menunjukkan bahwa model LightGBM efektif untuk diagnosis diabetes dan dapat membantu dalam mengenali dini penyakit di wilayah dengan sumber daya medis terbatas [10].

Penelitian ini bertujuan untuk menganalisis penggunaan algoritma Gradient Boosting dalam mempengaruhi masyarakat dalam memilih rumah sewa. Dengan menggunakan dataset prediksi sewa rumah, penelitian ini akan mengevaluasi akurasi dan performa model prediktif yang dibangun menggunakan algoritma tersebut. Diharapkan hasil penelitian ini dapat memberikan kontribusi yang signifikan dalam memprediksi harga sewa rumah secara lebih akurat dan efisien, serta membantu masyarakat dalam membuat keputusan yang lebih tepat terkait pemilihan rumah sewa.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining atau penambangan data, dapat diartikan sebagai suatu rangkaian proses yang dirancang untuk mengeksplorasi dan mengekstrak informasi yang berharga, pola, serta relasi yang tersembunyi dalam suatu basis data. Proses ini melibatkan penggunaan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk menganalisis dan memahami data dengan lebih dalam[11].

Data mining dapat pula diartikan sebagai suatu proses analisis yang dilakukan pada kumpulan data besar untuk mengidentifikasi pola, hubungan, atau informasi yang tidak terlihat secara langsung. Tujuannya adalah untuk menemukan pengetahuan baru atau wawasan yang dapat bermanfaat bagi pemilik data, baik dalam konteks bisnis, ilmu pengetahuan, atau sektor lainnya. *Data mining* digunakan dalam berbagai konteks, seperti analisis pasar, prediksi tren konsumen, identifikasi kecurangan, penelitian ilmiah, dan banyak lagi. Pemahaman yang mendalam tentang *data mining* memungkinkan organisasi atau individu untuk mengoptimalkan penggunaan data mereka dan membuat keputusan yang lebih informasional[12].

Data mining adalah bagian integral dari proses *Knowledge Discovery in Databases* (KDD). KDD adalah pendekatan sistematis untuk mengekstrak informasi yang berguna, pengetahuan, dan pola tersembunyi dari data yang besar dan kompleks. Ini melibatkan serangkaian langkah yang terkoordinasi untuk mengubah data mentah menjadi pengetahuan yang dapat digunakan untuk membuat keputusan yang lebih baik. Proses *Knowledge Discovery in Databases* (KDD) merupakan serangkaian langkah atau tahapan yang dilakukan untuk mengubah data mentah menjadi pengetahuan yang bermanfaat. Berikut adalah rangkuman umum dari tahapan dalam proses KDD:

a. Seleksi Data (*Selection*)

Memilih data yang relevan dari berbagai sumber untuk dimasukkan ke dalam proses KDD. Identifikasi sumber data, pilih data yang relevan untuk analisis.



b. Preprocessing Data (*Preprocessing*)

Merapikan dan membersihkan data untuk menghilangkan noise, menangani nilai yang hilang, dan mengubah format data. Pembersihan data, normalisasi, penanganan nilai yang hilang, transformasi data.

c. Penggalian Data (*Data Mining*)

Menerapkan teknik data mining untuk menemukan pola atau pengetahuan baru dari data yang telah diproses. Penggunaan algoritma data mining seperti klasifikasi, klusterisasi, asosiasi, atau regresi untuk menggali pengetahuan.

d. Evaluasi (*Evaluation*)

Menilai hasil dari langkah penggalian data untuk memastikan keandalan dan relevansi pola yang ditemukan. Pengukuran performa model, evaluasi signifikansi pola, validasi hasil.

e. Penginterpretasian (*Interpretation*)

Menginterpretasikan hasil penggalian data dan menafsirkan pola yang ditemukan. Analisis pola dan hubungan, penafsiran konteks bisnis, pembentukan pengetahuan baru.

f. Penggunaan (*Application*)

Menggunakan pengetahuan atau pola yang ditemukan untuk pengambilan keputusan atau perbaikan proses bisnis. Implementasi hasil, integrasi pengetahuan ke dalam sistem atau proses bisnis, pengambilan keputusan berbasis data.

Proses KDD bersifat iteratif, dan langkah-langkah tersebut dapat melibatkan revisi dan perbaikan berdasarkan hasil evaluasi dan interpretasi. Penting untuk memahami bahwa proses ini melibatkan kolaborasi antara komputer dan manusia, dengan manusia memainkan peran penting dalam interpretasi hasil dan membuat keputusan[13].

2.1 Gradient Boosting

Gradient boosting merupakan satu dari sekian algoritma *machine learning* yang termasuk kategori *ensemble learning* untuk menggabungkan model prediksi yang sederhana menjadi model prediksi yang lebih kuat atau lengkap[14]. *Gradient Boosting* adalah suatu metode dalam machine learning yang digunakan

untuk mengatasi masalah prediksi atau klasifikasi. Ini termasuk dalam kategori ensemble learning, di mana model yang kuat dibangun dari beberapa model yang lebih lemah.

Konsep dasar dari *Gradient Boosting* adalah membangun serangkaian model prediktif secara berurutan, dan setiap model mencoba untuk mengoreksi kesalahan prediksi model sebelumnya. Dalam konteks ini, "*gradient*" merujuk pada penyesuaian yang dilakukan pada model berdasarkan gradien (turunan) dari fungsi kerugian yang ingin dioptimalkan[4].

Algoritma *Gradient Boosting* dapat membentuk *decision tree* dengan penambahan jumlah iterasi untuk meningkatkan kinerja model lebih baik lagi. Adapun tahapan pengujian yang dilakukan[15]:

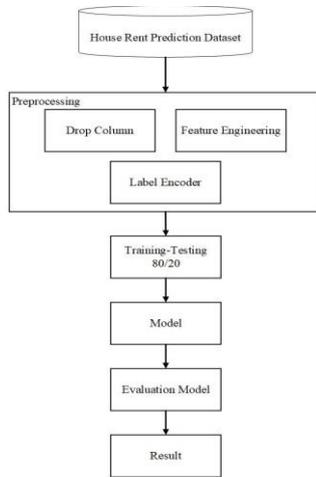
1. Menentukan dataset *training* D :
 $D = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$
2. Sampel T set elemen n berasal dari D (*replacement*)
 $D_1, D_2, \dots, \dots, \dots, D_T \rightarrow T$ dataset pelatihan
3. *Training* tiap
 $D_i, i=1, \dots, T$ dan urutan T output $F_1(x), \dots, F_T(x)$
4. *Aggregate classifier* dapat diterapkan dalam regresi dan klasifikasi dengan rumus:
Regresi: $f(x) = \sum_{i=1}^T F_i(x)$

$$\text{Klasifikasi: } f(X) = \text{sign} \left(\sum_{i=1}^T F_i(X) \right)$$

3. METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Tahapan penelitian merupakan kegiatan yang dilakukan untuk memperoleh pengetahuan baru atau memecahkan permasalahan dengan menggunakan pendekatan ilmiah, sistematis, dan logis. Proses ini melibatkan serangkaian tahapan yang dirancang untuk memastikan bahwa penelitian dilakukan dengan metode yang tepat dan dapat diandalkan[16]. Berikut merupakan langkah-langkah pelaksanaan penelitian yang diterapkan:



Gambar 1. Tahapan Penelitian

3.2. House Rent Prediction Dataset

Housing Rent Prediction Dataset adalah kumpulan data yang berisi informasi yang relevan untuk memprediksi harga sewa rumah atau properti. Dataset semacam ini sering digunakan dalam proyek-proyek machine learning atau analisis data untuk mengembangkan model yang dapat memperkirakan atau memprediksi biaya sewa berdasarkan berbagai fitur atau atribut[12].

Dalam penelitian ini data yang diterapkan bersumber dari repository kaggle (<https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset>). Data ini memiliki informasi 4745 rumah yang pernah disewa dan dataset tersebut memiliki 11 field diantaranya:

Tabel 1. Field Dataset House Rent Prediction

No.	Nama Atribut	Type	Keterangan
1	BHK	int64	Jumlah kamar, ruang tamu dan dapur
2	Rent	object	Harga Sewa
3	Size	object	Ukuran Rumah

			dalam persegi
4	Floor	object	Letak Lantai dan jumlah lantai
5	Area Type	object	Ukuran Rumah sewa yang dihitung berdasarkan Super Area, Carpet Area, Build Area
6	Area Locality	object	Lokasi rumah sewa
7	City	object	Kota dimana letak rumah sewa
8	Furnishing Status	object	Status furnishing rumah sewa baik Furnished, Semi-Furnished atau Unfurnished
9	Tenant Preferred	object	Jenis penyewa yang diutamakan untuk menyewa rumah sewa
10	Bathroom	int64	Jumlah kamar mandi
11	Point Of Contact	object	Siapa yang harus dihubungi jika ingin menyewa atau mengetahui info lebih lanjut terkait rumah sewa

Dari penjabaran informasi field yang diperoleh pada data diatas, maka berikut merupakan tampilan data yang akan diuji pada penelitian ini :



	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
0	2022-05-10	2	10000	1100	Ground out of 2	Super Area	Bundel	Kolkata	Unfurnished	Bachelors/Family	2	Contact Owner
1	2022-05-13	2	20000	800	1 out of 3	Super Area	Phool Bagan, Kankarguchi	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
2	2022-05-16	2	17000	1000	1 out of 3	Super Area	Sat Lake City Sector 2	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
3	2022-07-04	2	10000	800	1 out of 2	Super Area	Dumdum Park	Kolkata	Unfurnished	Bachelors/Family	1	Contact Owner
4	2022-05-09	2	7500	650	1 out of 2	Carpet Area	South Dum Dum	Kolkata	Unfurnished	Bachelors	1	Contact Owner
...
4740	2022-05-10	2	15000	1000	3 out of 5	Carpet Area	Bandam Kommu	Hyderabad	Semi-Furnished	Bachelors/Family	2	Contact Owner
4741	2022-05-15	3	28000	2000	1 out of 4	Super Area	Mankonda, Hyderabad	Hyderabad	Semi-Furnished	Bachelors/Family	3	Contact Owner
4742	2022-07-10	3	35000	1750	3 out of 5	Carpet Area	Himayath Nagar, NH 7	Hyderabad	Semi-Furnished	Bachelors/Family	3	Contact Agent
4743	2022-07-06	3	45000	1500	23 out of 34	Carpet Area	Gachibowli	Hyderabad	Semi-Furnished	Family	2	Contact Agent
4744	2022-05-04	2	15000	1000	4 out of 5	Carpet Area	Suchitra Circle	Hyderabad	Unfurnished	Bachelors	2	Contact Owner

Gambar 2. Isi Data House Rent Prediction

3.3. Preprocessing

1. Drop Column

Penelitian ini melakukan *drop column* untuk memilih beberapa kolom yang tidak akan digunakan sehingga kolom tersebut juga akan dihapus. Kolom yang akan dihapus antara lain *Posted On, Floor, Area Locality, Point of Contact*.

2. Feature Engineering

Feature engineering adalah proses mengubah atau meningkatkan fitur-fitur dalam dataset untuk meningkatkan kinerja model. Dalam hal ini, penelitian yang melakukan *binning* pada kolom *Rent* dan *Size* dengan mengambil nilai tengah dari data tersebut.

3. Label Encoder

Label Encoding adalah proses konversi nilai-nilai kategori menjadi angka-angka atau bilangan bulat. Ini umumnya digunakan pada kolom-kolom yang berisi data kategori untuk mempersiapkannya untuk pelatihan model *machine learning*.

3.3 Training Testing 80/20

Pada tahapan ini dilakukan *training testing* guna membagi data antara testing dan training, adapun besaran data yang dibagi yakni 80% data *training* sedangkan 20% merupakan data *testing*.

3.4 Model

Tahapan selanjutnya adalah data yang akan siap diolah diuji terlebih dahulu menggunakan model yakni *Gradient Boosting*.

3.5 Evaluation Model

Tahapan evaluasi dilaksanakan untuk mempresentasikan kinerja algoritma yang diterapkan, seperti *Confusion Matrix, Recall, Precision*, serta *Accuracy*. Selain itu penelitian ini menambahkan *Feature importance* untuk menganalisis fitur-fitur mana yang paling berpengaruh dalam proses kerja model.

4. HASIL DAN PEMBAHASAN

Sebelum dilakukan pemodelan, dataset yang terdiri dari 4744 entri dan 11 atribut ini telah akan melalui proses preprocessing. Proses ini meliputi penanganan nilai kosong cara mengidentifikasinya terlebih dahulu dengan fungsi `isnull()`. Berikut adalah hasil identifikasi data serta informasi tipe data pada dataset.

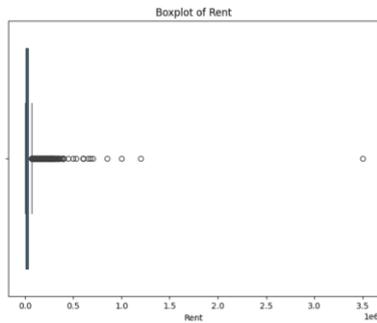
#	Column	Non-Null Count	Dtype
0	Posted On	4745 non-null	datetime64[ns]
1	BHK	4745 non-null	int64
2	Rent	4745 non-null	int64
3	Size	4745 non-null	int64
4	Floor	4745 non-null	object
5	Area Type	4745 non-null	object
6	Area Locality	4745 non-null	object
7	City	4745 non-null	object
8	Furnishing Status	4745 non-null	object
9	Tenant Preferred	4745 non-null	object
10	Bathroom	4745 non-null	int64
11	Point of Contact	4745 non-null	object

Gambar 3. Informasi Field Dataset

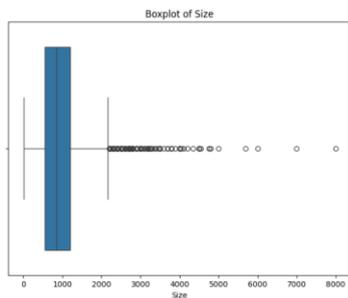
Terlihat pada Gambar 3 bahwa pada setiap *field* data tidak memiliki data kosong.

4.1. Proses Binning Data

Proses *binning* dilakukan untuk membuat rentang nilai tertentu. Proses *Binning* dapat meningkatkan performa model serta mengurangi risiko *overfitting*. Proses *binning* divalidasi dengan melihat *outliers* untuk melihat konsistensi antar atribut.



Gambar 4. Hasil Outlier pada Field Rent



Gambar 5. Hasil Outlier pada Field Size

Pada Gambar 4 dan Gambar 5 memperlihatkan tidak konsistennya data. Untuk menyeimbangkan data kita lakukan proses binning dengan mengambil nilai median dari dua field tersebut dengan menggunakan fungsi `describe()`.

Tabel 2. Hasil `describe()` field Rent dan Size

index	Rent	Size
std	78113.47070882423	634.2156387975206
min	1200.0	10.0
mean	34999.66701791359	967.6103266596417
max	3500000.0	8000.0
count	4745.0	4745.0
75%	33000.0	1200.0
50%	16000.0	850.0
25%	10000.0	550.0

Hasil yang ditampilkan pada Tabel 2 memperlihatkan bahwa nilai median yang bisa kita ambil untuk proses binning pada field Rent adalah 16.000. Sedangkan field Size, nilai median yang dapat diambil adalah 850.

Berikut adalah ketentuan yang akan diambil pada proses binning:

Tabel 3. Binning pada Kolom Rent dan Size

Rent	0 - 16000	Low
	16001-3500000	High

Size	0 – 850	Low
	851-8000	High

Ketika binning diterapkan, maka hasil data yang dihasilkan akan seperti gambar berikut:

```

housing['Rent_new'] = pd.cut(housing['Rent'], bins=[0,16000,3500000], labels=["Low", "High"])
housing['Size_new'] = pd.cut(housing['Size'], bins=[0,850,8000], labels=["Low", "High"])
housing.drop(['Rent','Size'], inplace=True, axis=1)
    
```

BHK	Area Type	City	Furnishing Status	Tenant Preferred	Bathroom	Rent_new	Size_new
0	2	Super Area	Kolkata	Unfurnished	Bachelors/Family	2	Low
1	2	Super Area	Kolkata	Semi-Furnished	Bachelors/Family	1	High
2	2	Super Area	Kolkata	Semi-Furnished	Bachelors/Family	1	High
3	2	Super Area	Kolkata	Unfurnished	Bachelors/Family	1	Low
4	2	Carpet Area	Kolkata	Unfurnished	Bachelors	1	Low
...
4740	2	Carpet Area	Hyderabad	Semi-Furnished	Bachelors/Family	2	Low
4741	3	Super Area	Hyderabad	Semi-Furnished	Bachelors/Family	3	High
4742	3	Carpet Area	Hyderabad	Semi-Furnished	Bachelors/Family	3	High
4743	3	Carpet Area	Hyderabad	Semi-Furnished	Family	2	High
4744	2	Carpet Area	Hyderabad	Unfurnished	Bachelors	2	Low

4745 rows x 8 columns

Gambar 6. Hasil Data Binning

4.2. Proses Label Encoder

Proses label encoder menjadi proses yang biasa dilakukan sebelum diproses oleh algoritma machine learning dalam mengubah nilai-nilai string atau kategori menjadi numerik. Proses ini dilakukan pada seluruh kolom. Berikut proses label encoder yang dilakukan pada data.

```

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
housing['BHK'] = le.fit_transform(housing['BHK'])
housing['Rent_new'] = le.fit_transform(housing['Rent_new'])
housing['Size_new'] = le.fit_transform(housing['Size_new'])
housing['Area Type'] = le.fit_transform(housing['Area Type'])
housing['City'] = le.fit_transform(housing['City'])
housing['Furnishing Status'] = le.fit_transform(housing['Furnishing Status'])
housing['Tenant Preferred'] = le.fit_transform(housing['Tenant Preferred'])
housing['Bathroom'] = le.fit_transform(housing['Bathroom'])
    
```

BHK	Area Type	City	Furnishing Status	Tenant Preferred	Bathroom	Rent_new	Size_new
0	1	2	4	2	1	2	1
1	1	2	4	1	1	0	0
2	1	2	4	1	1	0	0
3	1	2	4	2	1	0	1
4	1	1	4	2	0	0	1
...
4740	1	1	3	1	1	2	1
4741	2	2	3	1	1	3	0
4742	2	1	3	1	1	3	0
4743	2	1	3	1	2	2	0
4744	1	1	3	2	0	2	1

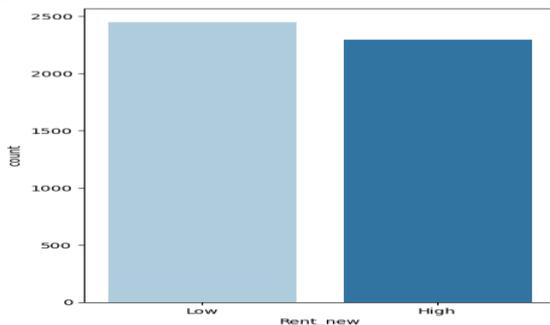
4745 rows x 8 columns

Gambar 7. proses dan hasil Label Encoder



4.3. Proses Pengujian Dataset

Prosess pengujian dataset menggunakan model *Gradient Boosting*. 4744 data dilakukan pembagian data sebesar 80% (3795 data) proses training data dan 20% (949 data) *testing data*. Kolom yang menjadi label yaitu *rent_new* yang sebelumnya dilakukan proses binning yang menentukan tinggi rendahnya harga sewa yang ditawarkan. Distribusi data yang didapatkan dari hasil binning pada kolom *rent_new* sebagai berikut.



Gambar 8 . Distribusi Data Pada Label *Rent_new*

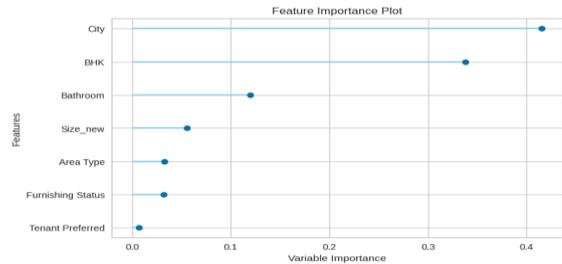
Pada diagram diatas dapat terlihat distribusi data masih berimbang. Hal ini dilakukan agar data tidak bias atau kecenderungan pada hasil. Hasil akhir dapat dilihat pada hasil model *Gradient Boosting* berikut.

Tabel 4. Hasil Model *Gradient Boosting*

<i>Accuracy</i>	Precisi on	<i>Recall</i>	<i>AUC</i>
84.38	83.33	87.53	92.65

Pada tabel diatas dapat dilihat hasil dari model *Gradient Boosting* memiliki performa yang cukup baik. Dengan hasil akurasi 84,38% , model dapat memprediksi keakuratan prediksi sebesar 84,38%. Selain itu , *AUC* menunjukkan hasil sebesar 92.65% *Gradient Boosting* memiliki kemampuan yang baik dalam memprediksi data positif dan negatif dengan akurasi yang tinggi. Hal ini juga berkaitan dengan *feature importance* dalam menentukan fitur apa yang berpengaruh dengan model yang

digunakan. Hal ini juga dapat mempengaruhi prediksi masyarakat dalam memilih rumah sewa untuk ditinggali. Berikut hasil *feature importance* dari model *Gradient Boosting*.



Gambar 9. Hasil *Feature Importance*

Hasil diatas memperlihatkan bahwa *feature* yang berpengaruh dalam model ini adalah *City* dan *BHK* yaitu lokasi dimana tempat rumah sewa ini disewakan serta jumlah kamar ruang tamu dan dapur dalam menentukan rumah sewa.

5. KESIMPULAN DAN SARAN

Dalam penelitian ini, *Gradient Boosting* menunjukkan performa yang menjanjikan dalam memprediksi harga sewa rumah. Dengan akurasi sebesar 84,38% dan *AUC* 92,65%, model ini mampu memprediksi harga dengan baik dan membedakan data positif dan negatif dengan akurasi tinggi. *Analisis feature importance* mengungkapkan bahwa lokasi (*City*) dan jumlah kamar (*BHK*) merupakan faktor paling berpengaruh, menunjukkan pentingnya kedua aspek tersebut dalam menentukan harga sewa. Secara keseluruhan, penelitian ini menunjukkan potensi *Gradient Boosting* sebagai alat bantu dalam memprediksi harga sewa rumah, terutama bagi masyarakat yang ingin memilih rumah sewa yang sesuai dengan budget dan kebutuhan mereka.

6. UCAPAN TERIMA KASIH

Penelitian ini dapat berjalan dengan baik karena bantuan dari berbagai pihak yang bersedia membantu penelitian ini. Khususnya kepada kedua orang tua yang senantiasa memberikan dukungan moril dan materiil, serta rekan-rekan peneliti yang telah bekerja sama dengan baik. Semoga penelitian ini dapat membawa manfaat sebagaimana fungsinya.



DAFTAR PUSTAKA:

- [1] A. S. Najmi and T. Fitriyani, "Permasalahan Urbanisasi Dalam Pembangunan Ekonomi Di Indonesia," *Work Pappers*, no. 2008204086, 2023, [Online]. Available: [http://repository.syekhnrurjati.ac.id/id/eprint/13001%0Ahttps://repository.syekhnrurjati.ac.id/13001/1/10_Pembangunan dan Urbanisasi - Copy.pdf](http://repository.syekhnrurjati.ac.id/id/eprint/13001%0Ahttps://repository.syekhnrurjati.ac.id/13001/1/10_Pembangunan%20dan%20Urbanisasi%20-%20Copy.pdf).
- [2] E. Situmorang and D. Purba, "Perancangan Aplikasi Pengujian Validitas dan Reliabilitas Instrumen Penelitian," vol. 2, pp. 54–58, 2019.
- [3] I. Mahendra and P. K. Putri, "Implementasi Metode Analytical Hierarchy Process (Ahp) Dalam Sistem Pendukung Keputusan Pembelian Rumah Di Kota Tangerang," *J. Teknoinfo*, vol. 13, no. 1, p. 36, 2019, doi: 10.33365/jti.v13i1.238.
- [4] R. Sun, G. Wang, W. Zhang, L. T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft Comput. J.*, vol. 86, no. xxxx, p. 105942, 2020, doi: 10.1016/j.asoc.2019.105942.
- [5] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient Boosting Feature Selection with Machine Learning Classifiers for Intrusion Detection on Power Grids," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 1, pp. 1104–1116, 2021, doi: 10.1109/TNSM.2020.3032618.
- [6] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [7] R. Dahlia and C. I. Agustyaningrum, "Perbandingan Gradient Boosting dan Light Gradient Boosting Dalam Melakukan Klasifikasi Rumah Sewa," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 6, pp. 1016–1020, 2022, doi: 10.32672/jnkti.v5i6.5460.
- [8] T. Z. Jasman, M. A. Fadhlullah, A. L. Pratama, and R. Rismayani, "Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 392–402, 2022, doi: 10.28932/jutisi.v8i2.4906.
- [9] A. Handayani, "Model Algoritma Boosted Gradient Trees untuk Penentuan Kelayakan Pemberian Kredit Koperasi," *Sci. Sacra J. Sains*, vol. 2, no. 2, pp. 300–305, 2022, [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>.
- [10] D. D. Rufo, T. G. Debelee, A. Ibenhal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (Lightgbm)," *Diagnostics*, vol. 11, no. 9, pp. 1–14, 2021, doi: 10.3390/diagnostics11091714.
- [11] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 846, 2019, doi: 10.30645/senaris.v1i0.91.
- [12] R. Sepriyanda, R. P. Sari, and Y. Fitriasia, "Sistem Computer Based Test Try Out Uji Kompetensi Ners Berbasis Website Menggunakan Metode Prototyping (Studi Kasus: Fakultas Ilmu Keperawatan Universitas Riau)," *J. Komput. Terap.*, vol. 5, no. 2, pp. 13–21, 2019, doi: 10.35143/jkt.v5i2.3337.
- [13] L. Muflikhah, D. E. Ratnawati, and R. regasari M. Putri, *Data Mining*. UB Press, 2018.
- [14] M. R. Supriadi and R. Andarsyah, *Deteksi Halaman Website Phishing Menggunakan Algoritma Machine Learning Gradient Boosting Classifier*. Bandung: Buku Pedia, 2023.
- [15] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification," *Techno.COM*, vol. 20, no. 3, pp. 400–410, 2021.
- [16] Lady Agustine and S. Seimahuira, "Penerapan Metode SAW dalam Analisa Perbandingan Performa Web server (Apache, Nginx, Lighttpd, Iis) pada Bahasa Pemrograman PHP," *Remik*, vol. 7, no. 1, pp. 409–420, 2023, doi: 10.33395/remik.v7i1.12075.