



Komparasi Penggunaan *Information Gain* Pada *Machine Learning* untuk Memprediksi Harga Rumah di Jabodetabek

Nabila Agustina Cahyani Putri^{1*}, Dede Brahma Arianto²

¹Sistem Informasi, Universitas Narotama, Surabaya, Indonesia

²Informatika, Universitas Islam Indonesia, Yogyakarta, Indonesia

Email Penulis Korespondensi: 1agustinacahyani5@gmail.com

Abstrak– Dengan kepadatan penduduk di Jabodetabek menyebabkan peningkatan kebutuhan akan rumah, baik digunakan untuk kebutuhan dasar maupun dijadikan sebagai investasi. Maka dari itu, perlu dilakukannya penelitian akan analisis dari harga rumah di Jabodetabek untuk mengurangi kerugian dalam berinvestasi. Tujuan dilakukannya penelitian ini adalah untuk memprediksi harga rumah di Jabodetabek sesuai dengan preferensi yang dimiliki oleh pembeli dan membantu para investor dalam memprediksi harga rumah agar dapat mengurangi risiko kerugian. Metode yang digunakan untuk analisis harga rumah di Jabodetabek adalah dengan membandingkan algoritma *multiple linear regression* dan *random forest regression* dengan menerapkan metode *information gain* untuk seleksi fitur. Kerangka kerja yang digunakan pada penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Hasil penelitian menunjukkan bahwa model terbaik adalah *random forest regression* dengan 10 variabel bebas yang didasarkan pada RMSE (*Root Mean Squared Error*) terendah yaitu sebesar 561272163 yang kemudian diaplikasikan ke dalam website dengan menggunakan *framework streamlit*. Meskipun demikian, hasil menunjukkan RMSE (*Root Mean Squared Error*) pada model *machine learning* masih cukup tinggi, sehingga disarankan untuk penelitian selanjutnya mencoba menerapkan model algoritma selain dari *multiple linear regression* dan *random forest regression*, serta mencoba menerapkan seleksi fitur selain metode *information gain* yang diharapkan dapat menurunkan nilai RMSE (*Root Mean Squared Error*) pada model sistem prediksi harga rumah di Jabodetabek.

Kata Kunci: Multiple Linear Regression, Random Forest Regression, Information Gain, Prediksi Harga Rumah, CRISP-DM.

Abstract– The population density in Jabodetabek causes an increase in the need for houses, both used for basic needs and used as an investment. For that reason, it is necessary to research the analysis of house prices in Jabodetabek to reduce the losses in investing. The purpose of this research is to predict house prices in Jabodetabek according to the preferences of buyers and help investors predict house prices to reduce the risk of losses. The method used to analyze house prices in Jabodetabek is the comparison of multiple linear regression and random forest regression algorithms by applying the information gain method for feature selection. This research uses the CRISP-DM (*Cross-Industry Standard Process for Data Mining*) framework. Results showed that the best model was random forest regression with 10 independent variables based on the lowest RMSE (*Root Mean Squared Error*) of 561272163 which was then applied to the website using the streamlit framework. However, the results show that the RMSE (*Root Mean Squared Error*) on the machine learning model is still quite high, so it is recommended for further research to try to apply algorithm models other than multiple linear regression and random forest regression and try to apply feature selection other than the information gain method which is expected to reduce the RMSE (*Root Mean Squared Error*) in the model of house price prediction system in Jabodetabek.

Keywords: Multiple Linear Regression, Random Forest Regression, Information Gain, House Price Prediction, CRISP-DM.

I. PENDAHULUAN

Jabodetabek adalah kawasan megapolitan yang memiliki jumlah penduduk yang banyak dengan pendidikan, perbankan, dan ekonomi lebih maju dibandingkan wilayah lain di Indonesia [1]–[3]. Kawasan megapolitan terbentuk dari beberapa wilayah metropolitan yang memiliki jumlah penduduk lebih dari 5 juta jiwa [4]. Pada tahun 2019 kawasan Jabodetabek memiliki kepadatan penduduk dengan rata-rata 4.523 jiwa/ km^2 [5]. Hal ini membuat orang lebih sadar akan kebutuhan dasar, termasuk tempat tinggal seperti rumah. [6]. Rumah adalah tempat untuk istirahat, berkumpul, berlindung, dan menyimpan barang, serta menunjukkan status sosial [7]. Rumah juga dapat dijadikan investasi jangka panjang [8]. Dalam investasi jangka panjang seperti rumah, diperlukan analisis harga rumah untuk mengurangi risiko kerugian. Analisis harga rumah juga dapat membantu pembeli untuk memprediksi harga berdasarkan preferensi yang dimiliki pembeli. Oleh karena itu, dibutuhkan sistem dengan penerapan pemodelan *machine learning*.

Dalam pemodelan prediksi harga rumah di Kawasan Jabodetabek dengan menggunakan *machine learning* diperlukan dataset yang memuat data-data akan kriteria rumah beserta harga rumah. Terdapat banyak kriteria dalam penentuan harga rumah, untuk memberikan hasil prediksi rumah yang lebih efektif diperlukan teknik seleksi fitur. Seleksi fitur adalah mengurangi dimensi atribut untuk mendapatkan atribut yang relevan dan tidak berlebihan [9]. Seleksi fitur yang dilakukan pada penelitian ini dilakukan dengan dua cara yaitu secara manual dan dengan menggunakan *information gain*.

Information gain adalah metode seleksi fitur ke dalam metode filter yang dapat menentukan fitur yang relevan dari sekumpulan fitur dataset dengan mengukur nilai *entropy* tiap fitur sebelum diperingkatkan sesuai dengan relevansi dalam menentukan kelas yang berbeda [10]. Dalam penggunaan *information gain* dapat menghasilkan nilai-nilai korelasi terhadap harga rumah. Setelah dilakukannya proses *information gain*, selanjutnya dilakukannya pemodelan dengan menggunakan algoritma *machine learning*.

Dalam penelitian untuk memprediksi harga rumah sebelumnya telah dilakukan oleh Linda Uswatun Hasanah, Inayatul Maula, dan Abu Tholib dengan judul Analisis Prediksi Harga Rumah di Jabodetabek Menggunakan *Multiple Linear Regression*. Penelitian tersebut menggunakan algoritma *Multiple Linear Regression* yang menghasilkan tingkat akurasi cukup baik sebesar 85% dengan MAE dan RMSE yang cukup tinggi [11].

Dalam penelitian sebelumnya juga untuk memprediksi harga rumah telah dilakukan oleh Rinabi Tanamal, Nathalia Minoque, Trianggoro Wiradinata, Yosua Soekamto, dan Theresia Ratih dengan judul *House Price Prediction Model Using Random Forest in Surabaya City*. Penelitian tersebut menggunakan algoritma *random forest* yang menghasilkan tingkat akurasi sebanyak 88% [12].

Penelitian terdahulu turut membahas akan prediksi harga rumah menggunakan algoritma *Multiple Linear Regression* yang dilakukan oleh Cep Haryanto, Nining Rahaningsih, dan Fadhil Muhammad Basysyar dengan judul *Komparasi Algoritma Machine Learning Dalam Memprediksi Harga Rumah*. Penelitian tersebut menggunakan algoritma *multiple linear regression* dan *random forest* yang menghasilkan algoritma *multiple linear regression* mendapatkan akurasi sebesar 78,5%, algoritma *random forest* mendapatkan 81,6% yang dapat disimpulkan model algoritma *random forest* lebih akurat pada penelitian ini [13].

Dalam penelitian terdahulu juga, terdapat penelitian yang menerapkan algoritma *random forest* menggunakan seleksi fitur yaitu metode *information gain* yang ditujukan untuk mengoptimasi akurasi algoritma *random forest*. Penelitian yang menerapkan algoritma *random forest* dan *information gain* sebagai seleksi fiturnya dilakukan oleh Taghfirul Azhima Yoga Siswa dan Wawan Joko Pranoto dengan judul *Implementasi Seleksi Fitur Information Gain Ratio Pada Algoritma Random Forest untuk Model Data Klasifikasi Pembayaran Kuliah*. Penelitian tersebut menghasilkan akurasi sebesar 78,47% [14].

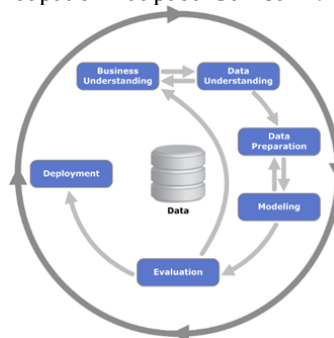
Pada penelitian ini algoritma yang akan dipakai adalah dengan menggunakan algoritma *multiple linear regression* dan *random forest regression*. *Multiple linear regression* merupakan teknik statistik yang digunakan untuk memodelkan hubungan antara variabel *dependen* dan variabel *independen* yang menggunakan lebih dari satu variabel *independen* [15], [16]. *Random forest regression* merupakan model regresi berdasar dengan teknik *supervised learning* menggunakan fitur dari *history data* yang menggabungkan prediksi dari gabungan prosedur pemecahan *machine learning* yang menghasilkan model prediksi yang akurat [17], [18].

Berdasarkan permasalahan dan penjelasan akan teori model *machine learning* dan teknik seleksi fitur di atas, penelitian ini secara garis besar akan membandingkan penerapan seleksi fitur dengan *information gain* dalam algoritma *multiple linear regression* dan *random forest regression* untuk memprediksi harga rumah di Jabodetabek. Penelitian ini bertujuan untuk memprediksi harga rumah sesuai dengan preferensi mereka dan juga

membantu para investor untuk lebih akurat memprediksi harga rumah, sehingga dapat mengurangi risiko kerugian.

II. METODE PENELITIAN

Dalam penelitian ini akan menggunakan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan proses penganalisaan data yang ditujukan untuk menghasilkan *insight* yang bermanfaat dalam pemecahan masalah dalam bisnis [19], [20]. Dalam kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) terdapat 6 proses tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment* [20]. Tahapan kerangka kerja pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja CRISP-DM

A. Business Understanding

Business understanding merupakan langkah awal CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang ditujukan untuk mengidentifikasi permasalahan dalam bisnis [21], [22].

B. Data Understanding

Data understanding merupakan langkah yang ditujukan untuk mengenali data yang meliputi tahapan pengumpulan data, pemahaman data, pendeskripsian data, pengidentifikasian data serta pengevaluasian kualitas data [20], [21].

C. Data Preparation

Data preparation merupakan langkah pengolahan data yang meliputi tahapan seperti pembersihan data, pemilihan data, memeriksa atribut-atribut dengan tujuan data siap digunakan pada tahapan selanjutnya yaitu tahap *modelling* [20], [21], [23].

D. Modelling

Pada tahap *modelling* akan membagi data menjadi 2 bagian yaitu data *training* dan data *testing* [23]. *Modelling* merupakan langkah penerapan algoritma *machine learning* pada data *training* yang telah diolah sebelumnya [20], [23].

E. Evaluation

Evaluation merupakan tahap yang berfokus pada model yang telah dilatih menggunakan algoritma *machine learning* pada tahapan sebelumnya yang akan dilakukan evaluasi terhadap kualitas model [21], [22].

F. Deployment

Deployment merupakan langkah terakhir dari kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang ditujukan untuk disebarluaskan yang dapat digunakan pengguna [22].

III. HASIL DAN PEMBAHASAN

A. *Business Understanding*

Permasalahan yang akan dipecahkan dalam kasus prediksi harga rumah di Jabodetabek sebagai berikut:

1. Meminimalisir kerugian investasi dalam sektor properti yaitu pada investasi pembelian rumah.
2. Memudahkan pembeli dalam memperkirakan harga rumah sesuai dengan keinginan mereka.

B. *Data Understanding*

1. Pengumpulan Data

Data dikumpulkan pada penelitian ini berjenis data sekunder. Data sekunder merupakan informasi yang dikumpulkan bukan dari sumber utama, tetapi informasi dikumpulkan melalui sumber perantara [24], [25]. Data sekunder yang digunakan pada penelitian ini diperoleh dari website Kaggle.com.

2. Penjelasan Kolom

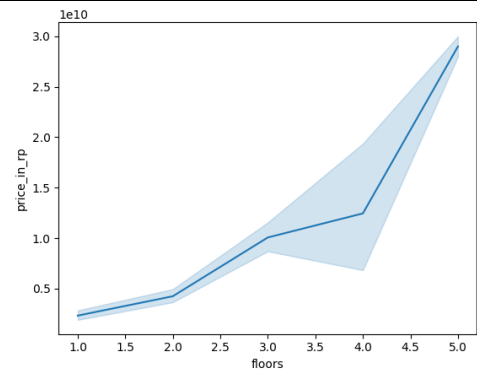
Data yang telah diambil dari website Kaggle.com dengan atribut dan keterangan yang dijelaskan pada Tabel 1 berikut:

Tabel 1. Penjelasan Kolom

Atribut	Keterangan
url	Berisikan data unique dari tautan halaman detail penjualan rumah yang berada pada website rumah123.com.
price_in_rp	Berisikan data harga rumah.
title	Berisikan judul rumah yang disajikan pada website rumah123.com.
address	Berisikan keterangan alamat dari rumah yang akan dijual.
district	Berisikan keterangan kabupaten rumah yang akan dijual.
city	Berisikan keterangan kota rumah yang akan dijual.
lat	Berisikan keterangan latitude berdasarkan lokasi kabupaten.

long	Berisikan keterangan longitude berdasarkan lokasi kabupaten.
facilities	Berisikan fasilitas yang terdapat pada rumah.
property_type	Berisikan type properti.
ads_id	Berisikan id unique ads.
bedrooms	Berisikan jumlah kamar tidur di rumah.
bathrooms	Berisikan jumlah kamar mandi di rumah.
land_size_m2	Berisikan luas dari tanah.
building_size_m2	Berisikan luas dari bangunan rumah.
carports	Berisikan luas carport.
certificate	Berisikan sertifikat yang dilampirkan pada proses jual beli rumah.
electricity	Berisikan tentang kapasitas tegangan listrik pada rumah.
maid_bedrooms	Berisikan jumlah kamar tidur untuk pembantu.
maid_bathrooms	Berisikan jumlah kamar mandi untuk pembantu.
floors	Berisikan jumlah lantai pada rumah.
building_age	Berisikan usia bangunan.
year_built	Berisikan tahun pembuatan rumah.

property_condition	Berisikan keterangan kondisi properti.
building_orientation	Berisikan orientasi gedung.
garages	Berisikan luas garasi.
furnishing	Berisikan keadaan perabotan di rumah.

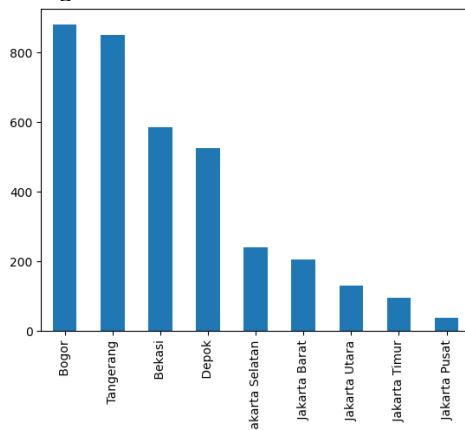


Gambar 3. Harga Rumah Berdasarkan Jumlah Lantai Rumah

3. Eksplorasi data

Eksplorasi data bertujuan untuk memahami komponen dan konten data yang digunakan, dengan tujuan mendapatkan informasi penting [26].

a. Histogram



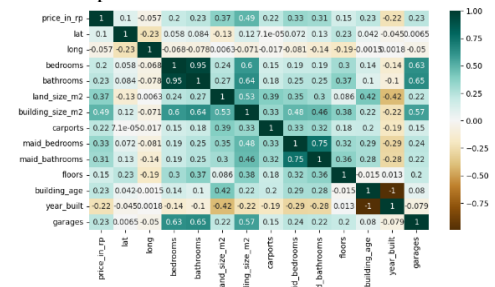
Gambar 2. Persebaran Data Berdasarkan Kota

Pada diagram histogram, seperti ditunjukkan pada Gambar 2, dapat dilihat bahwa data yang digunakan dalam penelitian ini memiliki jumlah data terbanyak di kota Bogor, mencapai 881 data.

b. Line Plot

Pada diagram line plot, seperti ditunjukkan pada Gambar 3, terlihat bahwa harga rumah cenderung meningkat seiring dengan peningkatan jumlah lantai rumah.

c. Heatmap



Gambar 4. Hubungan Korelasi Setiap Kolom Numeric

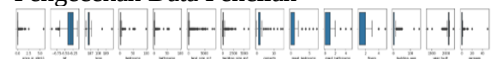
Pada diagram heatmap, seperti ditunjukkan pada Gambar 4, terlihat bahwa korelasi tertinggi dengan target kolom 'price_in_rp' adalah kolom 'building_size_m2' dengan nilai sebesar 0.49.

4. Pengecekan Kualitas Data

a. Pengecekan Duplikasi Data

Data yang dipakai pada penelitian ini tidak memiliki duplikasi data.

b. Pengecekan Data Pencilan



Gambar 5. Pengecekan Data Pencilan

Pada diagram barplot, seperti yang terlihat pada Gambar 5, data yang digunakan dalam penelitian ini menunjukkan adanya nilai pencilan pada beberapa kolom, seperti 'price_in_rp,' 'lat,' 'long,' 'bedrooms,' 'bathrooms,' 'land_size_m2,' 'building_size_m2,' 'carports,' 'maid_bedrooms,' 'maid_bathrooms,' 'floors,' 'building_age,' 'year_built,' dan 'garages'.

c. Pengecekan Nilai NaN

Data yang dipakai pada penelitian ini memiliki nilai NaN pada kolom *property_type*, *ads_id*, *bedrooms*, *bathrooms*, *land_size_m2*, *building_size_m2*, *certificate*, *floors*, *building_age*, *year_built*, *property_condition*, *building_orientation*, dan *furnishing*.

C. Data Preparation

1. Penghapusan Kolom Manual

Penghapusan kolom secara manual ini didasarkan karena terdapat kolom yang memiliki nilai yang sama dengan kolom lainnya, nilai kolom hanya berisi nilai identifikasi unik, dan kolom memiliki nilai NaN lebih dari 40%. Oleh karena itu, perlu dilakukan penghapusan kolom secara manual. Kolom yang dihapus meliputi kolom *url*, *title*, *address*, *district*, *lat*, *long*, *property_type*, *ads_id*, *building_age*, *year_built*, *building_orientation* dan *facilities*.

2. Menghapus Nilai NaN dan Data *Outlier*

Langkah selanjutnya adalah penghapusan nilai NaN dan data *outlier* yang telah teridentifikasi dalam tahap *data understanding*. Dalam tahapan ini sebanyak 33,60% atau sebanyak 1194 baris data terhapus yang menyisakan 2359 baris data.

3. Manipulasi Data

Langkah selanjutnya adalah merubah data yang memiliki tipe data kategorik menjadi numerik dengan proses *encoding*. Kolom yang diubah dari tipe data kategorik menjadi numerik adalah kolom *city*, *certificate*, *property_condition* dan *furnishing*. Manipulasi data yang dilakukan pada penelitian ini juga dilakukan pada kolom *electricity* dengan menghapus keterangan mah dan merubah tipe datanya menjadi *float*.

4. Penghapusan Duplikat

Setelah data dilakukan penghapusan kolom manual, penghapusan nilai NaN, penghapusan data outlier dan manipulasi data memunculkan data duplikasi pada dataset sebanyak 435 data yang selanjutnya dihapus dan menyisakan 1814 baris data.

5. Seleksi Fitur dengan *Information Gain*

Setelah dilakukan penghapusan data kolom secara manual, data penelitian memiliki 15 kolom yang terdiri atas 14 kolom fitur dan 1 kolom target. Selanjutnya diterapkan metode seleksi fitur *information gain* yang didasarkan pada pengurutan nilai *entropy* tinggi ke rendah yang menghasilkan 10 jenis dataset baru dengan varian jumlah fitur dari 5 fitur sampai dengan 14 fitur.

6. Scaling data

Semua variansi dataset baru selanjutnya dilakukan *scaling* dengan metode *standard scaller*.

D. Modelling

Tahapan ini dilakukan penerapan model algoritma *machine learning* yang telah ditetapkan sebelumnya yaitu algoritma *multiple linear regression* dan *random forest regression* yang bertujuan untuk memprediksi harga rumah pada dataset yang telah terbentuk setelah kegiatan *data preparation*. Pada tahap modelling ini semua dataset akan dibagi menjadi 80% data *training* dan 20% data *testing*.

E. Evaluation

Tahapan ini proses evaluasi dari hasil yang dilakukan pada tahap sebelumnya yaitu *modelling* dengan memilih model dengan RMSE (*Root Mean Squared Error*) terendah dari keseluruhan model yang terbentuk pada tahap sebelumnya.

model	Linear Regression	Random Forest
rmse_5_feature	719313354	722167151
rmse_6_feature	680287852	578565389
rmse_7_feature	672085786	568533992
rmse_8_feature	674486151	574505066
rmse_9_feature	673517731	565289974
rmse_10_feature	671411012	561272163
rmse_11_feature	671666888	566459657
rmse_12_feature	672702524	580398725
rmse_13_feature	669234836	571067021
rmse_14_feature	668780035	565074051

Gambar 6. Hasil Evaluasi Model

Dari hasil yang terlihat pada Gambar 6, dapat disimpulkan bahwa model terbaik adalah model algoritma *random forest regression* dengan 10 variabel bebas, yang memiliki RMSE (*Root Mean Squared Error*) sebesar 561,272,163. Evaluasi model algoritma *random forest regression* dengan 10 variabel bebas secara detail dapat dilihat pada Gambar 7.

```
print('rmse: ', round(np.sqrt(mean_squared_error(y_test, y_pred))))
print('mse: ', round(mean_squared_error(y_test, y_pred)))
print('mae: ', round(mean_absolute_error(y_test, y_pred)))
print('mape: ', mean_absolute_percentage_error(y_test, y_pred))
print('r2: ', r2_score(y_test, y_pred))

rmse: 561272163
mse: 315026440571417216
mae: 329673439
mape: 0.23848362397897027
r2: 0.8465141937802718
```

Gambar 7. Evaluasi Model Random Forest Regression dengan 10 Variabel Bebas

F. Deployment

Tahap ini melibatkan implementasi model dengan RMSE (*Root Mean Squared Error*) terendah agar dapat digunakan oleh pengguna dengan bahasa pemrograman Python menggunakan *framework Streamlit*. Detail

implementasi dapat dilihat pada Gambar 8.



Gambar 8. Website Prediksi Harga Rumah di Jabodetabek

IV. KESIMPULAN

Berdasarkan hasil penelitian terkait dengan komparasi penggunaan *information gain* pada *machine learning* untuk memprediksi harga rumah di Jabodetabek menggunakan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dapat disimpulkan bahwa pada model *machine learning multiple linear regression* nilai RMSE (*Root Mean Squared Error*) terendah pada dataset yang memiliki 14 variabel bebas yaitu sebesar 668780035 dan pada model *machine learning random forest regression* nilai RMSE (*Root Mean Squared Error*) terendah pada dataset yang memiliki 10 variabel bebas yaitu sebesar 561272163.

Prediksi harga rumah menggunakan model *multiple linear regression* dan *random forest regression* menunjukkan bahwa model terbaik adalah *random forest regression* dengan 10 variabel bebas yang didasarkan pada RMSE (*Root Mean Squared Error*) terendah yaitu sebesar 561272163. Prediksi harga rumah menggunakan model *random forest regression* dengan 10 variabel bebas menghasilkan MSE (*Mean Squared Error*) sebesar 315026440571417216, MAE (*Mean Absolute Error*) sebesar 329673439, MAPE (*Mean Absolute Percentage Error*) sebesar 0.23848362397897027 dan r^2 (*R-squared*) sebesar 0.8465141937802718.

Model prediksi yang telah dipilih dari RMSE (*Root Mean Squared Error*) terendah dari semua dataset yang tersedia dengan model *multiple linear regression* dan *random forest regression* yaitu model *random forest regression* dengan dataset yang memiliki 10 variabel bebas selanjutnya diaplikasikan ke dalam website dengan menggunakan *framework streamlit*. Penelitian yang telah dilakukan dengan model *machine learning* masih kurang bagus untuk memprediksi harga rumah di Jabodetabek karena memiliki RMSE (*Root Mean Squared Error*) yang cukup tinggi.

Dengan hasil yang menunjukkan RMSE (*Root Mean Squared Error*) yang cukup tinggi, saran untuk dilakukan pada penelitian selanjutnya adalah dengan mencoba menerapkan model algoritma selain dari *multiple linear regression* dan *random forest regression* serta mencoba menerapkan seleksi fitur selain metode *information gain* yang diharapkan bisa menurunkan nilai RMSE (*Root Mean*

Squared Error) pada model sistem prediksi harga rumah di Jabodetabek.

REFERENSI

- [1] E. D. Viana, F. Febrianti, and F. R. Dewi, "Literasi Keuangan, Inklusi Keuangan dan Minat Investasi Generasi Z di Jabodetabek," *Jurnal Manajemen dan Organisasi*, vol. 12, no. 3, pp. 252–264, Jan. 2022, doi: 10.29244/jmo.v12i3.34207.
- [2] N. D. Permata and others, "Evaluasi Fungsi Pemanfaatan Hutan Kota Sebagai Kawasan Rekreasi Masyarakat Jabodetabek," IPB University, 2019.
- [3] F. Febyanti, "Pemodelan faktor-faktor yang mempengaruhi harga rumah di Jabodetabek menggunakan metode regresi probit," *Jurnal Riset Statistika*, pp. 51–57, 2022.
- [4] L. Agustino, F. Manan, and I. Akbar, *Laporan Penelitian BNN RI-IKoTAN 2019*. BNN RI, 2019.
- [5] InfoJabodetabek.com, "Luas Wilayah dan Jumlah Penduduk Jabodetabek," InfoJabodetabek.com.
- [6] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *Journal of Applied Computer Science and Technology*, vol. 4, no. 1, pp. 58–64, Jul. 2023, doi: 10.52158/jacost.v4i1.491.
- [7] S. Utomo and T. Mardiono, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN RUMAH PADA PERUMAHAN MENGGUNAKAN METODE ANALYTICAL HIERARCHY PROCESS (STUDI KASUS: KEC. NGAMPRAH KAB. BANDUNG BARAT)," *FIKI (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 1, pp. 47–59, May 2018.
- [8] M. A. Verdiansyah and Suwanda, "Penerapan Metode Regresi Komponen Utama Kernel untuk Prediksi Harga Rumah," *Bandung Conference Series: Statistics*, vol. 3, no. 2, pp. 653–661, Aug. 2023, doi: 10.29313/bcss.v3i2.9084.
- [9] S. H. A. Aini, Y. A. Sari, and A. Arwan, "Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 2546–2554, 2018.
- [10] R. A. Azizah, F. Bachtiar, and S. Adinugroho, "Klasifikasi Kinerja Akademik Siswa Menggunakan Neighbor Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 3, pp. 605–614, Jun. 2022, doi: 10.25126/jtiik.2022935751.
- [11] I. Maula, L. U. Hasanah, and A. Tholib, "ANALISIS PREDIKSI HARGA RUMAH DI JABODETABEK MENGGUNAKAN MULTIPLE LINEAR REGRESSION," *Jurnal*

- Informatika Kaputama (JIK)*, vol. 7, no. 2, pp. 216–224, 2023.
- [12] R. Tanamal, N. M. K. S. Rasyid Jr, T. Wiradinata, Y. S. Soekanto, and T. R. D. Saputri, “House Price Prediction Model Using Random Forest in Surabaya City,” 2023.
- [13] C. Haryanto, N. Rahaningsih, and F. M. Basysyar, “KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI HARGA RUMAH,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 533–539, 2023.
- [14] W. J. Pranoto, “IMPLEMENTASI SELEKSI FITUR INFORMATION GAIN RATIO PADA ALGORITMA RANDOM FOREST UNTUK MODEL DATA KLASIFIKASI PEMBAYARAN KULIAH,” 2023.
- [15] H. Akbar, M. Ayomi, and Y. Haryanto, “ANALISIS PERBANDINGAN MODEL MACHINE LEARNINGDALAM PREDIKSI SUHU PERMUKAAN LAUT MENGGUNAKAN DATA MODEL REANALYSIS ECMWF,” *Prosiding Seminar Nasional Hasil Penelitian Kelautan dan Perikanan*, pp. 96–100, Mar. 2023.
- [16] A. K. Baskara, A. Nazir, M. Irsyad, Y. Yusra, and F. Insani, “Implementasi Data Mining Memprediksi Penjualan Crude Palm Oil Berdasarkan Kapasitas Tangki Menggunakan Multiple Linear Regression,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 3, p. 493, Mar. 2023, doi: 10.30865/json.v4i3.5665.
- [17] S. Fachid and A. Triayudi, “Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 68, Jan. 2022, doi: 10.30865/mib.v6i1.3492.
- [18] Diana Tri Susetianingtiyas, Eka Patriya, and Rodiah, “Model Random Forest Regression Untuk Peramalan Penyebaran Covid-19 Di Indonesia,” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 2, no. 2, pp. 84–95, Sep. 2022, doi: 10.51454/decode.v2i2.48.
- [19] D. U. Iswavigra, L. E. Zen, H. Hanim, and others, “Marketing Strategy UMKM Dengan CRISP-DM Clustering & Promotion Mix Menggunakan Metode K-Medoids,” *Jurnal Informasi dan Teknologi*, pp. 45–54, 2023.
- [20] Y. Yudiana, A. Y. Agustina, N. Khofifah, and others, “Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan,” *Indonesian Journal of Islamic Economics and Business*, vol. 8, no. 1, pp. 1–20, 2023.
- [21] H. Halawa, O. J. Harmaja, W. sandi Hulu, and others, “Implementasi Algoritma K-Means Clustering Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Pulo Brayan,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 150–157, 2023.
- [22] A. Rizkiawan and T. Wahyudi, “IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MEMBER BARU MENGGUNAKAN LINEAR REGRESSION PADA PT. GSI,” *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, vol. 6, no. 1, pp. 118–126, 2023.
- [23] M. Fajar, D. Mahdiana, A. Diana, and G. Triyono, “PERAMALAN HARGA SAHAM UBER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE,” in *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2023, pp. 304–308.
- [24] S. Sutisna and M. N. Yuniar, “Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve bayes,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 243–246, 2023.
- [25] R. Rasiban, S. Bila, and T. Triwahyudi, “Pemilihan Rute Terbaik Dari Lokasi Calon Pelanggan Ke Terminal Akses Menggunakan Metode A* Star PT. Bahtera Anugrah Electical,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 331–335, 2023.
- [26] T. M. Khalil and S. Sahid, “Klasifikasi informasi hoaks pada media sosial twitter menggunakan algoritma random forest berbasis particle swarm optimization,” *Jurnal Kajian dan Terapan Matematika*, vol. 8, no. 3, pp. 199–209, 2022.