

Analisis Penggunaan Algoritma Random Forest Regressor untuk Prediksi Harga Properti di Kebayoran Baru

Adhelio Reyhandro¹, Aura Lintang Pembayun Pinasti Aji², Fritz Filemon Siregar³, Sultan Adyatma Rangga Setiawan⁴

Sistem Informasi, Universitas Multimedia Nusantara, Tangerang, Indonesia

¹adhelio.reyhandro@student.umn.ac.id, ²aura.lintang@student.umn.ac.id,

³fritz.filemon@student.umn.ac.id, ⁴sultan.adyatma@student.umn.ac.id

Abstrak—Banyak properti di Kebayoran Baru dijual dengan berbagai harga atas berbagai faktor. Model prediksi akan diciptakan menggunakan algoritma Random Forest Regressor yang berdasarkan dengan metodologi CRISP-DM. Data di filter berdasarkan kecamatan Kebayoran Baru serta dihapus nilai kosong dan outlier pada setiap atribut. Perancangan model diawali dengan tidak merubah parameter agar dapat melakukan hyperparameter tuning pada tahap post-pruning. Model over fit dengan nilai akurasi data latih 0.98 dan dilakukan tuning parameter. Kriteria yang diuji sebagai parameter untuk model adalah Squared Error dan Friedman MSE. Parameter lain juga diuji, seperti Max Depth, Min Sample Split, dan Number of Estimator. Hasil yang didapat merupakan sebuah model prediksi yang cukup akurat dengan akurasi pada test set 0.90 dan train set sebesar 0.92, yang menandakan bahwa model tidak over fit seperti sebelumnya. Model di-deploy dengan menggunakan Streamlit yang berbahasa Python melalui Github untuk dijadikan sebuah aplikasi web yang dapat diakses secara massal.

Kata Kunci—CRISP-DM; Harga Properti; Kebayoran Baru; Random Forest Regressor.

I. PENDAHULUAN

Tiap tahun, harga properti akan selalu mengalami kenaikan, terutama pada harga properti di Kebayoran Baru. Beberapa faktor yang menyebabkan peningkatan harga properti adalah jumlah penduduk yang bertambah dan lokasi yang strategis seperti berpusat pada keramaian Jakarta serta mendapatkan lokasi yang strategis dalam mengakses berbagai fasilitas [1]. Walaupun tiap tahun harga properti mengalami kenaikan, tidak membuat masyarakat menurunkan minat untuk membeli properti di Kebayoran Baru. Berdasarkan data yang diambil dari situs web Rumah.com, permintaan masyarakat terhadap rumah diatas 1 miliar meningkat sebanyak 62% di kuartal pertama tahun 2023. Beberapa alasan meningkatnya peminat properti di Kebayoran Baru walaupun memiliki harga yang cukup fantastis, meliputi lokasi properti di Kebayoran baru strategis karena mudah untuk mengakses pusat perbelanjaan; pusat bisnis; serta fasilitas umum, keamanan yang

terjamin, dan dapat dijadikan sebagai investasi jangka panjang [2].

Pada penelitian ini peneliti mengambil data properti di daerah Kebayoran Baru dari situs web *open source* Kaggle yang terdiri dari 12 kolom dan 594708 baris. Tetapi, hanya 7 kolom dan 4460 baris yang digunakan untuk memprediksikan harga properti di Kebayoran Baru. Di dalam data tersebut mencakup data tentang harga, lokasi, ukuran, dan fitur properti, serta faktor-faktor yang mempengaruhi permintaan dan harga properti di Kebayoran Baru. Pada penelitian ini digunakan juga metode Random Forest Regressor untuk memprediksi harga properti [3]. Dari latar belakang tersebut, rumusan masalah dari penelitian ini adalah: (1) Bagaimana sebuah model prediksi harga properti dapat membantu pembeli atau penjual properti Kebayoran Baru?; (2) Apakah criterion Squared Error lebih baik dalam memprediksi harga properti Kebayoran baru dibandingkan dengan Friedman MSE?; (3) Bagaimana akurasi algoritma Random Forest Regressor dalam memprediksi harga properti di Kebayoran Baru berdasarkan atribut-atribut yang tersedia?.

II. KAJIAN LITERATUR

A. Random Forest Regression

Random Forest Regression merupakan algoritma yang memiliki lebih dari satu *decision tree*. Pada random forest, *decision tree* bergantung kepada nilai yang didapatkan dalam satu model vektor acak pada sampel independen [4]. Subset acak dari data pelatihan akan diproses oleh *decision tree* pada random forest agar prediksi dapat dihasilkan.

Squared error adalah metrik yang digunakan untuk mengukur kesalahan atau perbedaan antara nilai prediksi dan nilai sebenarnya dalam analisis statistik atau pembelajaran mesin. Ini diperoleh dengan mengkuadratkan selisih antara nilai prediksi dan nilai sebenarnya, dan kemudian menjumlahkan kesalahan-kesalahan kuadrat tersebut [5].

$$MSE = \frac{\sum_{t=1}^n (At - Ft)^2}{n}$$

MSE = Mean Squared Error

At = Nilai aktual

Ft = Nilai hasil prediksi

n = Banyaknya data

Friedman MSE (Mean Squared Error) adalah metrik evaluasi yang digunakan dalam pembelajaran mesin untuk mengukur kualitas prediksi dalam masalah regresi. Ini adalah variasi dari squared error yang mempertimbangkan perbedaan antara prediksi dan rata-rata prediksi dari beberapa model yang berbeda [6].

B. Penelitian Pendahulu

Terdapat dua penelitian pendahulu yang menjadi landasan dari penelitian ini untuk menjawab rumusan masalah yang terbentuk dalam kasus harga properti Kebayoran Baru. Penelitian pertama merupakan literatur nasional sebagai referensi dalam menjawab kasus nasional, dan penelitian kedua merupakan literatur internasional sebagai referensi dalam pembentukan model prediksi dengan algoritma Random Forest Classifier.

Penelitian pertama merupakan karya Haryanto dengan judul “Komparasi Algoritma Machine Learning dalam Memprediksi Harga Rumah”. Pada penelitian nasional ini, digunakan data properti dengan atribut yang serupa, namun tidak dengan ketinggian tanah. Penelitian ini memiliki tujuan untuk membandingkan algoritma multiple linear regression dengan random forest regressor. Hasil yang didapat dari penelitian ini merupakan dua model prediksi yang dimana random forest regressor lebih unggul dengan nilai akurasi 81,6% bila dibandingkan dengan multiple linear regression yang menghasilkan nilai akurasi 78,5% [7].

Penelitian kedua merupakan karya Afonso dengan judul “Housing Prices Prediction with a Deep Learning and Random Forest Ensemble”. Pada penelitian internasional ini, digunakan data properti yang berisikan atribut seperti jumlah lantai, luas tanah, luas bangunan, jumlah kamar, serta gambar properti yang ada di Brazil. Penelitian ini memiliki tujuan untuk memprediksi harga dengan dua model algoritma berbeda yang digabungkan, yaitu random forest regressor dan bidirectional LSTM, atau Long Short-Term Memory. Kesimpulan yang didapat untuk model dengan random forest regressor adalah algoritma random forest regressor bekerja lebih baik dengan atribut-atribut numerik dan tidak dapat menangani raw data gambar ataupun teks, sehingga untuk data gambar dan teks digunakan model LSTM [8].

III. METODE PENELITIAN

A. Metodologi

Harga properti dipengaruhi oleh beberapa faktor, meliputi lokasi properti, ukuran dan tipe rumah, dan kondisi properti. Faktor tersebut dapat digunakan untuk menentukan harga properti dengan pengolahan data yang menggunakan metode CRISP-DM.

Metode CRISP-DM [9] memiliki 6 tahapan yang dilakukan dan dimulai dari pemahaman bisnis (*business understanding*) yaitu memahami tujuan dari pembuatan bisnis yang akan dilakukan; pemahaman data (*data understanding*) yaitu mengeksplorasi data, seperti pengumpulan data yang dibutuhkan dan memahami struktur dari data untuk mendapatkan kualitas data yang terjamin; pra proses data [10] (*data preparation*), pada tahap ini data yang telah dieksplorasi akan diidentifikasi dan dilakukan *cleaning data* sehingga format terbaru akan terbentuk tanpa ada nilai yang *missing* atau ambigu; pemodelan (*modeling*), data akan diterapkan ke dalam algoritma yang telah ditentukan sehingga pola dan struktur dalam data dapat diketahui. Pemilihan model umumnya ditentukan berdasarkan tipe data dan pada penelitian ini model yang digunakan adalah Random Forest Regressor; evaluasi (*evaluation*), model yang telah dipilih untuk data akan diuji menggunakan beberapa metrik evaluasi sehingga menghasilkan performansi model yang akurat. Metrik evaluasi yang digunakan pada penelitian ini adalah R^2 , MAE, MSE, dan RMSE; penerapan (*deployment*), tahap penerapan dapat dilakukan dengan membuat laporan atau aplikasi setelah kelima tahap telah selesai dilakukan.

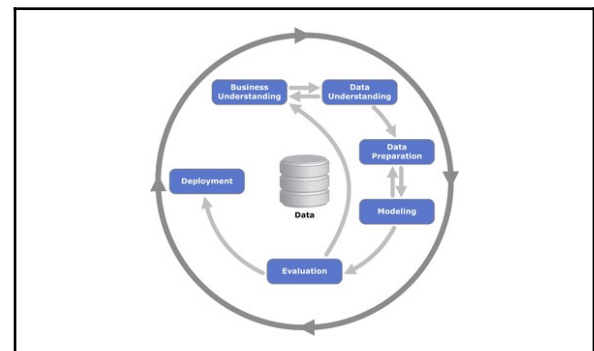


Fig. 1. Proses CRISP-DM

Tahap evaluasi dilakukan setelah tahap *business understanding* hingga *modeling* selesai dilakukan. R^2 Score, Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE) digunakan sebagai pengukur kesalahan serta selisih antara nilai prediksi dan nilai aktual sementara. R^2 (Coefficient Determination) berfungsi sebagai pengukur jarak antara hasil prediksi dengan nilai aktual, Mean Absolute Error (MAE) berfungsi agar nilai prediksi dan nilai aktual dapat terukur dengan rata-rata kesalahan absolut yang terhitung, Mean Squared Error (MSE) berfungsi untuk menghitung rata-rata dari selisih yang dikuadratkan dan lebih

sensitif terhadap kesalahan besar, serta *Root Mean Squared Error* (RMSE) berfungsi sebagai penghitung skala kesalahan pada penelitian [11].

Tahap deployment dilakukan pada penelitian dengan menggunakan framework berbasis web, yaitu Streamlit. Streamlit digunakan untuk mengimplementasikan model pembelajaran mesin yang menggunakan bahasa pemrograman Python, sehingga peneliti memilih untuk menggunakan Streamlit untuk deploy model prediksi harga properti di Kebayoran Baru menjadi sebuah aplikasi web yang dapat diakses secara massal dengan mudah [12].

B. Sumber Data

Data yang akan digunakan merupakan data properti di seluruh Indonesia dengan judul “property price” yang diambil melalui Kaggle [13]. Data properti diambil pada periode bulan November tahun 2022. Data harga properti terdiri dari 12 atribut, yaitu “price”, “buildingSize”, “landSize”, “bedRooms”, “bathRooms”, “garages”, “z”, “province”, “city”, “district”, “latitude”, dan “longitude”, serta lebih dari 500.000 baris data.

Populasi dan sampel dari data yang akan diteliti merupakan “district” atau kecamatan dalam Kebayoran Baru untuk menyesuaikan dengan hipotesis yang telah dibentuk pada latar belakang penelitian ini. Didapat sebesar 4.460 baris data properti yang ada di kecamatan Kebayoran Baru dari data yang kita gunakan.

C. Alur Kerja

Menggunakan metodologi CRISP-DM dan data yang telah kami kumpulkan, data akan dirapikan pada tahap pra proses data serta melakukan filter terhadap kecamatan Kebayoran Baru. Terdapat dua alur proses pengerjaan, yaitu alur proses pelatihan model dan alur proses prediksi harga properti untuk pengguna dari model kami.

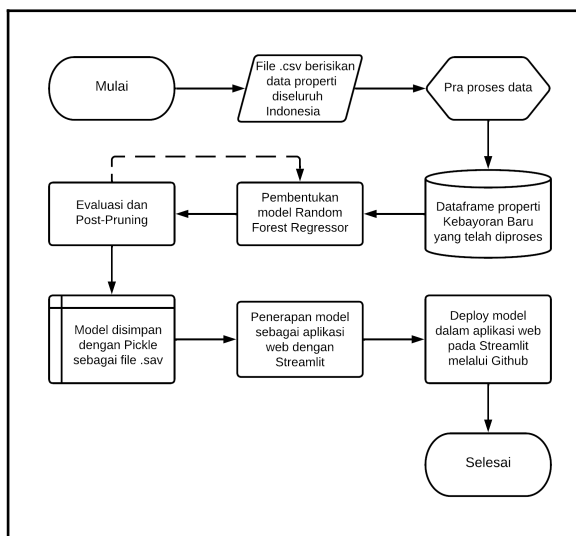


Fig. 2. Alur Proses Latih

Pada Figur 2, merupakan diagram alur dari proses pelatihan model prediksi harga properti. Proses ini diawali dari memasukkan data properti di Indonesia yang berbentuk .CSV. Data lalu diproses sehingga menghasilkan dataframe properti di Kebayoran Baru. Lalu, model prediksi dibentuk menggunakan random forest regressor dengan parameter *default*, atau belum diubah. Model akan dievaluasi dan dilakukan hyperparameter tuning serta dilatih ulang untuk menyesuaikan dengan data yang diberi. Model dievaluasi kembali, lalu diekspor dengan pickle sehingga menjadi sebuah file .SAV. Kemudian dibentuk aplikasi web menggunakan Python dan Streamlit untuk dimasukkan model prediksi yang telah dibuat. Terakhir, aplikasi web dengan model prediksi di-deploy pada Streamlit melalui Github agar dapat diakses secara publik.

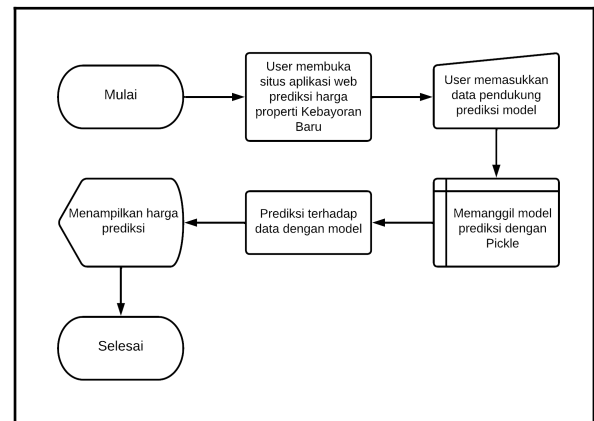


Fig. 3. Alur Proses Prediksi

Ketika model sudah diterapkan sebagai aplikasi web dan di-deploy secara massal, Figur 3 merupakan representasi dari diagram alur proses prediksi sebagai *end-user*. Dimulai dari membuka situs aplikasi web model prediksi kami lalu memasukkan data yang diminta pada aplikasi web. Lalu, akan dipanggil model prediksi dengan menggunakan Pickle untuk memprediksi harga dari data properti yang diberi dan menampilkannya pada bagian bawah dari situs aplikasi web.

IV. HASIL DAN PEMBAHASAN

Proses analisis penggunaan algoritma Random Forest Regressor untuk prediksi harga properti di Kebayoran Baru akan menggunakan tahap-tahap dari metodologi CRISP-DM sebagai berikut:

A. Business Understanding

Penelitian ini membutuhkan informasi tentang pasar properti yang ada di Kebayoran Baru. Dibutuhkan berbagai atribut atau faktor yang mempengaruhi harga properti di daerah tersebut. Dengan analisis menggunakan algoritma Random Forest Regressor, jurnal ini akan menyediakan pemahaman yang lebih baik tentang harga properti di Kebayoran Baru.

B. Data Understanding

Dalam penelitian ini, data yang digunakan merupakan data yang mencatat harga-harga properti di Indonesia. Data yang digunakan memiliki lebih dari 500.000 baris data properti. Data ini memiliki atribut seperti pada tabel I.

TABEL I. TABEL PENJELASAN ATRIBUT DATA PROPERTI

| Atribut | Penjelasan |
|--------------|--|
| price | Nilai jual properti dalam rupiah |
| buildingSize | Luas bangunan meter persegi |
| landSize | Luas tanah dalam meter persegi |
| bedRooms | Jumlah kamar tidur |
| bathRooms | Jumlah kamar mandi |
| garages | Jumlah tempat parkir mobil |
| z | Ketinggian tanah dari permukaan laut dalam dekameter |
| province | Provinsi dari lokasi properti |
| city | Kota dari lokasi properti |
| district | Kecamatan dari lokasi properti |
| latitude | Nilai garis lintang bumi |
| longitude | Nilai garis bujur bumi |

Langkah selanjutnya adalah melakukan visualisasi terhadap missing values dengan menggunakan heatmap. Visualisasi dilakukan agar peneliti dapat melihat atribut yang memiliki mayoritas missing values.



Fig. 4. Missing Values pada keseluruhan data

Setelah itu peneliti melakukan visualisasi terhadap outlier dengan menggunakan boxplot pada setiap atribut.

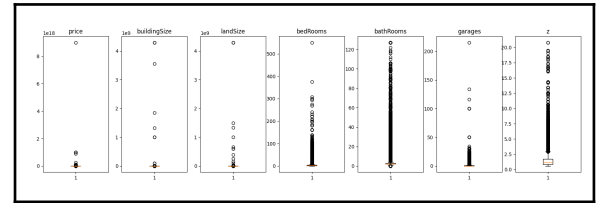


Fig. 5. Outliers pada keseluruhan data

Setelah melakukan visualisasi terhadap outlier, tahap selanjutnya adalah melakukan pra proses data atau *data preprocessing*. Pada tahap tersebut, terdapat penggantian nama atribut, penghapusan NaN, dan pemilihan kecamatan yang akan di prediksi harga propertinya.

C. Data Preprocessing

Langkah pertama pada tahap Data Preprocessing mengubah nama kolom “z” menjadi “landHeight”. Setelah nama kolom diubah, langkah yang dilakukan adalah menghapus missing values dan menggunakan visualisasi heatmap.

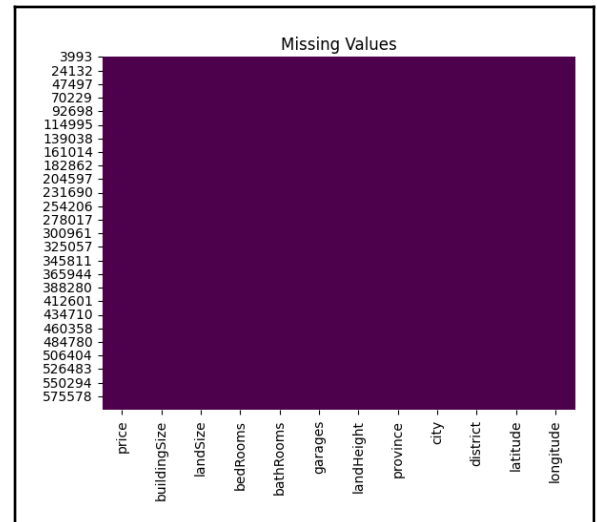


Fig. 6. Penghapusan Missing Values

Langkah selanjutnya adalah memilih kecamatan yaitu Kebayoran Baru serta melakukan filter terhadap atribut-atribut yang diperlukan dan menghapus atribut-atribut yang tidak diperlukan. Data yang digunakan memiliki atribut seperti pada tabel II.

TABEL II. TABEL ATRIBUT YANG DIGUNAKAN

| Atribut | Penjelasan |
|--------------|----------------------------------|
| price | Nilai jual properti dalam rupiah |
| buildingSize | Luas bangunan meter persegi |

| | |
|------------|--|
| landSize | Luas tanah dalam meter persegi |
| bedRooms | Jumlah kamar tidur |
| bathRooms | Jumlah kamar mandi |
| garages | Jumlah tempat parkir mobil |
| landHeight | Ketinggian tanah dari permukaan laut dalam dekameter |

| | |
|-----------------------|-------|
| MSE | 0.00 |
| RMSE | 0.06 |
| Accuracy of Train Set | 0.989 |
| Accuracy of Test Set | 0.930 |

Setelah data berhasil difilter dengan memilih kecamatan Kebayoran Baru sebagai data yang akan digunakan, langkah selanjutnya adalah memvisualisasikan outlier dengan menggunakan boxplot pada setiap atributnya.

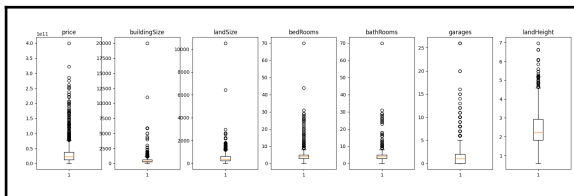


Fig. 7. Outliers pada Data Kebayoran Baru

Setelah outlier terlihat, peneliti menghapus outlier dengan menggunakan metode IQR pada semua atribut yang memiliki outlier. Setelah outlier terhapus peneliti membuat visualisasi outlier boxplot kembali untuk melihat hasil boxplot setelah outlier dihapus.

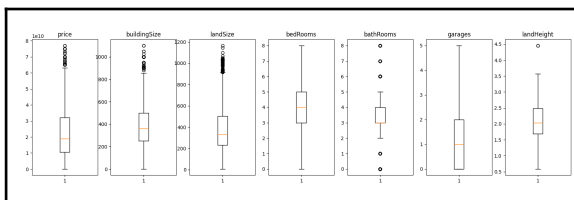


Fig. 8. Penghapusan Outliers

Hasil pembentukan data yang telah melalui proses penghapusan outlier, kemudian dilakukan *data scaling* menggunakan *Min Max Scaler*. *Data scaling* ini bertujuan agar nilai data dibatasi dari 0 hingga 1.

D. Modeling

Pemodelan pembelajaran mesin menggunakan data train dengan jumlah 2438, dan test sebesar 607, algoritmanya Random Forest Regressor tanpa mengubah parameter dan menghasilkan skor-skor seperti pada tabel III.

TABEL III. TABEL METRIK EVALUASI DAN AKURASI PREDIKSI MODEL

| | Nilai |
|----------------------|-------|
| R ² Score | 0.930 |
| MAE | 0.02 |

Hasil yang didapat dari model *Random Forest Regressor* tanpa tuning parameter sangat baik dengan R² skor yang tinggi dan skor MAE, MSE, serta RMSE yang sangat rendah, namun bila model diberi data latih untuk diprediksi, tingkat akurasi sangat tinggi. Hal ini bisa terjadi karena model over fit, dimana model bukan mempelajari dari data latih namun menghafalkan data latih. Maka, tahap selanjutnya akan dilakukan evaluasi lebih lanjut dengan tuning parameter.

E. Evaluation

Langkah pertama dalam tahap evaluasi adalah membuat tabel antara harga asli dengan harga prediksi. Pada Tabel IV, menunjukkan bahwa antara harga asli dengan harga prediksi memiliki selisih yang sedikit dan bahkan terdapat data dengan harga yang sama.

TABEL IV. TABEL EVALUASI SEBELUM PRUNING

| Price | Predicted |
|----------|-----------|
| 0.519243 | 0.518983 |
| 0.584210 | 0.584210 |
| 0.870066 | 0.864738 |
| 0.324342 | 0.220135 |
| 0.239885 | 0.203281 |

Visualisasi pohon dibuat dengan menggunakan salah satu estimator dari model Random Forest Regressor. Berdasarkan figur 9, visualisasi pohon menunjukkan bahwa *branch* terlalu dalam dan kompleks, hal ini dapat menjadi penyebab over fit terhadap model.

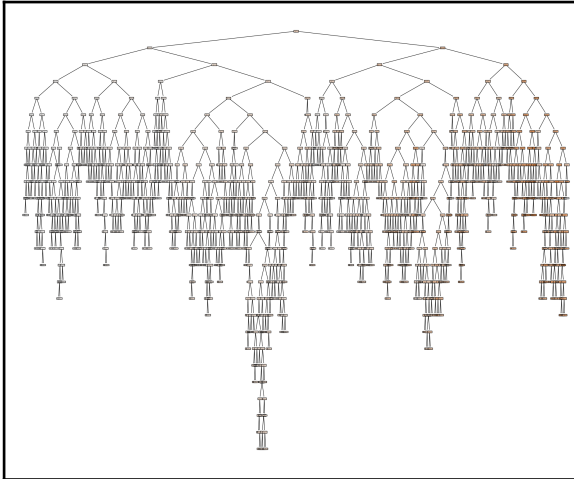


Fig. 9. Visualisasi Pohon Estimator

Melakukan *Grid Search Cross Validation* dengan parameter berupa *Criterion* Friedman MSE dan *squared_error*; *Max Depth* dengan range 4, 6, dan 8; untuk *Minimum Sample Split* berupa range 50, 100, dan 150; *n_estimator* dengan range 100, 150, dan 200. Tampilan hasil dari proses *Grid Search Cross Validation* menampilkan ranking berupa:

TABEL V. TABEL EVALUASI SEBELUM PRUNING

| Rank | Criterion | max_depth | min_samples_split | n_estimators | Score |
|------|---------------|-----------|-------------------|--------------|----------|
| 1 | friedman_mse | 8 | 50 | 100 | 0.896314 |
| 2 | squared_error | 8 | 50 | 100 | 0.896276 |
| 3 | friedman_mse | 8 | 50 | 150 | 0.896249 |
| 4 | squared_error | 8 | 50 | 150 | 0.896219 |
| 5 | friedman_mse | 8 | 50 | 200 | 0.896046 |

Setelah itu dilakukan pemodelan kembali dengan parameter yang terbaik, hal tersebut dilakukan agar menghasilkan model yang tidak *overfitting*. Pada tabel VI dapat terlihat bahwa akurasi training menurun cukup besar dan akurasi test menurun sedikit.

TABEL VI. TABEL METRIK EVALUASI DAN AKURASI PREDIKSI MODEL DENGAN PARAMETER TERBAIK

| | Nilai Sebelum | Nilai Setelah |
|-----------------------|---------------|---------------|
| R ² Score | 0.930 | 0.902 |
| MAE | 0.02 | 0.03 |
| MSE | 0.00 | 0.00 |
| RMSE | 0.06 | 0.07 |
| Accuracy of Train Set | 0.989 | 0.920 |
| Accuracy of Test Set | 0.930 | 0.902 |

bahwa, data harga prediksi mengalami perubahan setelah dilakukan pruning dengan selisih harga yang berbeda tetapi tidak mengalami *overfitting*.

TABEL VII. TABEL EVALUASI SETELAH PRUNING

| Harga | Prediksi Sebelum | Prediksi Setelah |
|----------|------------------|------------------|
| 0.519243 | 0.518983 | 0.493328 |
| 0.584210 | 0.584210 | 0.579944 |
| 0.870066 | 0.864738 | 0.846247 |
| 0.324342 | 0.220135 | 0.199467 |
| 0.239885 | 0.203281 | 0.167641 |

Visualisasi pohon menggunakan parameter terbaik hasil *grid search* berupa Friedman MSE dengan jumlah estimasi 100, minimal pembagian data 50, dan kedalaman 8, menghasilkan visualisasi Random Tree Forest sebagai berikut.

Data harga prediksi sebelum dilakukan pruning akan dibandingkan dengan harga prediksi setelah dilakukan pruning. Data pada Tabel VII menunjukkan

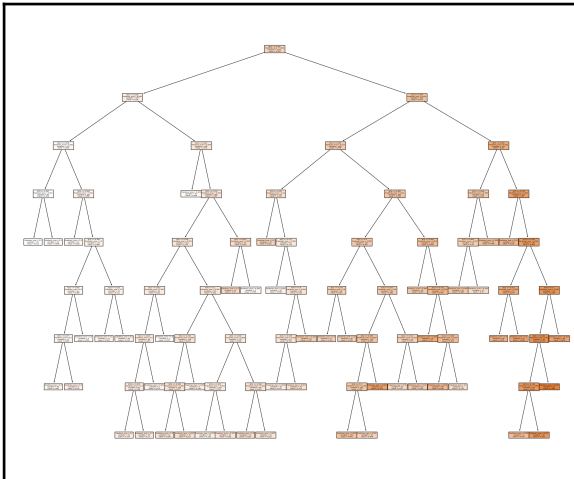


Fig. 10. Visualisasi pohon estimator setelah pruning

F. Deployment

Setelah mendapatkan model prediksi harga yang bagus, maka selanjutnya adalah mengeksport model serta scaler untuk pembentukan aplikasi web dengan python Streamlit. Source code python untuk deployment diunggah ke Github repository untuk di publish secara massal melalui Streamlit. Tampilan akhir yang didapat ada di figur 9.

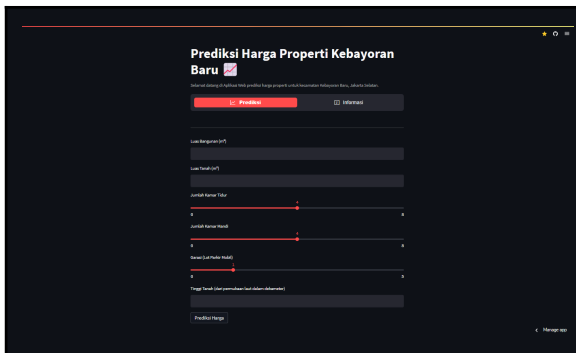


Fig. 11. Tampilan Aplikasi Web model prediksi

Penggunaan model dapat dilakukan dengan memasukkan atribut-atribut yang diperlukan untuk model membuat hasil prediksi harga properti. Data yang kami uji merupakan salah satu rumah yang sedang dijual di Kebayoran Baru dengan harga asli Rp 45 miliar seperti pada figur 12 [14] dan dengan ketinggian tanah diantara 25 hingga 26 meter melalui situs mapcoordinates.net seperti pada figur 13.

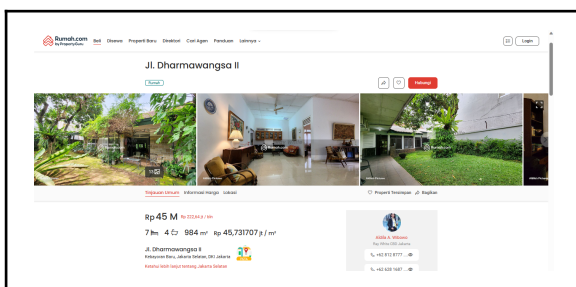


Fig. 12. Situs rumah.com untuk data properti asli untuk pengujian model



Fig. 13. Situs mapcoordinates.net untuk ketinggian tanah

Setelah mengetahui data dan mengisi seluruh atribut pada aplikasi web yang telah kami rancang, maka akan ada output harga dibagian bawah seperti pada figur 14.

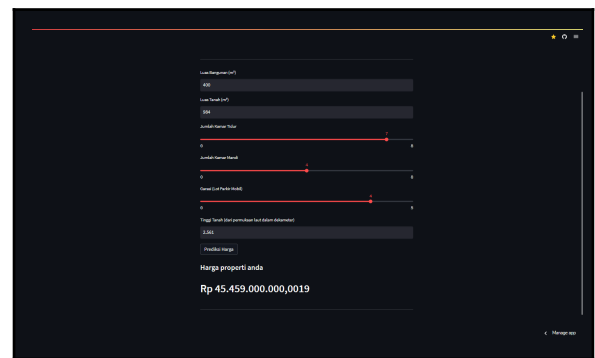


Fig. 14. Hasil prediksi model pada Aplikasi Web

Hasil prediksi harga properti dari aplikasi web kami tidak berbeda jauh dengan harga asli yang tertera pada Rumah.com, hal ini menandakan bahwa model yang dibuat sudah baik dan telah diterapkan secara benar dengan Github dan Streamlit. Aplikasi web ini juga dapat diakses secara massal melalui tautan pada Lampiran A.

V. KESIMPULAN

Penelitian ini menganalisis harga properti pada kecamatan Kebayoran Baru dengan menggunakan algoritma Random Forest Regressor. Model algoritma dilatih dan dibentuk melalui proses *grid search cross validation*, yang menghasilkan parameter *criterion* terbaik berupa Friedman MSE untuk melakukan prediksi harga properti. Hasil *grid search* ini memotong akurasi model, dimana akurasi awal berupa 0.93 dengan status model *over fit* menjadi 0.90 dan memperbaiki model yang *over fit* tadi. Model setelah itu di-deploy menggunakan *Streamlit* melalui Github. Dengan menggunakan aplikasi web Stremlit yang telah dibentuk pada penelitian, pembeli dapat menampilkan prediksi harga properti yang diinginkan berdasarkan dengan atribut yang terikat pada properti.

Penelitian ini juga tidak jauh dari berbagai kekurangan. Model tidak dapat memprediksi berdasarkan mebel yang ada di properti. Selain itu, harga properti juga selalu meningkat tiap tahunnya, maka model harus dilatih ulang apabila ada perubahan tren harga properti secara signifikan. Kami berharap pada penelitian yang melanjutkan dari topik

ini dapat membawakan hasil yang lebih baik dengan menggunakan data yang lebih lengkap serta algoritma dan parameter yang lebih tepat.

PERAN DAN TANGGUNG JAWAB

TABEL VIII. TABEL PERAN DAN TANGGUNG JAWAB

| Nama | Peran |
|-----------------------------------|--|
| Adhelio Reyhandro | Meninjau literatur pendahulu, mengerjakan laporan, pra proses data, pemodelan, evaluasi, dan deployment. |
| Aura Lintang Pembayun Pinasti Aji | Mengerjakan laporan, pra proses data, pemodelan, evaluasi, dan ppt. |
| Fritz Filemon Siregar | Mengerjakan laporan, pra proses data, pemodelan, evaluasi, dan ppt. |
| Sultan Adyatma Rangga Setiawan | Mengerjakan laporan, pra proses data, pemodelan, evaluasi, dan deployment. |

REFERENSI

- [1] Administrator, "Harga Rumah (Memang) Terus Naik, Jabodetabek Tertinggi," Persatuan Perusahaan Realestat Indonesia, <https://www.rei.or.id/newrei/berita-harga-rumah-memang-terus-naik-jabodetabek-tertinggi.html> (accessed May 29, 2023).
- [2] A. R. Nurdifa, "Wow! Harga Rumah di Atas rp1 miliar makin diminati, Ini Alasannya," Bisnis.com, <https://ekonomi.bisnis.com/read/20230519/47/1657314/wow-harga-rumah-di-atas-rp1-miliar-makin-diminati-ini-alasannya> (accessed May 29, 2023).
- [3] D. J. Crosss Sihombing, D. C. Othernima, J. Manurung, and J. R. Sagala, "Comparative models of price estimation using multiple linear regression and random forest methods," 2023 *International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, 2023. doi:10.1109/iccosite57641.2023.10127705
- [4] N. KURNIAWATI, A. NOVITRI, and Y. K. NINGSIH, "Prediksi channel gain threshold untuk modulasi ADAPTIF V2V Menggunakan ALGORITMA random forest regression," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 3, p. 544, 2022. doi:10.26760/elkomika.v10i3.544
- [5] U. Azmi, Z. N. Hadi, and S. Soraya, "ARDL method: Forecasting data Curah Hujan Harian NTB," *Jurnal Varian*, vol. 3, no. 2, pp. 73–82, 2020. doi:10.30812/varian.v3i2.627
- [6] D. Kurniasih, S. Mariani, and Sugiman, "Efisiensi Relatif Estimator Fungsi Kernel Gaussian Terhadap Estimator Polinomial Dalam Peramalan Usd Terhadap Jpy," *UNNES Journal of Mathematics*, vol. 2, Nov. 2013.
- [7] 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, 2023. doi:https://doi.org/10.36040/jati.v7i1.6343
- [8] [1] B. Afonso, L. Melo, W. Oliveira, S. Sousa, and L. Berton, "Housing prices prediction with a deep learning and Random Forest Ensemble," *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 2019. doi:10.5753/eniac.2019.9300
- [9] Sylvia Jane A. Sumarauw and M. G. Maukar, "Fuzzy c-Means Clustering untuk Pengenalan PolaStudi kasus Data Saham," *Jurnal Axioma: Jurnal Matematika dan Pembelajaran*, vol. 7, pp. 97–106, Jul. 2022.
- [10] A. Pambudi, Z. Abidin, and Permata, "Penerapan Crisp-Dm Menggunakan Mlrk-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm)(Studi Kasus: Bursaefek Indonesia Tahun 2015-2022)," *Jurnal Data Mining dan Sistem Informasi*, vol. 4, pp. 1–14, 2023.
- [11] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, Mae, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, <https://peerj.com/articles/cs-623/> (accessed May 29, 2023).
- [12] A. Putranto, N. L. Azizah, and I. R. I. Astutik, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023. Accessed: May 29, 2023. [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/180>
- [13] HEYINGXU, "Property Price," Kaggle, <https://www.kaggle.com/datasets/yingxuhe/property-price> (accessed Jun. 4, 2023).
- [14] A. A. Wibowo, "Rumah dijual JL. Dharmawangsa II, 400 m2 mulai RP 45 m | rumah.com," Rumah.com by PropertyGuru, <https://www.rumah.com/listing-properti/dijual-jl-dharmawangsa-ii-oleh-aldila-a-wibowo-21652750> (accessed May 29, 2023).

LAMPIRAN

- A. Prediksi Harga Properti Kebayoran Baru. Kelompok 5 Mata Kuliah Data Modelling Kelas A-HY. <https://4dhelio-model-kebayoranbaru-rfr-kebayoran-baru-8a9dvc.streamlit.app/>

TAMBAHAN: Halaman ini (9) dapat dipotong untuk menyesuaikan dengan template IEEE.

B. Hasil Turnitin melalui UMN Library

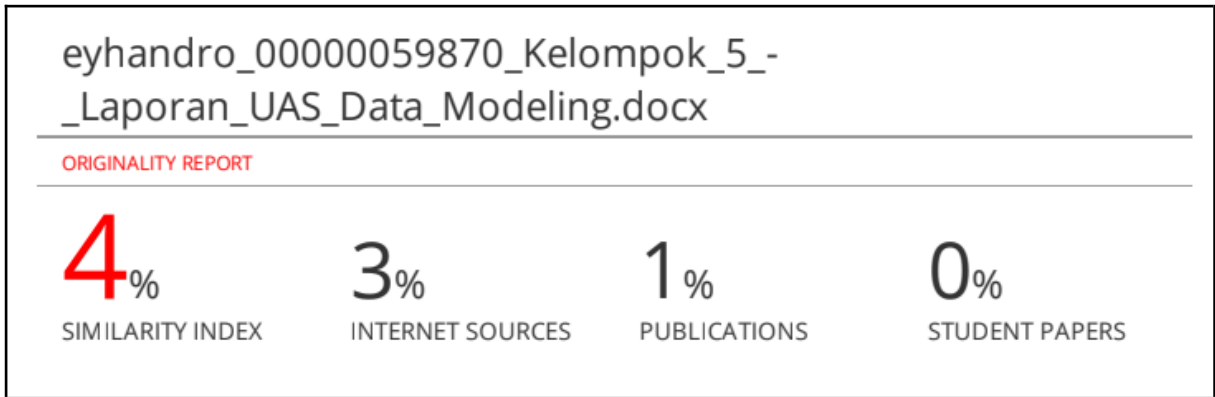


Fig. 15. Hasil Turnitin UMN Library

Penelitian ini telah diuji plagiarisme dengan Turnitin yang disediakan oleh UMN Library melalui email. Nilai plagiarisme yang didapat pada penelitian ini adalah 4%, dengan 3% dari sumber Internet dan 1% dari publikasi. Berikut adalah tautan untuk laporan penuh hasil Turnitin dalam bentuk .pdf terhadap penelitian ini: <https://drive.google.com/file/d/1VXad3N4cLoNtqdxOp14exLkZYgq2ph/view?usp=sharing>

C. Hasil AI GPT melalui ZeroGPT

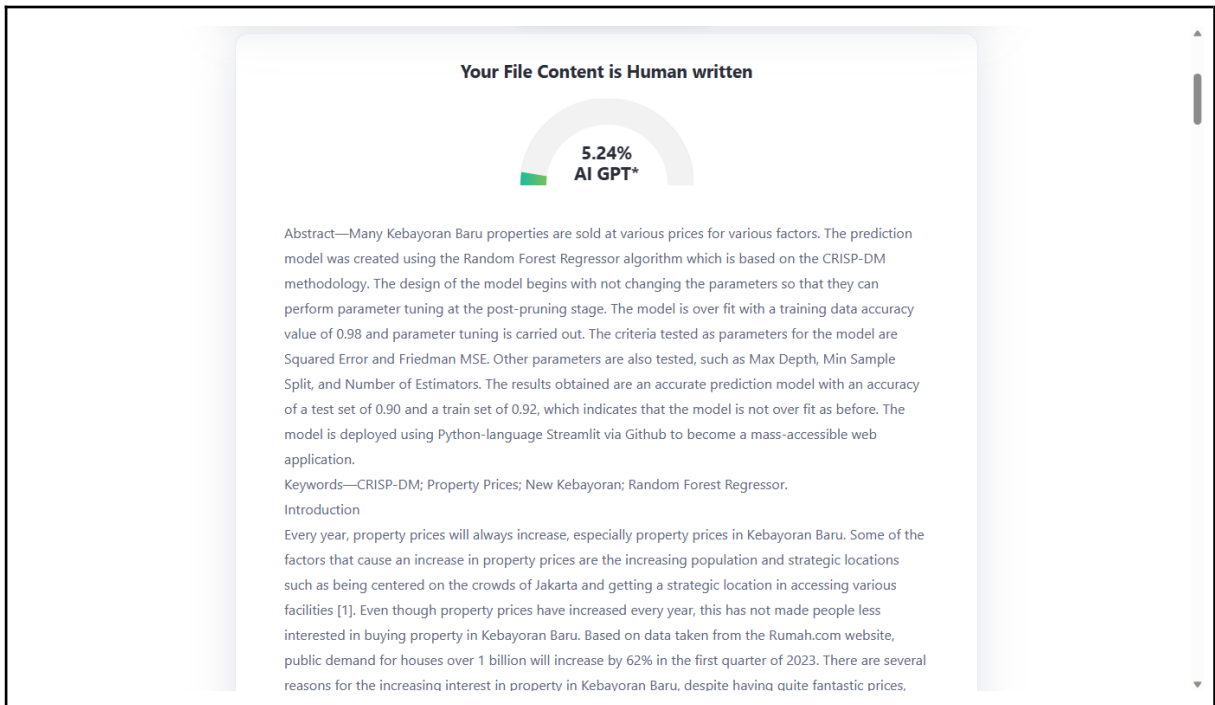


Fig. 16. Hasil ZeroGPT

Penelitian ini juga telah dilakukan pengujian oleh para peneliti terhadap pengerjaan AI GPT melalui ZeroGPT dengan skor 5.24% AI GPT. Untuk model milik ZeroGPT membaca teks pengerjaan AI GPT, penelitian ini harus diterjemahkan ke bahasa Inggris agar pengecekan dapat dilakukan secara optimal, karena ZeroGPT hanya dapat bekerja dengan teks berbahasa Inggris dan tidak dapat bekerja dengan bahasa Indonesia.