

PREDIKSI HARGA REAL ESTATE MENGGUNAKAN METODE REGRESI LINEAR BERBASIS MACHINE LEARNING

Nanda Khalijah Zainal

STMIK Royal, Kisaran

Email : Nanda Khalijah Zainal1@gmail.com

ABSTRAK

Abstrak: Kenaikan dan penurunan harga real estate yang cepat dan tidak stabil dapat berdampak pada pertumbuhan pendapatan masyarakat, akses perumahan, dan menciptakan ketidaksetaraan pembangunan antar wilayah. Diperlukan upaya prediksi untuk meminimalisir fluktuasi harga real estate yang tidak stabil. Kebijakan regulasi, seperti pembatasan spekulasi dan pemantauan ketat terhadap peminjaman berisiko, diperlukan untuk menjaga stabilitas pasar. Prediksi menggunakan metode regresi dimana dalam metode regresi terdapat 7 metode lagi meliputi (1) Linear Regression, (2) Support Vector Regression – Linear, (3) Support Vector Regression – RBF, (4) Decision Tree Regression, (5) Random Forest Regressor, (6) Gradient Boosting Regression, (7) NLP Regressor yang diterapkan dalam penelitian ini. Berdasarkan 7 metode regresi tersebut akan dicari 1 metode terbaik dengan nilai akurasi paling baik yang akan digunakan dalam proses pengolahan deploy untuk memprediksi harga rumah per unitnya. Hasilnya metode MLP-Regressor keluar sebagai metode terbaik karena metode tersebut paling banyak menghasilkan nilai akurasi terbaik pada saat pengujian akurasi dengan menggunakan 4 pengujian akurasi dengan rasio 90:10, 80:20, 70:30 dan 60:40. MLP-Regressor pada pengujian akurasi rasio 80:20 menghasilkan 2 nilai akurasi terbaik yaitu nilai estimasi pada MSE sebesar 1.00796350 dan pada RMSE sebesar 1.00397386. Kemudian pada rasio 70:30 MLP-Regressor menghasilkan 1.43239002 dengan nilai pengukuran MSE, 1.19682498 dengan RMSE. Kemudian pada rasio 60:40 MLP-Regressor menghasilkan 1.74633185 dengan nilai pengukuran MSE, 1.32148850 dengan RMSE. Dengan demikian metode regresi yaitu MLP-Regressor digunakan sebagai metode regresi yang akan di implementasikan dalam pembuatan deploy aplikasi untuk memprediksi harga rumah per unitnya.

Kata Kunci: Regresi Linier, Machine Learning, Harga

ABSTRACT

Abstract: Rapid and unstable increases and decreases in real estate prices can impact people's income growth, access to housing, and create development inequality between regions. Predictive efforts are needed to minimize unstable real estate price fluctuations. Regulatory policies, such as restrictions on speculation and strict monitoring of risky lending, are necessary to maintain market stability. Predictions use the regression method where in the regression method there are 7 more methods including (1) Linear Regression, (2) Support Vector Regression – Linear, (3) Support Vector Regression – RBF, (4) Decision Tree Regression, (5) Random Forest Regressor, (6) Gradient Boosting Regression, (7) NLP Regressor applied in this research. Based on the 7 regression methods, we will look for the best method with the best accuracy value which will be used in the deploy processing process to predict house prices per unit. As a result, the MLP-Regressor method came out as the best method because this method produced the best accuracy values when testing accuracy using 4 accuracy tests with ratios of 90:10, 80:20, 70:30 and 60:40. The MLP-Regressor in the 80:20 ratio accuracy test produces the 2 best accuracy values, namely the estimated value for MSE of 1.00796350 and RMSE of 1.00397386. Then at a ratio of 70:30 MLP-Regressor produces 1.43239002 with MSE measurement values, 1.19682498 with RMSE. Then at a ratio of 60:40 MLP-Regressor produces 1.74633185 with MSE measurement values, 1.32148850 with RMSE. Thus, the regression method, namely MLP-Regressor, is used as a regression method that will be implemented in deploying applications to predict house prices per unit.

Keywords: Linear Regression, Machine Learning, Pricing

PENDAHULUAN



Rumah bukan hanya suatu struktur fisik, mereka adalah tempat di mana kenangan dibuat, keluarga dibesarkan, dan kehidupan dijalani.[1] Kondisi pasar real estate rumah sangat tergantung pada sejumlah faktor, termasuk permintaan dan penawaran di pasar, kebijakan pemerintah terkait perumahan, tingkat suku bunga, dan tren gaya hidup masyarakat. Real estate, atau properti, membentuk landasan yang kokoh dalam struktur ekonomi global dan merupakan elemen vital dalam kehidupan sehari-hari masyarakat.[2] Keberadaan dan nilai properti mencerminkan lebih dari sekadar struktur fisik, mereka juga menciptakan lingkungan sosial, ekonomi, dan budaya yang memengaruhi kehidupan manusia secara langsung. Seiring dengan perkembangan zaman, real estate telah berkembang menjadi lebih dari sekadar tempat tinggal atau tempat bisnis. Investasi dalam properti menjadi salah satu bentuk investasi yang paling signifikan dan berdampak luas, mencakup segmen perumahan, komersial, dan industri. Keberlanjutan dan pertumbuhan pasar real estate terkait erat dengan dinamika ekonomi, perkembangan sosial, dan perubahan kebijakan pemerintah.

Harga real estate rumah memainkan peran kunci dalam ekosistem properti, mengingat bahwa rumah bukan sekadar tempat tinggal, tetapi juga mencerminkan kestabilan finansial dan kesejahteraan bagi banyak individu dan keluarga. Fenomena harga real estate rumah mencakup interaksi yang kompleks antara faktor-faktor ekonomi, sosial, dan lingkungan, menciptakan suatu dinamika pasar yang selalu berubah. Pasar real estate rumah sering menjadi cermin dari kesehatan ekonomi suatu negara atau wilayah. Naik turunnya harga rumah dapat menjadi indikator kestabilan atau ketidakstabilan ekonomi, mempengaruhi keputusan pembelian, investasi, dan bahkan keputusan kebijakan pemerintah.[3]

Berdasarkan data dari harga real estate rumah memiliki dampak yang signifikan pada dinamika ekonomi, keuangan pribadi, dan kesejahteraan masyarakat secara keseluruhan. Melalui analisis data harga real estate rumah, kita dapat memperoleh wawasan yang mendalam tentang tren pasar, faktor-faktor pendorong, dan tantangan yang dihadapi oleh pelaku dalam industri properti. Data-data ini tidak hanya mencerminkan nilai fisik dari suatu properti, tetapi juga mencatat cerita ekonomi suatu wilayah atau negara. Dalam beberapa tahun terakhir, harga real estate rumah telah menjadi fokus perhatian, karena fluktuasi mereka dapat memberikan gambaran tentang kesehatan dan stabilitas ekonomi secara keseluruhan. Analisis data harga real estate rumah melibatkan pemahaman terhadap tren jangka panjang dan pendek, identifikasi area dengan pertumbuhan harga yang signifikan, dan penelaahan faktor-faktor makroekonomi yang berdampak pada harga properti. Selain itu, data ini dapat memberikan gambaran tentang respons pasar terhadap perubahan kebijakan pemerintah, perubahan demografis, dan dinamika ekonomi global.

Permasalahannya adalah jika terjadi kenaikan dan penurunan harga real estate rumah yang cepat dan tidak stabil itu berpengaruh pada pertumbuhan pendapatan masyarakat. Hal ini dapat menciptakan kesulitan akses perumahan bagi sebagian besar penduduk, terutama generasi muda. Data harga real estate yang menunjukkan pertumbuhan yang tidak seimbang antar wilayah dapat menciptakan ketidaksetaraan pembangunan. Wilayah tertentu mungkin mengalami pertumbuhan harga yang lebih tinggi daripada yang lain, menciptakan ketidaksetaraan dalam akses dan pemanfaatan sumber daya perumahan. Pada tahun 2000 hingga 2010 beberapa wilayah, terutama di beberapa negara maju, mengalami kenaikan harga real estate rumah yang cepat selama periode ini. Hal ini dapat menciptakan kekhawatiran akan terjadinya gelembung properti dan risiko koreksi harga yang tajam di masa depan.[4] Dan pada peristiwa global seperti pandemi COVID-19 pada tahun 2020 menciptakan ketidakpastian ekonomi yang dapat memengaruhi harga real estate rumah. Pandemi ini dapat memicu perubahan dalam kebijakan pemerintah dan perilaku konsumen, yang semuanya dapat memainkan peran dalam harga properti.[5]

Berdasarkan permasalahan diatas, bahwa tindakan pencegahan diperlukan sebagai upaya untuk meminimalisir terjadi kenaikan dan penurunan harga real estate rumah yang cepat dan tidak stabil. Implementasikan kebijakan regulasi yang efektif untuk mencegah kenaikan harga real estate rumah yang tidak stabil. Hal ini dapat mencakup pembatasan spekulasi properti, pemantauan ketat terhadap praktik peminjaman yang berisiko, dan penegakan hukum terhadap kecurangan atau manipulasi pasar. Pengujian harga real estate nantinya akan menggunakan dataset prediksi harga real estate. Prediksi harga real estate rumah tersebut akan menggunakan metode regresi. Penggunaan metode regresi adalah untuk mengetahui hubungan pengaruh dari variabel (x) yaitu variabel independen dengan variabel y yaitu variabel dependen. Variabel independen adalah variabel bebas yang mempengaruhi variabel dependen. Variabel dependen merupakan variabel yang terikat yang dipengaruhi oleh variabel dependennya. Dimana terdapat 3 variabel independen yang akan diolah yaitu variabel X1 Usia Rumah, X3 Jumlah Toko, Y Jarak MRT dan variabel dependennya adalah X2 Harga Unit.

Metode regresi pada harga real estate berbasis machine learning digunakan untuk memodelkan hubungan antara variabel input yang relevan dengan harga real estate dan variabel target yang dapat mencakup tingkat harga real estate. Dalam konteks ini, regresi linear seringkali merupakan pilihan karena sederhana dan dapat memberikan pemahaman yang baik tentang hubungan antara variabel. Selain itu Metode Regresi Linear

merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X). Variabel dependen merupakan variabel akibat atau variabel yang dipengaruhi, sedangkan variabel independen merupakan variabel sebab atau variabel yang mempengaruhi. Prediksi terhadap nilai variabel dependen dapat dilakukan jika variabel independennya diketahui.

Oleh karena itu, penelitian ini bertujuan mendapatkan hasil estimasi dan mengetahui metode manakah yang menghasilkan nilai akurasi yang lebih akurat dengan membandingkan 7 model yang terdapat dalam regresi diantaranya adalah (1) Linear Regression, (2) Support Vector Regression Linear, (3) Support Vector Regression RBF, (4) Decision Tree Regression, (5) Random Forest Regressor, (6) Gradient Boosting Regression, (7) NLP Regressor. Serta untuk hasil pengujiannya menggunakan 4 model pengukuran nilai estimasi yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), dan R2-Score.

METODE

Metode riset memuat hal-hal penting terkait pelaksanaan penelitian dan tahapan yang akan ditempuh dalam penelitian ini untuk mencapai suatu hasil atau output yang diinginkan. Berikut ini hal-hal yang mencakup ke dalam metode riset antara lain :

Peralatan yang akan digunakan dalam penelitian ini adalah berupa (1) Komputer dengan spesifikasi yang mendukung, (2) software yang akan digunakan dalam proses pengolahan berupa software Jupyter notebook dalam Anaconda Navigator dimana bahasa pemrograman yang akan digunakan adalah bahasa pemrograman Python, Visual Studio Code sebagai teks editor dan menggunakan HTML sebagai bahasa markup yang digunakan untuk membangun deploy aplikasinya.

Dalam machine learning, penelitian ini dikategorikan sebagai penelitian supervised learning yaitu sebuah teknik model dimana data yang akan diproses memiliki label/target/class dengan tujuan untuk mengetahui hubungan kausalitas antara variabel bebas (variabel independen) dan variabel yang menjadi target/labelnya (variabel y). Oleh sebab itu, variabel yang terdapat dalam penelitian ini terdiri dari variabel independen atau variabel x meliputi X1 Usia Rumah, X3 Jumlah Toko dan Y Jarak MRT. Serta yang menjadi variabel dependennya adalah X2 Harga Unit.

Tahapan yang akan ditempuh dalam penelitian ini digambarkan dengan flowchart sehingga dapat terlihat dengan jelas tahapan yang akan dilakukan.

Data riset yang akan digunakan adalah dataset harga real estate yang berasal dari kaggle.com pada tahun 2021. Data tersebut berjumlah sebanyak 415 baris data. Terdapat 3 kolom (meliputi X1 Usia Rumah, X3 Jumlah Toko dan Y Jarak MRT) yang akan digunakan untuk pengujian dan 1 kolom (X2 Harga Unit) sebagai kolom targer/label

Pembagian data dalam penelitian ini menggunakan data training dan data testing dengan rasio sebesar 90:10 Dimana data training memiliki kisaran data sebanyak 90 dari dataset serta data testing sebanyak 10 dari dataset. Algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasi dapat diperoleh sebesar 99,40%. Adapun rasio 80:20, 70:30, dan 60:40 yang digunakan dalam penelitian ini sehingga rasio pembagian data lebih bervariasi dimana tujuan yang diharapkan adalah agar memperoleh model regresi manakah yang memiliki hasil estimasi terbaik.

HASIL DAN PEMBAHASAN

Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar data dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel x terhadap variabel y. Selanjutnya, data akan diuji dengan 4 bentuk pembagian data dengan rasio yang berbeda. Dimana didalamnya akan dilakukan pengujian data menggunakan 7 model regresi dan diukur dengan 4 model nilai evaluasi. Hal tersebut bertujuan untuk menemukan model regresi manakah yang paling baik dengan menghasilkan nilai akurasi yang paling akurat untuk digunakan dalam estimasi harga rumah per unitnya. Tahapan pembahasan yang dilakukan adalah sebagai berikut :

Visualisasi Data

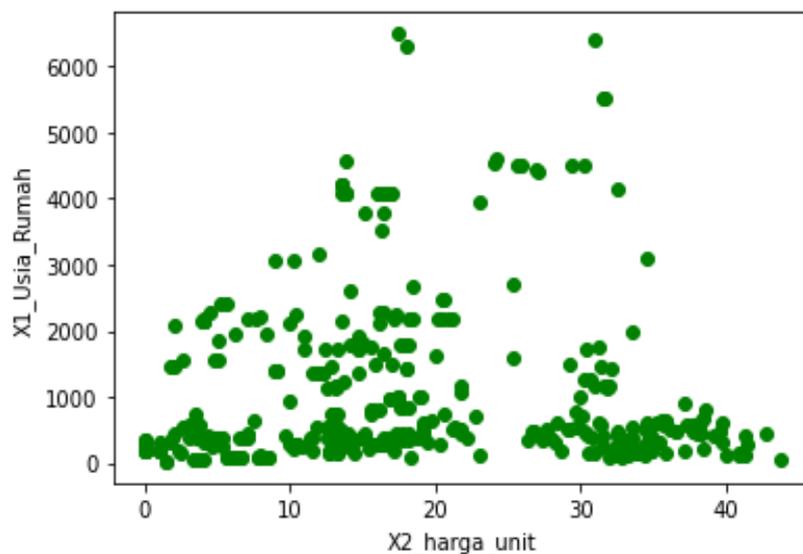
Visualisasi data bertujuan untuk memudahkan seseorang dalam mengamati dan memahami sebuah data. Terdapat berbagai model yang dapat digunakan untuk menyajikan sebuah data kedalam bentuk visualnya. *Scatter*



plot menggunakan *library* yaitu `matplotlib.pyplot`. Pada penelitian ini visualisasi data menggunakan *scatter plot* yang bertujuan untuk melihat plot sebaran data dan mengamati seberapa besar hubungan pengaruh antara sumbu x dengan sumbu y. Data *fuel consumptions rating* yang digunakan untuk mengestimasi karbon dioksida kendaraan telah divisualisasikan terhadap sumbu x nya. Visualisasi data disajikan dalam bentuk antar variabel x dengan variabel y dan penyajian *scatter* tiap variabel menggunakan warna yang berbeda agar lebih mudah dipahami dan mudah diingat. Visualisasinya data tersebut sebagai berikut ini :

Scatter Plot X1_Usia_Rumah Terhadap X2_harga_unit

Berikut ini tampilan visualisasi data pada X1_Usia_Rumah sebagai sumbu y terhadap X2_harga_unit sebagai sumbu x.

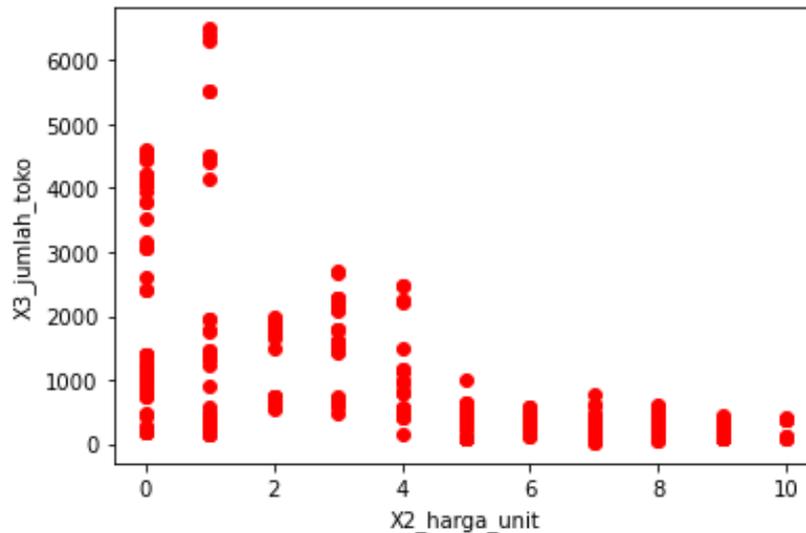


Gambar 1. Scatter X1_Usia_Rumah terhadap X2_harga_unit

Pada gambar diatas, plot titik sebaran X1_Usia_Rumah (sumbu y) divisualisasikan dengan titik berwarna hijau. Berdasarkan gambar diatas, dapat diketahui bahwa ukuran usia sebuah rumah berjalan dengan setabil meski masih terjadi naik turun pada usia rumahnya, maka harga rumah per unitnya yang dihasilkan relatif setabil. Hubungan yang ditunjukkan oleh gambar diatas ialah hubungan positif karena adanya peningkatan dan penurunan yang ditunjukkan oleh sumbu x (X2_harga_unit) yang diikuti dengan peningkatan sumbu y (X1_Usia_Rumah).

Scatter Plot X3_jumlah_toko Terhadap X2_harga_unit

Selanjutnya tampilan visualisasi data pada X3_jumlah_toko (sumbu y) terhadap X2_harga_unit (sumbu x).

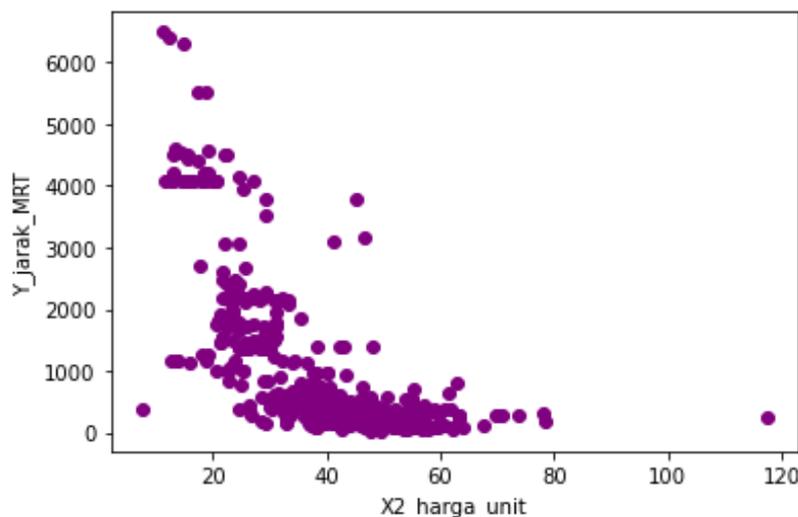


Gambar 2. Scatter $X3_jumlah_toko$ terhadap $X2_harga_unit$

Visualisasi plot titik sebaran dari sumbu y ($X3_jumlah_toko$) terhadap sumbu x ($X2_harga_unit$) disajikan kedalam *scatter* berwarna merah. Hubungan yang ditunjukkan mulai bersifat negative karena menunjukkan penurunan dari variabel $X3_jumlah_toko$ ke variabel $X2_harga_unit$.

Scatter Plot Y_jarak_MRT Terhadap $X2_harga_unit$

Tampilan visualisasi data pada Y_jarak_MRT (sumbu y) terhadap $X2_harga_unit$ (sumbu x).



Gambar 3. Scatter Y_jarak_MRT terhadap $X2_harga_unit$

Visualisasi pada sumbu y (Y_jarak_MRT) terhadap sumbu x ($X2_harga_unit$) menggunakan *scatter* plot berwarna purple. Berdasarkan gambar diatas, plot titik sebaran tampak membentuk sebuah garis lurus ke bawah, hal tersebut memberikan makna bahwa sumbu x mengalami penurunan yang derastis terhadap sumbu y.

Pengujian Data dengan 7 Model Regresi

Pada penelitian ini menggunakan metode regresi dimana akan dilakukan perbandingan dari 7 metode regresi agar diperoleh model dengan hasil akurasi terbaik untuk digunakan dalam proses harga per unit dari sebuah rumah. Penelitian ini menggunakan pembagian data untuk diuji sebanyak 4 bentuk diantaranya pembagian rasio pertama yaitu 90:10, kedua 80:20, ketiga 70:30 dan rasio keempat yaitu 60:40. Pengujian data dengan rasio yang bervariasi untuk mencari serta menentukan model regresi manakah yang paling sering muncul dalam

menghasilkan nilai estimasi yang terbaik. Sehingga, nantinya akan diketahui model regresi yang terbaik dalam melakukan estimasi terhadap dataset untuk menentukan harga rumah perunitnya. Berikut pembahasan penelitian untuk menemukan model regresi terbaik yang diuraikan kedalam rincian poin sebagai berikut :

Pengujian Akurasi Pada Rasio Data 90:10

Pengujian data pertama menggunakan rasio 90:10, dimana *data training* sebesar 90% dari total data dan *data testing* sebesar 10% menggunakan 7 model regresi diperoleh hasil seperti pada tabel dibawah ini :

Tabel 1. Pengujian Akurasi Data Rasio 90:10

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	5.39847558	5.83610867	7.63944283	0.30203022
SVR-Linear	4.45107440	5.58767347	7.47507423	0.33174184
SVR-RBF	5.00278335	8.80301906	9.38244055	-0.05279762
DecisionTree	5.60954791	6.64303920	2.57740940	1
RandomForest	1.08916457	3.83512942	1.95834865	0.95413375
GradientBoosting	2.15576979	1.21996225	3.49279579	0.85409854
MLP-Regressor	6.50198397	8.80036198	9.38102445	-0.05247984

Berdasarkan tabel diatas, MAE sebesar 1.08916457 menandakan bahwa rata-rata kesalahan prediksi pada harga rumah per unit adalah sekitar 1.09. Semakin rendah nilai MAE, semakin baik kinerja model. MSE sebesar 1.21996225 mengukur rata-rata dari kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin baik kinerja model. MRSE sebesar 1.95834865 adalah akar kuadrat dari MSE dan memberikan gambaran intuitif tentang kesalahan prediksi. Semakin rendah nilai MRSE, semakin baik kinerja model. R2-Score sebesar 1 mengindikasikan bahwa model DecisionTree dapat menjelaskan sekitar 96.74% dari variasi dalam harga rumah per unit. Semakin mendekati 1, semakin baik kinerja model. Berdasarkan nilai-nilai tersebut, dapat disimpulkan bahwa RandomForest memiliki performa terbaik untuk MAE dan MRSE, sementara GradientBoosting memiliki performa terbaik untuk MSE. Selain itu, model DecisionTree memiliki nilai estimasi terbaik untuk R2-Score.

Pengujian Akurasi Pada Rasio Data 80:20

Pengujian data kedua menggunakan pembagian data dengan rasio 80:20. Terdapat sebanyak 80% *data training* dan 20% *data testing* dari dataset, diperoleh hasil seperti tabel dibawah ini :

Tabel 2. Pengujian Akurasi Data Rasio 80:20

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	6.25177877	7.26032317	8.52075300	0.22088182
SVR-Linear	5.21478346	6.61987445	8.13626109	0.28960951
SVR-RBF	5.92705357	1.07489319	1.03677056	-0.15348699
DecisionTree	6.51330841	9.00353246	3.00058869	1
RandomForest	1.25336907	6.80405460	2.60845828	0.92698448
GradientBoosting	1.68821381	7.73620809	2.78140398	0.91698138
MLP-Regressor	7.08107036	1.00796350	1.00397386	-0.08166355

Berdasarkan tabel diatas, MAE sebesar 1.25336907 menandakan bahwa rata-rata kesalahan prediksi pada harga rumah per unit adalah sekitar 1.25. Semakin rendah nilai MAE, semakin baik kinerja model. MSE sebesar 1.00796350 mengukur rata-rata dari kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin baik kinerja model. MRSE sebesar 1.00397386 adalah akar kuadrat dari MSE dan memberikan gambaran intuitif tentang kesalahan prediksi. Semakin rendah nilai MRSE, semakin baik kinerja model. R2-Score sebesar 1 mengindikasikan bahwa model DecisionTree dapat menjelaskan sekitar 93.14% dari variasi dalam harga rumah per unit. Semakin mendekati 1, semakin baik kinerja model. Berdasarkan nilai-nilai tersebut, dapat disimpulkan bahwa RandomForest memiliki performa terbaik untuk MAE, MLP-Regressor memiliki performa terbaik untuk MSE dan MRSE, sementara DecisionTree memiliki nilai estimasi terbaik untuk R2-Score.

Pengujian Akurasi Pada Rasio Data 70:30

Pengujian data ketiga dengan menggunakan rasio data 70:30 yaitu 70% *data training* dan 30% *data testing* menggunakan 7 model regresi diperoleh hasil sebagai berikut :

Tabel 3. Pengujian Akurasi Data Rasio 70:30

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	6.23345853	6.78710192	8.23838693	0.50173191
SVR-Linear	5.58356384	7.25191383	8.51581695	0.46760823
SVR-RBF	7.07082184	1.55432249	1.24672471	-0.14108983
DecisionTree	7.89491929	1.57070971	3.96321803	1
RandomForest	9.82379953	5.24364585	2.28990084	0.96150431
GradientBoosting	1.37497588	4.20917758	2.05162803	0.96909876
MLP-Regressor	8.66026232	1.43239002	1.19682498	-0.05157436

Berdasarkan tabel diatas, MAE sebesar 1.37497588 menandakan bahwa rata-rata kesalahan prediksi pada harga rumah per unit adalah sekitar 1.37. Semakin rendah nilai MAE, semakin baik kinerja model. MSE sebesar 1.43239002 mengukur rata-rata dari kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin baik kinerja model. MRSE sebesar 1.19682498 adalah akar kuadrat dari MSE dan memberikan gambaran intuitif tentang kesalahan prediksi. Semakin rendah nilai MRSE, semakin baik kinerja model. R2-Score sebesar 1 mengindikasikan bahwa model DecisionTree dapat menjelaskan sekitar 97.15% dari variasi dalam harga rumah per unit. Semakin mendekati 1, semakin baik kinerja model. Berdasarkan nilai-nilai tersebut, dapat disimpulkan bahwa GradientBoosting memiliki performa terbaik untuk MAE, MLP-Regressor memiliki performa terbaik untuk MSE dan MRSE, sementara DecisionTree memiliki nilai estimasi terbaik untuk R2-Score.

Pengujian Akurasi Pada Rasio Data 60:40

Pengujian data keempat dengan menggunakan rasio data 60:40 yakni diman 60% digunakan sebagai *data training* dan 40% digunakan sebagai *data testing*. Diperoleh hasil sebagai berikut ini :

Tabel 4. Pengujian Akurasi Data Rasio 60:40

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	6.31560507	7.34699967	8.57146409	0.55064746
SVR-Linear	5.99063539	9.31262445	9.65019402	0.43042716
SVR-RBF	8.61043810	1.99149148	1.41120214	-0.21802341

DecisionTree	5.51877853	7.84265599	2.80047424	1
RandomForest	1.07723570	6.73737454	2.59564530	0.9587933
GradientBoosting	1.51040234	5.00860839	2.23799205	0.96936667
MLP-Regressor	9.50439842	1.74633185	1.32148850	-0.06808043

Berdasarkan tabel diatas, MAE sebesar 1.07723570 menandakan bahwa rata-rata kesalahan prediksi pada harga rumah per unit adalah sekitar 1.08. Semakin rendah nilai MAE, semakin baik kinerja model. MSE sebesar 1.74633185 mengukur rata-rata dari kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin baik kinerja model. MRSE sebesar 1.32148850 adalah akar kuadrat dari MSE dan memberikan gambaran intuitif tentang kesalahan prediksi. Semakin rendah nilai MRSE, semakin baik kinerja model. R2-Score sebesar 1 mengindikasikan bahwa model DecisionTree dapat menjelaskan sekitar 96.95% dari variasi dalam harga rumah per unit. Semakin mendekati 1, semakin baik kinerja model. Berdasarkan nilai-nilai tersebut, dapat disimpulkan bahwa RandomForest memiliki performa terbaik untuk MAE, MLP-Regressor memiliki performa terbaik untuk MSE dan MRSE, sementara DecisionTree memiliki nilai estimasi terbaik untuk R2-Score.

Perbandingan 3 Metode Regresi Terbaik

Setelah melakukan pengujian akurasi dengan 4 bentuk rasio data, kemudian diperoleh 2 metode terbaik yang paling banyak menghasilkan nilai estimasi. 2 metode tersebut adalah *RandomFores*, dan *MLP-Regressor*. Masing-masing metode memiliki tingkat akurasi yang tinggi diantaranya dalam rasio 90:10 *RandomFores* sebagai metode terbaik dalam rasio tersebut memiliki tingkat akurasi sebesar 98%, pada rasio 80:20, 70:30, dan 60:40 *MLP-Regressor* sebagai metode terbaik dalam rasio tersebut memiliki nilai akurasi sebesar 95-97%. Berikut adalah hasil dari perbandingan 2 metode terbaik dengan tingkat akurasi yang dihasilkan :

Tabel 5. Perbandingan 2 Metode Terbaik

Metode Regresi	90:10	80:20	70:30	60:40
<i>RandomForest</i>	98%	-	-	-
<i>MLP-Regressor</i>	-	97%	96%	95%

Pada tabel di atas, dapat diketahui bahwa *MLP-Regressor* menghasilkan nilai akurasi terbaik sebanyak 3 kali yakni pada pengujian akurasi dengan rasio 80:20, 70:30 dan 60:40. Sementara metode lainnya yaitu *RandomForest* menghasilkan nilai akurasi terbaik sebanyak 1 kali yaitu pada rasio 90:10. Dengan demikian bahwa berdasarkan hasil pengujian dan perbandingan metode terbaik yang tersisa dihasilkanlah metode *MLP-Regressor* sebagai metode yang terbaik diantara 7 metode regresi lainnya untuk memprediksi harga rumah per unit dengan nilai akurasi yang dihasilkan yakni sebesar 97%, 96% dan 95%.

KESIMPULAN

Berdasarkan hasil pengolahan data menggunakan 7 metode regresi diperoleh 2 metode terbaik yang paling banyak menghasilkan nilai estimasi. 2 metode tersebut adalah *RandomFores*, dan *MLP-Regressor*. Masing-masing metode memiliki tingkat akurasi yang tinggi diantaranya dalam rasio 90:10 *RandomFores* sebagai metode terbaik dalam rasio tersebut memiliki tingkat akurasi sebesar 98%, pada rasio 80:20, 70:30, dan 60:40 *MLP-Regressor* sebagai metode terbaik dalam rasio tersebut memiliki nilai akurasi sebesar 95-97%. Hasilnya metode *MLP-Regressor* keluar sebagai metode terbaik karena metode tersebut paling banyak menghasilkan nilai akurasi terbaik pada saat pengujian akurasi dengan menggunakan 4 rasio. *MLP-Regressor* pada pengujian akurasi rasio 80:20 menghasilkan 2 nilai akurasi terbaik yaitu nilai estimasi pada MSE sebesar 1.00796350 dan pada RMSE sebesar 1.00397386. Kemudian pada rasio 70:30 *MLP-Regressor* menghasilkan 1.43239002 dengan nilai pengukuran MSE, 1.19682498 dengan RMSE. Kemudian pada rasio 60:40 *MLP-Regressor* menghasilkan 1.74633185 dengan nilai pengukuran MSE, 1.32148850 dengan RMSE. Dengan demikian metode regresi yaitu *MLP-Regressor* digunakan sebagai metode regresi yang akan di implementasikan dalam pembuatan deploy aplikasi untuk memprediksi harga rumah per unitnya.

DAFTAR PUSTAKA

- [1] E. F. Rahayuningtyas, F. N. Rahayu, and Y. Azhar, "Prediksi Harga Rumah Menggunakan General Regression Neural Network," *J. Inform.*, vol. 8, no. 1, pp. 59–66, 2021, doi: 10.31294/ji.v8i1.9036.
- [2] S. Nur and C. Utomo, "Analisa Nilai Agunan Rumah Tinggal di Medokan Asri Utara XII Surabaya," vol. 2, no. 2, pp. 67–71, 2013.
- [3] D. A. N. Property and D. I. Bei, "real assets)".
- [4] A. A. Thobarry, "Analisis Pengaruh nilai tukar, suku bunga, laju inflasi dan pertumbuhan GDP terhadap indeks harga saham sektor properti (kajian empiris pada Bursa Efek Indonesia periode pengamatan tahun 2000-2008)," *Progr. Pascasarj. Universi-tas Diponegoro. Semarang*, p. 104, 2009, [Online]. Available: <http://eprints.undip.ac.id/19029/>
- [5] L. B. Masalah and M. Mulyadi, "BAB I," pp. 1–14, 2013.