

IMPLEMENTASI ALGORITMA REGRESI LINIER DAN K-NEAREST NEIGHBOR UNTUK PREDIKSI HARGA RUMAH

Ilham Kurniawan¹, Nining Rahaningsih², Tati Suprapti³

Teknik Informatika, STMIK IKMI Cirebon
Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B, Majasem, Cirebon Indonesia
Kurniawanchand07@gmail.com

ABSTRAK

Perkembangan pesat dibidang informatika telah membawa dampak signifikan pada berbagai aspek kehidupan manusia. Ini disebabkan oleh kemajuan teknologi komputer, perangkat mobile, jaringan internet, dan berbagai aplikasi perangkat lunak yang terus berkembang. Prediksi harga rumah menjadi suatu hal yang krusial karena dapat membantu calon pembeli, penjual, dan pihak terkait lainnya untuk membuat keputusan yang informasional dan cerdas. Tujuan penelitian ini yaitu untuk mengetahui faktor yang paling berpengaruh terhadap harga rumah, mengetahui perbandingan tingkat akurasi dari kedua algoritma dalam prediksi harga rumah, dan cara generalisasi model regresi linier ke pasar perumahan yang berbeda. Metode yang digunakan dalam prediksi harga rumah adalah metode KDD menggunakan algoritma regresi linear dan *k-nearest neighbor* dalam konteks data mining. Pada penelitian ini faktor yang paling berpengaruh terhadap harga rumah adalah luas bangunan dengan nilai koefisien sebesar 23047431.201. Algoritma KNN mempunyai hasil akurasi yang lebih baik dibandingkan dengan regresi linier, perbandingan tingkat akurasi ini diambil berdasarkan nilai relative error. Hasil nilai relative error regresi linier sebesar 71.95% +/- 82.63%. Sedangkan hasil relative error k-nearest neighbor pada gambar sebesar 30.95% +/- 29.87%. Model regresi linier dapat digeneralisasi dengan baik ke pasar perumahan yang berbeda melibatkan beberapa pendekatan. Diantaranya yaitu Pengumpulan data yang representatif, normalisasi dan standarisasi, pemahaman konteks pasar.

Kata kunci : Data Mining, Prediksi, Regresi Linier, K-Nearest Neighbor, Rumah

1. PENDAHULUAN

Perkembangan pesat di bidang informatika telah membawa dampak signifikan pada berbagai aspek kehidupan manusia. Ini disebabkan oleh kemajuan teknologi komputer, perangkat mobile, jaringan internet, dan berbagai aplikasi perangkat lunak yang terus berkembang. Perkembangan pesat di bidang informatika telah menjadi penggerak utama perubahan di berbagai sektor kehidupan, seperti kegiatan bisnis properti. Prediksi harga rumah menjadi suatu hal yang krusial karena dapat membantu calon pembeli, penjual, dan pihak terkait lainnya untuk membuat keputusan yang informasional dan cerdas. Data mining adalah suatu proses ekstraksi pola atau pengetahuan yang signifikan dari dataset besar, dan regresi linear merupakan salah satu teknik statistik yang berguna untuk memodelkan hubungan linier antara variabel dependen dan satu atau lebih variabel independen, sedangkan k-nearest neighbor adalah algoritma yang digunakan untuk klasifikasi dan regresi.

Pada penelitian sebelumnya [1] dengan menggunakan metode regresi linier untuk meramalkan penjualan properti di PT. XYZ. Evaluasi terhadap tingkat akurasi prediksi dilakukan melalui penggunaan metrik MSE, RMSE, dan MAPE. Hasil penelitian berdasarkan data penjualan properti selama lima tahun menunjukkan bahwa proyeksi penjualan pada bulan berikutnya untuk properti tipe rumah 59 diperkirakan sekitar 4 unit, tipe rumah sudut sekitar 2 unit, sementara itu, 5 tipe lainnya diperkirakan tidak terjual. Penelitian terdahulu selanjutnya [2] peneliti

membahas mengenai penerapan metode regresi linear berganda menggunakan algoritma, yang menghasilkan sebuah model regresi dengan nilai konstanta sebesar 284645.5891073216, serta nilai koefisien - 139837.38007863 dan 363493.06049751. Setelah proses pemisahan data, nilai RMSE untuk data pelatihan adalah 8447373.485 dan untuk data uji adalah 9769609.026. Sementara itu, nilai koefisien determinasi mencapai 0.579 untuk data pelatihan dan 0.662 untuk data uji. Penelitian terdahulu berikutnya [3] membahas mengenai peneliti mengadopsi pendekatan regresi linier untuk meramalkan jumlah permohonan paten yang terdaftar di Indonesia, dengan mengandalkan data pendaftaran dari tahun 2014 hingga tahun 2018. Hasil ini dapat berfungsi sebagai landasan untuk menentukan jumlah paten yang perlu diperiksa dan diterbitkan.

Tujuan penelitian ini prediksi yang dilakukan yaitu untuk mengetahui faktor apa saja yang paling memengaruhi harga rumah, perbandingan tingkat akurasi prediksi regresi linier dengan k-nearest neighbor, dan memastikan bahwa model regresi linier dapat digeneralisasi dengan baik ke pasar perumahan yang berbeda.

Penelitian ini menggunakan metode KDD, algoritma yang digunakan untuk memprediksi harga rumah adalah algoritma regresi linier dan k-nearest neighbor. Regresi linier adalah suatu metode statistika untuk memodelkan hubungan linier antara satu variabel terikat (variabel dependen) dengan satu bahkan lebih variabel independen (variabel prediktor).

Model ini digunakan untuk memahami hubungan antara variabel-variabel tersebut. Regresi linear sederhana melibatkan satu variabel independen. K-Nearest Neighbors (KNN) adalah salah satu classification dan regression algorithm yang sering digunakan. Algoritma ini supervised learning, di mana model mempelajari pola dari data pelatihan yang telah dilabeli untuk memprediksi data baru yang tidak dilabeli.

Penelitian ini memberikan beberapa implikasi yang dapat menjadi landasan penting bagi berbagai pemangku kepentingan, termasuk calon pembeli rumah, agen real estate, serta pengembang properti. Pemahaman mendalam tentang variabel-variabel yang signifikan dalam memprediksi harga rumah dapat memberikan pandangan yang lebih akurat kepada calon pembeli terkait faktor-faktor yang memengaruhi nilai properti. Hal ini dapat membantu membuat keputusan pembelian yang lebih terinformasi. Selain itu, agen real estate dapat menggunakan model ini sebagai alat penunjang dalam menilai nilai properti dan memberikan panduan harga yang lebih objektif kepada kliennya. Pemangku kepentingan dapat memanfaatkan informasi ini untuk menyesuaikan ekspektasi atau strategi dalam pasar properti. Implikasi lainnya mencakup kemungkinan pengembangan model yang lebih canggih atau integrasi variabel tambahan untuk meningkatkan performa prediktif.

- 1) Variabel apa saja yang paling berpengaruh terhadap harga rumah?
- 2) Berapa tingkat akurasi prediksi regresi linier dengan *k-nearest neighbor*?
- 3) Bagaimana memastikan bahwa model regresi linier dapat digeneralisasi dengan baik ke pasar perumahan yang berbeda?

2. TINJAUAN PUSTAKA

[4] Uji asumsi klasik pada analisis regresi linier berganda telah dilakukan pada kasus penyusunan persamaan allometrik kenari muda. Menghasilkan persamaan allometrik $Y = -941,765 + 399,903X_1 + 3,429X_2$.

[5] Metode regresi linier berganda digunakan untuk mengevaluasi laju pertumbuhan penduduk di BPS Deli Serdang. Temuan dari penelitian ini mencakup identifikasi atribut atau kriteria yang mempengaruhi laju pertumbuhan penduduk, termasuk penemuan pola keterkaitan erat antara jumlah laki-laki dan jumlah perempuan terhadap laju pertumbuhan penduduk.

[6] Prediksi persediaan obat jenis tablet menggunakan regresi linier. Dengan mengaplikasikan metode regresi linier, diperoleh tingkat akurasi sebesar 98,505%, menunjukkan bahwa metode ini dapat efektif diimplementasikan untuk memprediksi stok obat jenis tablet, seperti ibuprofen, di Klinik Teluk.

[7] Prediksi pemesanan bibit pohon dengan regresi linier berganda. Tantangan yang umum terjadi di BPDASHL melibatkan aspek estimasi, seperti proyeksi cuaca, kesulitan dalam proses penanaman,

keterbatasan tenaga kerja, kekurangan pengalaman dalam pembibitan pohon, serta variasi kondisi tanah yang beragam. Selain itu, kelemahan lain di instansi ini adalah ketidakberadaan sistem untuk meramalkan estimasi pemesanan bibit pohon setiap tahunnya.

[8] Estimasi penjualan produk Unilever dilakukan melalui penerapan metode regresi linier. Akurasi hasil prediksi penjualan berbagai produk Unilever dinilai dengan menggunakan MEA dan MAPE untuk mengevaluasi tingkat kesalahan prediksi. Pengujian tingkat kesalahan tersebut dilakukan untuk tiga periode, yakni bulan Januari hingga Maret 2022, pada produk seperti Axe Deodorant, Bango Kecap, Buavita, dll. Hasil pengujian menunjukkan bahwa prediksi penjualan untuk semua produk tersebut tergolong dalam kategori sangat akurat. Produk Sunsilk Conditioner mencapai tingkat kesalahan terendah dengan nilai MAPE sebesar 1%.

[9] Penggunaan metode regresi linier untuk prediksi penjualan smartphone. Pengujian menunjukkan bahwa MAPE memiliki nilai sebesar 0,032 dan MSE memiliki nilai sebesar 5.16. Dengan nilai-nilai tersebut, dapat disimpulkan bahwa prediksi penjualan smartphone menggunakan metode Regresi Linier di 82 Cell Mayang tergolong dalam kategori yang sangat baik.

[10] Penerapan algoritma K-Nearest Neighbor (KNN) untuk memproyeksikan kemungkinan terjadinya penyakit stroke dan mengevaluasi akurasi hasil prediksinya. Diketahui bahwa algoritma KNN mampu meramalkan kemungkinan terjadinya penyakit stroke berdasarkan beberapa faktor, seperti jenis kelamin, usia, hipertensi, dll. Akurasi yang diperoleh dari algoritma KNN untuk prediksi penyakit stroke adalah sekitar 95%, dengan nilai $k=9$.

3. METODE PENELITIAN

3.1. Metode Penelitian

Metode yang digunakan dalam penelitian ini menggunakan metode KDD (*Knowledge Discovery in Databases*) melibatkan serangkaian tahap yang dilakukan secara berurutan untuk menghasilkan pengetahuan yang berharga dari data.

3.2. Sumber Data

Sumber data yang digunakan dalam penelitian ini merupakan data publik yang bersumber dari website resmi Kaggle. Pada website ini tersedia banyak dataset publik yang bisa dijadikan untuk bahan penelitian karena data didapatkan dari Kaggle.

3.3. Populasi dan Sampel

Penelitian ini akan menggunakan dataset daftar harga rumah tahun 2022 berdasarkan harga, lokasi, luas bangunan, luas tanah, kamar mandi, kamar tidur, carport, dan tahun pembangunan. Yang terhitung mempunyai 710 record. Penelitian dilakukan untuk mengetahui prediksi harga rumah berdasarkan variabel terkait. Untuk sampel yang digunakan dalam

penelitian ini sejumlah populasi yaitu 710 sampel atau seluruh populasi.

3.4. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan teknik manual melakukan akses ke website dan mengunduh dataset daftar harga rumah berdasarkan variabel tertentu. Data yang di unduh berupa file dengan format Ms.Excel.

3.5. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah dengan menggunakan metode *Knowledge Discovery in Database* (KDD), dengan langkah-langkah analisis sebagai berikut.

3.5.1. Data Selection

Data Selection adalah pemilihan data yang paling relevan atau representatif untuk mencapai tujuan analisis atau model prediksi.

3.5.2. Data Cleaning

Pada tahap ini dataset yang digunakan difilter, pada tahapan ini data dihapus seperti data yang null, data duplikat ataupun data yang tidak relevan dengan penelitian.

3.5.3. Data Transformation

Setelah dataset dibersihkan maka lanjut ketahap berikutnya yaitu teknik Transformasi data, data yang sudah dibersihkan di proses untuk digabungkan atau di marge, merubah format tipe atribut atau mengekstrak fitur yang lebih representatif. Teknik seperti pemilihan atribut, normalisasi, pengurangan dimensi, dan transformasi lainnya dapat diterapkan di sini.

3.5.4. Data Mining

Data mining merupakan proses inti untuk mengidentifikasi pola, tren data dan pengetahuan informasi yang tersembunyi dalam data. Dalam Proses data mining terdapat beberapa teknik seperti klasifikasi, regresi, clusterig, asosiasi, dll. Untuk proses penelitian ini menggunakan teknik regresi dengan algoritma regresi linier dan k-nearest neighbor.

1) Prediction

Prediksi mencakup upaya untuk mengidentifikasi atau mengetahui peristiwa yang mungkin terjadi di masa depan. Dengan memproses dan menggambarkan data, pengguna dapat membuat estimasi atau prediksi tentang kemungkinan peristiwa yang akan terjadi berdasarkan pola yang teridentifikasi.

2) Linear Regression

Regresi linear merupakan metode yang digunakan untuk memodelkan korelasi antara satu atau lebih variabel dependen dengan satu (dalam regresi linear sederhana) atau lebih variabel independen (dalam regresi linear berganda). Jika kita mengasumsikan bahwa hubungan antara variabel-variabel tersebut dapat

diaproksimasi oleh suatu garis lurus, maka model yang mencocokkan hubungan tersebut dalam data tersebut dapat disebut sebagai model regresi linear



Gambar 1 Flowchart regresi linier

3) Classification

Klasifikasi adalah pengelompokan atau penempatan suatu objek atau data ke dalam kategori atau kelas tertentu berdasarkan atribut-atribut yang dimiliki.



Gambar 2 Flowchart knn

4) K-Nearest Neighbor

K-Nearest Neighbors (KNN) adalah salah satu algoritma yang digunakan untuk klasifikasi dan regresi. Algoritma ini termasuk dalam kategori supervised learning, di mana model mempelajari pola dari data pelatihan yang telah dilabeli untuk membuat prediksi terhadap data baru yang tidak dilabeli.

3.5.5. Evaluation

Tahap evaluasi digunakan untuk menentukan tingkat akurasi dan validasi data *mining*. Tahap ini dilakukan secara terus menerus untuk menemukan *performance* dan validasi data paling tepat, tentunya sesuai dengan tujuan yang sudah ditetapkan.

3.5.6. Knowledge

Setelah data dievaluasi maka dari hasil evaluasi tadi menghasilkan informasi ataupun pengetahuan. Hasil digunakan untuk membuat keputusan, mengembangkan strategi, atau meningkatkan kinerja.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

Data Selection adalah pemilihan data yang paling relevan atau representatif untuk mencapai tujuan analisis atau model prediksi. Gambar 3 merupakan contoh dataset yang didapatkan melalui website Kaggle.

price_in_rp	title	city	bedrooms	bathrooms	land_size_m2	building_size_m2	carports
1850000000	Chandra Rumah Uk 4,5x19m Harga Murah di Jelambar	Jakarta Barat	3	3	84	189	0
6800000000	Chandra Rumah 2 Lantai Uk 12,5x20,5m Lokasi Bagus Komplek Kavling Polri	Jakarta Barat	4	3	260	387	0
1650000000	Chandra Rumah Baru 3,5 Lantai Uk 4x13m Hadap Timur di Jelambar	Jakarta Barat	3	3	60	132	0
2400000000	Chandra Rumah Uk 5x18m Lokasi Bagus di Jelambar	Jakarta Barat	5	4	89	227	0
9550000000	Ohyana Rumah 2 Lt Uk 7x16m di Komplek Kavling Polri	Jakarta Barat	4	2	112	160	0
3800000000	Bintaro Jaya Sektor 3, Rumah Sudah Renovasi Dan Turun Harga, Am 7420	Jakarta Selatan	5	3	260	350	2
17000000000	Rumah bagus 3 lantai di petojor jakarta pusat	Jakarta Pusat	4	4	303	500	0
3800000000	Green Garden Rumah Lama Perlu Renovasi	Jakarta Barat	4	3	220	250	1
5650000000	Rumah Pinang Ranti Mansion Letak Strategis Maleser Jakarta Timur	Jakarta Timur	2	1	32	38	1
3200000000	Menyusu 111r Ukuran 15x25 Dkt Kebon Jeruk Rumah Bagus Depan Taman Bagus	Jakarta Barat	6	4	150	200	3

Gambar 3 Dataset daftar harga rumah

4.1.2. Data Cleaning

Gambar 4 Proses filtering data

Setelah melakukan proses pemilihan data, selanjutnya adalah tahap data *cleaning*. Data *cleaning* merupakan serangkaian proses untuk identifikasi, memperbaiki, dan mengelola kesalahan, inkonsistensi, dan anomali dalam dataset. Seperti pada gambar 4 dimana setiap atribut tidak teridentifikasi data yang hilang.

4.1.3. Data Transformation

Langkah selanjutnya adalah data *transformation*. Transformasi data adalah serangkaian

proses untuk mengubah atau mengolah data menjadi bentuk atau format yang lebih sesuai atau berguna untuk analisis maupun pemodelan. Pada penelitian ini atribut harga rumah dijadikan sebagai label seperti pada gambar 5.

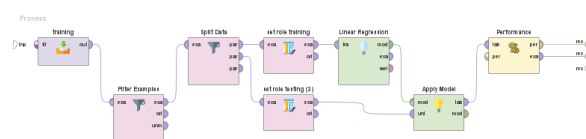
Gambar 5 Set role

Gambar 6 Split data

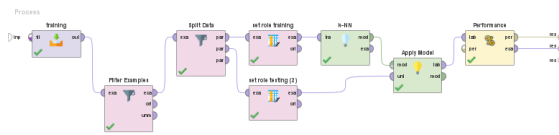
Split data merupakan proses pembagian data menjadi data training dan data testing dengan ratio yang digunakan 0.8 dan 0.2 seperti pada gambar 6. Kedua proses tersebut sangat penting dalam analisis maupun pemodelan.

4.1.4. Data Mining

Selanjutnya adalah proses data mining untuk mengidentifikasi pola atau tren tersembunyi dalam data yang tidak dapat dengan mudah diidentifikasi. Untuk mengolah data peneliti harus menentukan aplikasi yang digunakan, disini peneliti menggunakan Rapid Miner untuk mengolah data.



Gambar 7 Desain regresi linier

Gambar 8 Desain *k-nearest neighbor*

Dari kedua gambar diatas merupakan contoh desain dari algoritma regresi linier dan *k-nearest neighbor* menggunakan Rapid Miner. Desain tersebut digunakan untuk mengetahui faktor yang paling memengaruhi harga rumah dan perbandingan tingkat akurasi dari algoritma regresi linier dan *k-nearest neighbor* dengan langkah-langkah seperti pada gambar diatas.

Row No.	year_built	price_in_rp	prediction(p...	bedrooms	bathrooms	land_size_m2	building_size...	carports
1	2022	1590000000	2554175054	3	3	39	110	1
2	2021	2000000000	4560396181	5	5	125	160	2
3	2018	2400000000	7527296484	4	4	150	250	2
4	2022	2500000000	3764505352	4	3	120	220	1
5	2022	5500000000	1068435283	4	4	163	380	2
6	2022	2300000000	573789773	3	3	120	165	2
7	2022	4950000000	9823676141	6	6	223	320	3
8	1993	6000000000	1821112416	5	3	373	132	0
9	2022	4000000000	6571053092	6	6	128	270	2
10	2015	14000000000	2887741923	6	5	1431	500	4
11	2015	30000000000	3056792810	4	4	542	1100	0
12	2010	11000000000	1022406507	3	2	230	450	0
13	2021	6750000000	732425030	4	4	180	224	2
14	2021	3100000000	1645613936	3	2	100	115	1

Gambar 9 Example set linear regression

Row No.	year_built	price_in_rp	prediction(p...	bedrooms	bathrooms	land_size_m2	building_size...	carports
1	2022	1590000000	148158882	3	3	39	110	1
2	2021	2000000000	2243630198	5	5	125	160	2
3	2018	2400000000	4781636996	4	4	150	250	2
4	2022	2500000000	4919735091	4	3	120	220	1
5	2022	5500000000	5376456150	4	4	163	380	2
6	2022	2300000000	2466107443	3	3	120	165	2
7	2022	4950000000	9203853876	6	6	223	320	3
8	1993	6000000000	6503764891	5	3	373	132	0
9	2022	4000000000	464863019	6	6	128	270	2
10	2015	14000000000	225840780	6	5	1431	500	4
11	2015	30000000000	5125849288	4	4	542	1100	0
12	2010	11000000000	1962958898	3	2	230	450	0
13	2021	6750000000	4249175557	4	4	180	224	2
14	2021	3100000000	318644775	3	2	100	115	1
15	2015	8500000000	611311930	5	5	1000	2800	4

Gambar 10 Example set *k-nearest neighbor*

Pada gambar 9 dan 10 merupakan *example set* dari kedua algoritma yang digunakan dengan tujuan untuk mengidentifikasi perbandingan tingkat akurasi yang bisa dilihat dalam *performance*.

4.1.5. Evaluation

Criterion	root_mean_squared_error
root mean squared error	root_mean_squared_error: 5959169698.163 +/- 0.000
absolute error	
relative error	
squared error	

Gambar 11 Rmse regresi linier

Criterion	root_mean_squared_error
root mean squared error	root_mean_squared_error: 8778349353.540 +/- 0.000
absolute error	
relative error	
squared error	

Gambar 12 Rmse *k-nearest neighbor*

Dari kedua evaluasi diatas hasil *performance* yang diambil adalah Root Mean Squared Error. Pada gambar 11 hasil RMSE untuk regresi linier sebesar 5959169698.163, sedangkan pada gambar 12 hasil RMSE untuk KNN sebesar 8778349353.540. Kedua hasil RMSE tersebut di dapatkan melalui hasil *performance* dari desain regresi linier maupun KNN menggunakan Rapid Miner. Hasil tersebut bisa menjadi faktor dalam perbandingan tingkat akurasi dalam memprediksi harga rumah.

4.2. Pembahasan

4.2.1. Variabel Yang Paling Memengaruhi Harga Rumah

Setelah melakukan semua proses pengolahan data dengan benar, langkah selanjutnya yaitu menentukan variabel yang paling memengaruhi harga rumah. Peneliti menentukan variabel yang paling memengaruhi dengan melihat hasil dari *apply model* regresi linier seperti pada gambar 11.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	1.Stat	p-Value	Code
bedrooms	-2389953877.612	850162352.856	-0.550	0.931	-2.778	0.006	***
bathrooms	1625687180.914	85213364.291	0.428	0.937	2.142	0.033	**
land_size_m2	13146391.888	2501471.451	0.283	0.459	5.255	0.000	***
building_size_m2	23047431.201	230688.698	0.519	0.462	9.780	0	***
carports	85133259.668	47062190.146	0.073	0.799	1.788	0.075	*
(Intercept)	346938172.287	110088094.745	?	?	0.314	0.754	

Gambar 13 Hasil *apply model* regresi linier

Gambar 11 merupakan sebuah informasi hubungan setiap atribut dengan label. Nilai *coefficient* positif membuat harga rumah menjadi lebih tinggi, sedangkan nilai *coefficient* negatif membuat harga rumah menjadi lebih murah. Faktor lain untuk menentukan variabel yang paling memengaruhi bisa dengan melihat korelasi antara *P-Value* dengan code. Semakin banyak bintang dalam code semakin kecil juga *P-Value* nya dan semakin penting juga variabel dalam memengaruhi harga rumah.

Table 1 Hasil

No.	Variabel Yang Paling Memengaruhi Harga Rumah
1.	Luas Bangunan (<i>Building Size</i>)
2.	Luas Tanah (<i>Land Size</i>)
3.	Kamar Tidur (<i>Bedrooms</i>)
4.	Kamar Mandi (<i>Bathrooms</i>)
5.	<i>Carpots</i>

Variabel yang paling berpengaruh terhadap harga rumah yaitu luas bangunan (*building size*) dengan nilai koefisien sebesar 23047431201.

4.2.2. Tingkat Akurasi Prediksi Regresi Linier dengan K-Nearest Neighbor

Tingkat akurasi prediksi bisa dilihat dari *Root Mean Squared Error* seperti pada gambar 9 dan 10. Nilai RMSE pada Rapid Miner bisa dilihat pada hasil kriteria *performance*. Nilai RMSE sendiri merupakan nilai kesalahan rata-rata kuadrat, dan bisa dijadikan sebagai salah satu faktor untuk menilai tingkat akurasi. Untuk dijadikan perbandingan tingkat akurasi, peneliti

menggunakan hasil *Relative Error* atau kesalahan relatif.



Gambar 14 *Relative error* regresi linier



Gambar 15 *Relative error k-nearest neighbor*

Seperti pada gambar 12 hasil relatif error regresi linier sebesar 71.95% +/- 82.63%. Sedangkan hasil relatif error *K-Nearest Neighbor* pada gambar 13 sebesar 30.95% +/- 29.87%. Hasil kesalahan relatif ini dijadikan sebagai perbandingan tingkat akurasi oleh peneliti, dimana hasil *relative error K-Nearest Neighbor* lebih kecil dibandingkan hasil *relative error* regresi linier. Hasil ini menunjukkan bahwa tingkat akurasi prediksi menggunakan algoritma *K-Nearest Neighbor* lebih tinggi dibandingkan dengan akurasi prediksi regresi linier.

4.2.3. Memastikan Model Regresi Linier Dapat Digeneralisasi Dengan Baik ke Pasar Perumahan yang Berbeda

Memastikan bahwa model regresi linier dapat digeneralisasi dengan baik ke pasar perumahan yang berbeda-beda melibatkan beberapa strategi dan pendekatan. Diantaranya yaitu:

- Pengumpulan data yang representatif, pastikan dataset yang digunakan untuk melatih model mencakup variasi yang mencukupi dari berbagai pasar perumahan.
- Feature engineering* yang bijaksana, mengidentifikasi fitur-fitur yang memiliki pengaruh seragam di berbagai pasar perumahan. Hindari menggunakan fitur-fitur yang sangat spesifik untuk satu pasar tertentu yang mungkin tidak umum di tempat lain.
- Normalisasi dan standarisasi, pastikan data dinormalisasi atau distandarisasi agar memiliki skala yang seragam. Ini membantu model untuk tidak mendiskriminasi fitur-fitur dengan skala yang berbeda-beda.
- Pengujian pada pasar perumahan yang berbeda, pisahkan data menjadi set pelatihan dan set pengujian yang mencakup berbagai pasar perumahan. Evaluasi kinerja model pada set

pengujian untuk memastikan bahwa model dapat digeneralisasi dengan baik ke pasar yang tidak dilihat selama pelatihan.

- Pemahaman konteks pasar, pahami faktor-faktor spesifik yang memengaruhi pasar perumahan di berbagai lokasi, dan pastikan model dapat menangkap variabilitas tersebut. Dengan menerapkan langkah-langkah ini, dapat meningkatkan kemampuan model regresi linier untuk digeneralisasi dengan baik ke pasar perumahan yang berbeda-beda.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa: Prediksi menggunakan algoritma regresi linier dapat mengidentifikasi variabel yang paling berpengaruh terhadap harga rumah. Variabel yang paling berpengaruh terhadap harga rumah adalah luas bangunan dengan nilai koefisien sebesar 23047431.201. Nilai perbandingan tingkat akurasi menggunakan algoritma KNN lebih baik daripada Regresi Linier berdasarkan hasil *Relative Error*. Nilai *relative error* regresi linier sebesar 71.95% +/- 82.63%. Sedangkan hasil *relatif error K-Nearest Neighbor* sebesar 30.95% +/- 29.87%. Semakin kecil nilai *relative error* maka semakin tinggi tingkat akurasi. Meningkatkan kemampuan model regresi linier untuk digeneralisasi dengan baik ke pasar perumahan yang berbeda dengan beberapa strategi atau pendekatan, seperti pengumpulan data yang representatif, *feature engineering* yang bijaksana, pemahaman konteks pasar, dll. Pemangku kepentingan dapat memanfaatkan informasi ini untuk menyesuaikan ekspektasi atau strategi dalam pasar properti.

Diperlukan kelanjutan penelitian yang bertujuan untuk memperbaiki serta memperluas penerapan algoritma regresi linier dan KNN dalam lingkup yang lebih luas. Pentingnya menjalankan penelitian yang lebih komprehensif dengan mempertimbangkan penggunaan metode lain sebagai pembanding terhadap data yang dihasilkan dari setiap metode penelitian. Dengan demikian, dapat diperoleh pemahaman yang lebih mendalam tentang keefektifan kedua algoritma ini dalam memprediksi harga rumah berdasarkan variabel variabel tertentu, serta meningkatkan akurasi prediksi hasil penelitian untuk pemahaman mendalam tentang faktor-faktor yang memengaruhi nilai properti.

DAFTAR PUSTAKA

- [1] G. N. Ayuni and D. Fitriana, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>
- [2] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J.*

- Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.
- [3] D. Novianty, N. D. Palasara, and M. Qomaruddin, “Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia,” *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 81, 2021, doi: 10.26418/justin.v9i2.43664.
- [4] G.- MARDIATMOKO, “Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Berganda,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 3, pp. 333–342, 2020, doi: 10.30598/barekengvol14iss3pp333-342.
- [5] P. Purwadi, P. S. Ramadhan, and N. Safitri, “Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang,” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 55, 2019, doi: 10.53513/jis.v18i1.104.
- [6] Harsiti, Z. Muttaqin, and E. Srihartini, “Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet,” *JSii (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 12–16, 2022, doi: 10.30656/jsii.v9i1.4426.
- [7] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, “Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [8] A. Anggrawan, N. Azmi, U. Bumigora, and I. Anthonyangrawan, “Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear Sales Prediction of Unilever Products using the Linear Regression Method,” *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.
- [9] T. Indarwati, T. Irawati, and E. Rimawati, “Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone,” *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.
- [10] M. N. Maskuri, Harliana, K. Sukerti, and R. M. H. Bhakti, “Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 1, pp. 130–140, 2022.