

IMPLEMENTASI METODE LINEAR REGRESSION UNTUK PREDIKSI HARGA PROPERTI REAL ESTATE MENGGUNAKAN RAPIDMINER

Umi Lathifah ¹, Raditya Danar Dana ²

¹ Komputerisasi Akuntansi, STMIK IKMI Cirebon

² Manajemen Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No.10B, Karyamulya, Kec.Kesambi, Kota Cirebon, Jawa Barat

umilatifah692@gmail.com

ABSTRAK

Investasi adalah pilihan banyak orang untuk mengelola keuangan secara efektif, melibatkan emas, saham, property dan *real estate*. *Real Estate* mencakup tanah dan segala yang permanen diatasnya. Usaha *real estate* memiliki potensi jangka panjang, meskipun dihadapkan ketidakpastian dalam menilai harga property, dipengaruhi oleh faktor kompleks seperti lingkungan, pasar, dan ekonomi yang sulit diprediksi dengan metode konvensional. Untuk mengatasi masalah ini, pendekatan utama adalah metode Linear Regresi dengan RapidMiner untuk memperoleh estimasi harga yang lebih akurat. Metodologi melibatkan pengumpulan data historis transaksi property, identifikasi fitur relevan, pemrosesan data, dan implementasi model Linear Regresi. Pemilihan fitur, normalisasi data, evaluasi model menjadi fokus utama. Tujuan penelitian ini adalah mengembangkan model prediktif untuk estimasi harga property yang akurat dalam industri *real estate*. Dengan metode Linear Regresi, diharapkan model memberikan wawasan terinci tentang dinamika harga properti dalam pasar yang berubah. Hasil evaluasi dengan *Root Mean Square Error* (RMSE) 0.116 dan *Mean Absolute Error* (MAE) 0.088 menunjukkan performa yang baik, dengan nilai RMSE kecil dalam rentang 0-1. Selain hasil RMSE dan MAE yang baik juga menghasilkan prediksi harga yang cukup baik

Kata Kunci : Harga, Prediksi, Linear Regresi, Real Estate, RapidMiner

1. PENDAHULUAN

Konsep investasi melibatkan pengalokasian dana atau sumber daya pada asset tertentu dengan harapan mendapatkan keuntungan di masa depan. Investasi bisa berupa instrument keuangan seperti, saham, obligasi, property, komoditas, emas dan banyak lagi. Investasi dalam bentuk property bangunan yaitu *real estate* telah menjadi salah satu pilihan utama bagi banyak orang yang ingin membangun kekayaan jangka panjang. Menurut Otoritas Jasa Keuangan (OJK), *real estate* adalah termasuk tanag dan segala bangunan fisik yang terdapat di atasnya, termasuk segala benda yang terikat pada tanah seperti Gedung, pagar, dan struktur lainnya. Dari segi *historis*, properti *real estate* telah terbukti sebagai asset yang stabil dan memiliki potensi pertumbuhan nilai yang cukup signifikan

Tinggal di daerah metropolitan adalah janji yang pasti untuk semua karena harga investasi sangat tinggi dan harga tanah, rumah dan fasilitas umum sangat wajar dan harga bisa naik setiap tahun. Penelitian ini akan dilakukan untuk prediksi harga *real estate* menggunakan metode linear regresi yang bertujuan untuk menghasilkan model prediktf yang dapat membantu pemangku kepentingan dalam membuat keputusan investasi yang lebih informasi terkait harga properti di pasar *real estate*.

Beberapa penelitian terdahulu menunjukkan bahwa data mining dapat digunakan dengan baik dalam prediksi harga *real estate*. Penelitian yang dilakukan oleh [1] dengan judul “Prediksi Prospek Harga Real Estate di Masa Pandemi dengan 3 Atribut Berbasis Algoritma Linear Regression” pada tahun 2022. Menunjukkan hasil penggunaan metode regresi

linear untuk mendapatkan sebuah prediksi harga *real estate* terhadap 3 atribut yang mempengaruhi harga *real estate*, untuk harga prospek *real estate* di masa pandemic tidak mengalami kenaikan atau penurunan dari segi harga jual dengan akurasi yang didapatkan adalah 67,8% .[2]

Selanjutnya penelitian berjudul “Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ” Metode linear regresi dalam meramalkan penjualan properti, berdasarkan hasil penelitian, dapat dianggap cukup baik. Kesimpulan ini diperoleh dari uji *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) terhadap seluruh jenis properti di PT.XYZ yang memenuhi standar. Penelitian ini memberikan hasil berupa estimasi berdasarkan data pada periode tertentu. Meskipun demikian, hasil penelitian ini dapat menjadi pertimbangan bagi pengembang dalam pengambilan keputusan terkait pembangunan, dengan tujuan untuk mengurangi pembangunan berlebih disektor properti yang mengalami penurunan, sementara memaksimalkan di sektor yang mengalami peningkatan.[3]

Dari beberapa penelitian tersebut dapat disimpulkan bahwa, metode linear regresi dapat digunakan untuk memprediksi harga property real estate dengan baik

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah istilah yang digunakan untuk menemukan informasi yang tersembunyi dalam

sebuah basis data. Proses ini melibatkan ekstraksi pola yang bermanfaat atau pengetahuan berharga dari kumpulan data besar. *Data mining* memanfaatkan Teknik analisis statistic, matematika, teknologi kecerdasan buatan, dan pembelajaran mesin untuk mengeksplorasi serta menganalisa data yang ada. Tujuan utamanya adalah mengidentifikasi pola tersembunyi, relasi, atau informasi yang dapat digunakan untuk pengambilan keputusan atau perencanaan dimasa mendatang. *Data mining* memiliki aplikasi yang luas, dari bisnis dan keuangan hingga ilmu pengetahuan, kedokteran, dan pemerintah. Ini membantu organisasi untuk memahami pelanggan, mengidentifikasi tren pasar, meningkatkan efisiensi operasional, dan mengoptimalkan keputusan strategis. *Data mining* juga dikenal dengan istilah lain, seperti Knowledge Discovery in Databases (KDD) dan *Data Analysis*. Selain itu juga ada *Knowledge Extraction*, *Business Intelligence*, *Data Dreging*, *Information Harvesting*, *Data Archeologi*, dan lainnya. Proses ekstraksi data melibatkan serangkaian langkah dan metode, mulai dari pembersihan data, integrasi data, pemilihan data, dan transformasi data, hingga evaluasi pola untuk mendapatkan informasi dari data tersebut [4]

2.2. Peramalan / Prediksi

Prediksi atau peramalan adalah melibatkan perkiraan tentang apa yang akan mungkin terjadi di masa depan dengan melibatkan analisis data dari masa lalu atau saat ini, baik secara matematis atau statistic. Prediksi bisa berupa kualitatif (non-angka) atau kuantitatif (berupa angka). Prediksi Kualitatif lebih sulit dikarenakan variabel cenderung relative, sementara prediksi kuantitatif sangat bergantung pada metode yang digunakan. Berbagai metode akan menghasilkan nilai prediksi yang berbeda dan perancangan metode prediksi. Terdapat tiga tahap pada prediksi yaitu analisis data masa lampau untuk memahami pola, pemilihan metode yang sesuai dengan kebutuhan, dan mengukur hasil ramalan dengan tingkat kesalahan terkecil. Tidak ada metode yang pasti cocok untuk semua jenis data, dan terkadang tranformasi data diperlukan sesuai kebutuhan. [5]

Untuk menilai keakuratan suatu prediksi, dapat dilihat dari nilai *Root Mean Square Error* RMSE dan *Mean Absolute Error* (MAE). RMSE adalah akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya. RMSE memberikan penekanan pada kesalahan besar karena menggunakan kuadrat kesalahan, dan akar kuadrat digunakan untuk membawa skala kembali ke satuan awal. MAE adalah rata-rata dari nilai absolute dari selisih antara nilai prediksi dan nilai sebenarnya. MAE memberikan gambaran tentang kesalahan rata-rata tanpa memperhatikan arah kesalahan. Berikut ini adalah rumus RMSE dan MAE.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (1)$$

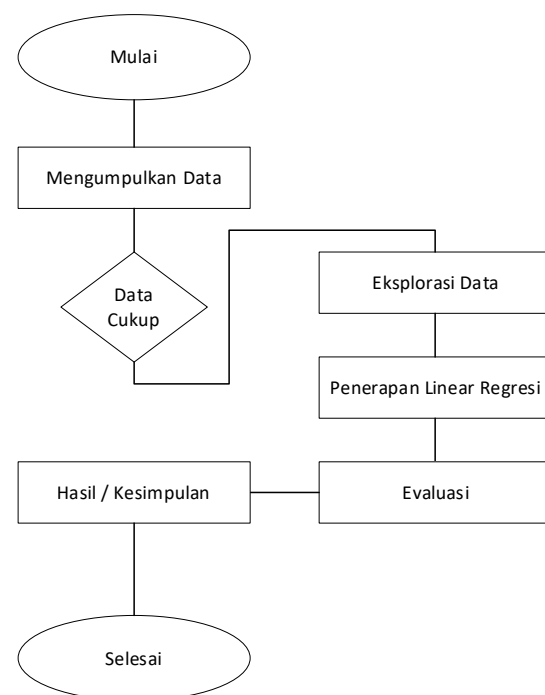
Di mana Y_t adalah nilai aktual pada periode t , dan y_t^1 adalah nilai prediksi pada periode t . Jumlah prediksi diwakili oleh nilai n .

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Dimana n adalah jumlah observasi atau sampel, kemudian y_1 adalah nilai sebenarnya untuk observasi ke- i lalu $\hat{y}_i$ adalah nilai prediksi untuk observasi ke- i , dan $|y_i - \hat{y}_i|$ merupakan nilai absolut dari selisih antara nilai sebenarnya dan nilai prediksi untuk setiap observasi.

2.3. Linear Regresi

Linear Regresi adalah alat statistik yang digunakan untuk meneliti dampak antara satu atau lebih variabel terhadap variabel lainnya. Variabel yang berpengaruh biasanya disebut variabel bebas atau independen, sementara variabel yang dipengaruhi disebut variabel terikat atau dependen.



Gambar 1. Flowchart Linear Regresi

Langkah awa dalm flowchart menunjukkan awal dari proses, selanjutnya adalah langkah proses pengumpulan data dari sampel lalu dirasa data yang dibutuhkan sudah cukup dilakukan tahap eksplorasi data yaitu proses analisis deskriptif data untuk memahami karakteristik dasar. Setelah selesai pada tahapan ini, selanjutnya adalah penerapan Linear Regresi dalam sebuah model bertujuan untuk mengasilkan model matematis yang dapat digunakan untuk membuat prediksi sebuah data, selanjutnya dilakukan evaluasi model seperti penggunaan data

pengujian dan matrik evaluasi, dan selesai. Metode ini hanya cocok untuk skala interval dan ratio. Regresi linear sederhana membantu mengetahui hubungan antara satu variabel bebas dan satu variabel terikat melalui rumus persamaan :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (3)$$

Dalam persamaan tersebut, Y melambangkan variabel dependen yang bergantung pada nilai variabel X, yang merupakan variabel independen. Nilai a merupakan suatu konstanta, sementara b merupakan koefisien regresi yang berkaitan dengan variabel X. Untuk memperoleh nilai a dan b sehubungan dengan nilai variabel X, kita dapat menyusun persamaan yang sesuai.[6]

2.4. RapidMiner

RapidMiner adalah perangkat lunak yang berperan sebagai alat pembelajaran untuk keperluan ilmu data mining. Platform ini dikembangkan oleh sebuah perusahaan yang berkomitmen pada segala aspek yang terkait dengan manipulasi data dalam skala besar, termasuk dalam konteks bisnis komersial, penelitian, Pendidikan, pelatihan, dan pembelajaran. RapidMiner menyediakan sekitar 100 solusi pembelajaran untuk tugas seperti pengelompokkan, klasifikasi, dan analisis regresi. Selain itu, platform ini mendukung sekitar 22 format file yang berbedad, seperti .xls, .csv, dan lainnya. Dengan membawa kecerdasan buatan ke dalam dunia perusahaan, RapidMiner hadir sebagai platform ilmu pengetahuan data yang terbuka, dapat diskalakan, dan dapat diintegrasikan dengan mudah melalui berbagai Bahasa Pemrograman, karena ditulis menggunakan bahasa Java. Lebih dari 625.000 profesional analitis memanfaatkan produk RapidMiner untuk meningkatkan pendapatan, mengurangi biaya, dan mengelola risiko.

RapidMiner adalah perangkat lunak independen yang digunakan untuk menganalisis data dan mesin penambahan data, yang dapat berinteraksi dengan berbagai bahasa pemrograman secara mudah. RapidMiner menyediakan UI untuk mendesain pipa analisis, di mana akan menghasilkan file XML yang dapat menjelaskan proses analisis yang ingin diterapkan oleh pengguna ke data. RapidMiner menyajikan antarmuka pengguna (UI) yang bersahabat, sehingga mempermudah pengguna dalam mengoperasikannya. Tampilan yang terdapat pada RapidMiner disebut Perspective, dengan 3 Perspective utama: Welcome Perspective, Design Perspective, dan Result Perspective. Saat pengguna membuka aplikasi RapidMiner, mereka akan melihat Welcome Perspective, yang merupakan tampilan pertama yang muncul. Design Perspective adalah tampilan utama di mana pengguna bekerja di RapidMiner. Ini digunakan sebagai tempat untuk merancang dan mengelola proses analisis. Tampilan ini menampilkan berbagai tampilan dengan fungsi masing-masing, yang mendukung proses analisis data. Sementara itu, Result

Perspective adalah tampilan di mana hasil analisis ditampilkan dalam berbagai bentuk, seperti grafik, teks, pohon keputusan, tabel, tergantung pada data dan metode analisis yang digunakan [7]

2.5. Real Estate

Real estate adalah istilah yang merujuk pada sektor ekonomi yang mencakup aktivitas penjualan, pembelian, serta penyewaan atas tanah dan bangunan, yang meliputi segala jenis properti mulai dari rumah tinggal, gedung perkantoran, pusat perbelanjaan, kompleks industri, hingga lahan kosong. Industri real estate tidak hanya terbatas pada transaksi properti, tetapi juga melibatkan proses pengembangan, manajemen properti, penilaian nilai properti, serta aktivitas investasi yang terkait dengan aset fisik ini. Dengan kompleksitasnya, industri real estate memiliki peran yang signifikan dalam perekonomian suatu negara karena berkaitan erat dengan pertumbuhan ekonomi, investasi, dan infrastruktur. [8]

3. METODE PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa kumpulan data harga real estate yang diperoleh dari situs <https://www.kaggle.com/datasets/oddyvirgantara/harga-real-estate> yang diakses pada hari Senin, 18 Desember 2023 pukul 07.15 WIB, serta data yang diperoleh melalui jurnal-jurnal terkait penelitian mengenai prediksi *data mining* dengan metode *Linear Regresi*.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data merupakan proses yang digunakan dalam mengumpulkan data untuk penelitian. Adapun Teknik pengumpulan data yang digunakan dalam penelitian ini adalah:

a. Observasi

Observasi dilakukan dengan mengamati langsung ke tempat penelitian untuk mengumpulkan informasi yang akan dilakukan penelitian. Namun untuk penelitian ini observasi tidak dilakukan secara langsung ke dalam tempat atau perusahaan tertentu, mengingat data yang digunakan diperoleh dari internet. Oleh karena itu, hanya dilakukan peninjauan data saja yang bertujuan untuk mendapatkan dataset yang sesuai dengan topik penelitian yang akan dibahas.

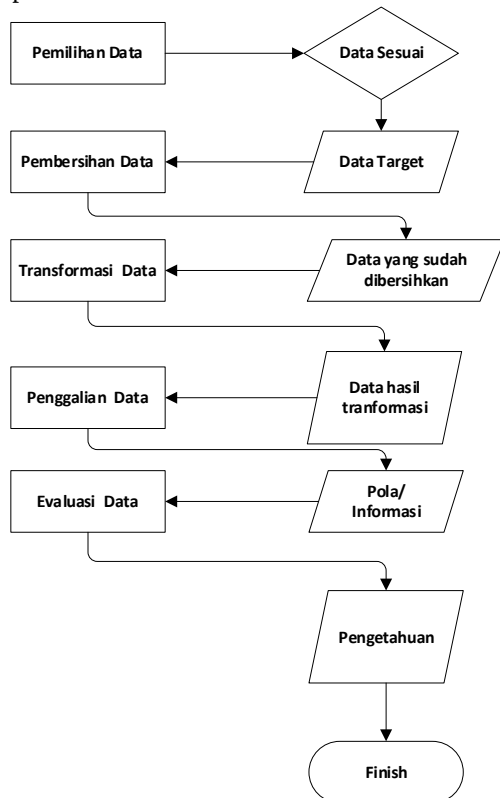
b. Studi Pustaka

Studi Pustaka dilakukan dengan mengumpulkan dan mempelajari berbagai jurnal ilmiah sebagai referensi yang berkaitan dengan topik permasalahan dan metode yang sesuai dengan penelitian ini.

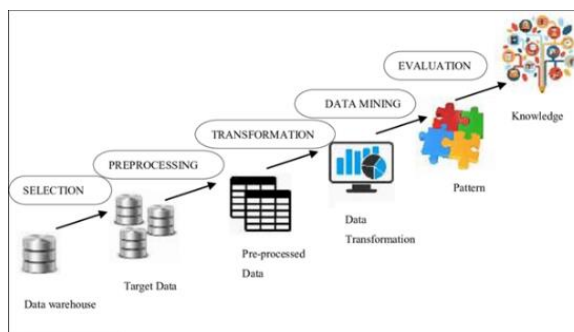
3.3. Tahapan Perancangan

Tahapan perancangan penerapan *data mining* dalam penelitian ini menggunakan tahapan *Knowledge Discovery in Database* (KDD) dan bantuan *tools*

RapidMiner. KDD adalah proses yang dilakukan untuk menemukan informasi dan pola yang ada dalam data. Berikut ini adalah *flowchart* yang menunjukkan tahapan KDD.



Gambar 2. Flowchart KDD



Gambar 3. Tahapan KDD

Ada lima tahapan utama KDD. Berikut ini adalah penjelasan tahapan tahapan KDD:

1. Data Selection (Pemilihan Data)

Pada tahap ini, data yang relevan akan dipilih dan dipersiapkan untuk dianalisis. Pada dataset *real estate* atribut-atribut yang akan dipilih sesuai dengan kebutuhan dalam penelitian yang sedang dilakukan. Pada dataset ini terdiri dari 4 atribut, yang mana seluruh atribut pada data ini akan digunakan dalam analisis.

2. Data Pre-Processing (Pembersihan Data)

Pada proses atau tahapan ini dilakukan pembersihan data terhadap data duplikat, data *missing value*, data yang bernilai sama, ataupun data yang tidak memiliki korelasi dengan variabel

yang lain. Pada dataset *real estate* ini tidak terdapat data data yang harus dibersihkan, maka proses akan dilanjutkan ke tahapan berikutnya.

3. Data Transformation (Transformasi Data)

Tahapan selanjutnya adalah dilakukannya transformasi data ke dalam format yang sesuai dengan tipe data yang akan digunakan dalam proses *data mining*. Proses transformasi dilakukan secara dengan mengubah role atribut yang akan dijadikan sebagai id dan label.

4. Data mining

Tahapan ini adalah proses utama untuk menemukan pola atau informasi dari data yang sudah disiapkan melalui proses sebelumnya adalah dengan menerapkan metode tertentu. Pada penelitian kali ini adalah menggunakan metode *linear regresi* untuk memprediksi harga *real estate* berdasarkan data tersebut.

5. Data Interpretation / Evaluation (Evaluasi)

Pada tahapan terakhir ini adalah proses evaluasi dari pola yang sudah dihasilkan dalam *data mining*. Evaluasi dilakukan dengan menggunakan performa *Root Mean Squared Error (RMSE)* dan *MAE (Mean Absolute Error)* untuk mendapatkan nilai akurasi. Penerapan metode *Linear Regresi* diharapkan penelitian ini dapat menghasilkan informasi tentang penerapan *data mining* untuk memprediksi harga *real estate* dengan nilai *Root Mean Squared Error (RMSE)* dan *MAE (Mean Absolute Error)* yang baik.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Implementasi metode Linear Regresi

Pada tahapan akan dibuat sebuah model data untuk memprediksi data *real estate* dengan menggunakan aplikasi *RapidMiner*. Berikut adalah tahapan tahapannya:

4.2. Hasil Pengumpulan Data (Data Collecting)

Data yang digunakan pada penelitian ini adalah data yang diperoleh dari website *Kaggle.com* tahun 2021. Pada data ini terdapat 4 atribut yaitu usia rumah, jarak mrt, jumlah toko, dan harga per unit. Banyaknya data yang diperoleh adalah 414 record.

Tabel 1. Daftar Harga *Real Estate*

NNo	Xusia rumah	X2 Jarak MRT	X3 Jumlah Toko	Y Harga Unit
11	32	84,87882	10	37.900
22	19,5	306,5947	9	42.200
33	13,3	561,9845	5	47.300
44	5	390,5684	5	43.100
55	32	84,87882	10	37.900
...	...	...	...	...

4.3. Data Selection

Untuk memilih atribut data dalam dataset harga *real estate* ini, semua atribut akan dipilih dan diproses pada tahapan berikutnya. Operator yang digunakan dalam proses seleksi data adalah sebagai berikut:

Gambar 4. Operator *Retrive*

Operator *Retrive* digunakan untuk membaca sebuah data dari berbagai sumber seperti database, file lokal, atau bahkan koneksi internet.

4.4. Pre-Processing Data

Processing Data adalah tahapan data mining dalam mengolah data dari mentah menjadi data yang valid dan siap untuk digunakan.



Gambar 5. Implementasi Operator Outlier

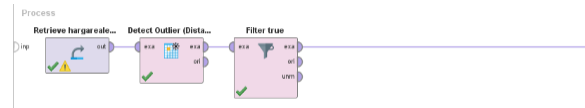
Tahap Pre-Processing yang pertama adalah deteksi Outlier menggunakan Operator Outlier berguna untuk mengidentifikasi dan menangani data data yang tidak biasa atau data yang ekstrem. Hal ini dapat membantu dalam membersihkan data dari gangguan atau nilai yang tidak biasa yang bisa mempengaruhi dalam proses analisis.

Row No.	Y_harga_unit	outlier	No	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_t...
1	37.900	false	1	32	84.879	10
2	42.200	false	2	20	306.595	9
3	47.300	false	3	13	561.985	5
4	54.800	false	4	13	561.985	5
5	43.100	false	5	5	390.568	5
6	32.100	false	6	7	2175.030	3
7	40.300	false	7	35	623.473	7
8	46.700	false	8	20	287.603	6
9	18.800	true	9	32	5512.038	1
10	22.100	false	10	18	1783.180	3
11	12.010	false	11	35	405.213	1
12	61.900	false	12	6	90.456	9
13	39.300	false	13	13	492.231	5
14	23.800	false	14	20	2469.645	4
15	48	false	15	13	1164.838	4

ExampleSet (414 examples, 2 special attributes, 4 regular attributes)

Gambar 6. Hasil Outlier

Hasil dari proses Outlier tersebut menunjukkan terdapat data yang menunjukkan keterangan false dan true. True berarti bahwa suatu data telah diidentifikasi sebagai outlier atau dianggap sebagai nilai yang tidak biasa, sedangkan False berarti data tidak dianggap sebagai outlier. Dan pada hasil tersebut terdapat beberapa data yang bernilai 0 sehingga perlu dilakukan tahapan selanjutnya yaitu Filter Example.



Gambar 7. Filter Example (True)

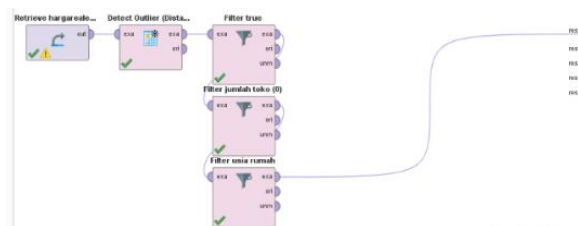
Tahapan selanjutnya adalah memfilter atau menghapus sebagian data, pada tahapan ini akan menghapus outlier bernilai True.

Row No.	Y_harga_unit	outlier	No	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_t...
1	37.900	false	1	32	84.879	10
2	42.200	false	2	20	306.595	9
3	47.300	false	3	13	561.985	5
4	54.800	false	4	13	561.985	5
5	43.100	false	5	5	390.568	5
6	32.100	false	6	20	2175.030	3
7	40.300	false	7	35	623.473	7
8	46.700	false	8	20	287.603	6
9	22.100	false	10	6	1783.180	3
10	12.010	false	11	30	405.213	1
11	61.900	false	12	37	90.456	9
12	39.300	false	13	4	492.231	5
13	23.800	false	14	7	2469.645	4
14	48	false	15	15	1164.838	4

ExampleSet (404 examples, 2 special attributes, 4 regular attributes)

Gambar 8. Hasil Filter example (true)

Tahapan selanjutnya adalah masih sama yaitu proses Filter, namun tahapan selanjutnya yaitu menghapus data bernilai 0 pada atribut jumlah toko dan usia rumah



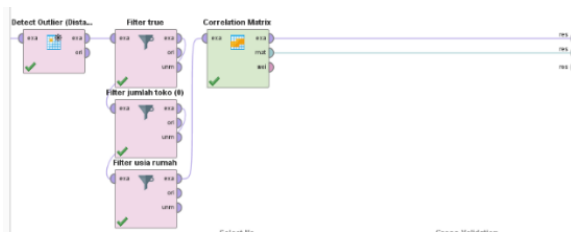
Gambar 9. Filter data bernilai 0

Row No.	Y_harga_unit	outlier	No	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_t...
1	37.900	false	1	32	84.879	10
2	42.200	false	2	20	306.595	9
3	47.300	false	3	13	561.985	5
4	54.800	false	4	13	561.985	5
5	43.100	false	5	5	390.568	5
6	32.100	false	6	20	2175.030	3
7	40.300	false	7	35	623.473	7
8	46.700	false	8	20	287.603	6
9	22.100	false	10	6	1783.180	3
10	12.010	false	11	30	405.213	1
11	61.900	false	12	37	90.456	9
12	39.300	false	13	4	492.231	5
13	23.800	false	14	7	2469.645	4
14	48	false	15	15	1164.838	4

ExampleSet (334 examples, 2 special attributes, 4 regular attributes)

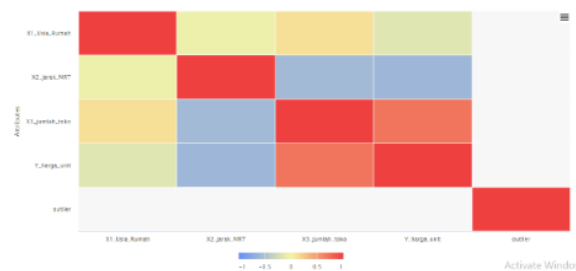
Gambar 10. Hasil Filter data bernilai 0

Setelah selesai tahapan *Outlier* dan Filterisasi selanjutnya mengukur Correlation Matrix.



Gambar 11. Correlation Matrix

Operator *Correlation Matrix* berguna untuk menghitung korelasi matrix antara variabel variabel yang ada didataset. Selain untuk menghitung korelasi matrix operator ini juga digunakan untuk menganalisis hubungan antar variabel.



Gambar 12. Hasil Correlation Matrix

Hasil *Correlation Matrix* dimana nilai yang mendekati 1 menunjukkan hubungan korelasi yang kuat antara variabel Y dengan variabel X sementara nilai yang mendekati 0 yaitu hubungan korelasi yang lemah. Pada hasil visualisasi hasil *Correlation Matrix* tersebut menunjukkan bahwa warna yang paling gelap menunjukkan hubungan korelasi yang lebih kuat. Pada hasil *Correlation* tersebut menghasilkan nilai terbesar adalah 0.709 yang berarti variabel X_jumlah_toko yang paling berpengaruh pada variabel Y_harga_unit dimana variabel X dan Y menunjukkan adanya hubungan linier positif yang kuat antara kedua variabel tersebut, ketika variabel X meningkat maka nilai variabel Y cenderung meningkat juga. Selanjutnya nilai -0.618 untuk korelasi Y_harga_rumah dengan X2_jarak_MRT hal ini berarti hasil korelasi variabel X dan Y menunjukkan adanya hubungan linier negative yang cukup kuat antar kedua variabel tersebut. Angka negative menunjukkan arah hubungan, dalam hal ini, hubungan linier negative. Ketika nilai variabel X meningkat, nilai variabel Y cenderung menurun.

LinearRegression

$$\begin{aligned}
 &- 0.193 * X1_Usia_Rumah \\
 &- 0.523 * X2_jarak_MRT \\
 &+ 0.423 * X3_jumlah_toko \\
 &+ 0.355
 \end{aligned}$$

Gambar 13. Formula Linear Regresi

Hasil yang terdapat dalam Gambar 4.7 adalah representasi dari model Regresi Linear dalam bentuk persamaan matematika yang menghubungkan

variabel-variabel prediktor ($X1_Usia_Rumah, X2_jarak_MRT, X3_jumlah_toko$) dengan variabel yang ingin diprediksi, yaitu variabel Y_Harga_Unit . Tiap koefisien yang menyertainya memberikan petunjuk seberapa besar pengaruh variabel tersebut terhadap nilai target. Formula tersebut bisa digunakan untuk menghitung data baru mendatang.

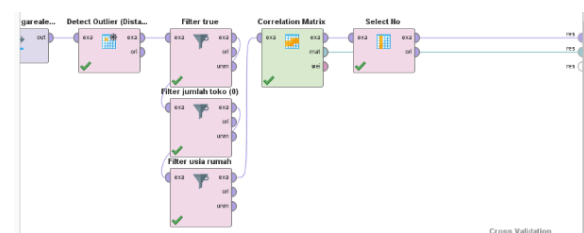
$-0.193 * X1_Usia_Rumah$: Setiap kali Usia Rumah ($X1$) mengalami perubahan satu unit, akan ada perubahan sebesar 0.193 unit dalam nilai target. Misalnya, jika usia rumah meningkat satu unit, variabel target akan berubah sebesar 0.283.

$-0.523 * X2_jarak_MRT$: Tiap kali Jarak ke Stasiun MRT ($X2$) berubah satu unit, akan terjadi perubahan sebesar 0.523 unit dalam variabel target. Artinya, jika jarak ke stasiun MRT meningkat satu unit, variabel target akan naik sebesar 0.005.

$0.423 * X3_jumlah_toko$: Setiap kali Jumlah Toko ($X3$) meningkat satu unit, akan terjadi kenaikan sebesar 0.423 unit dalam nilai target. Sebagai contoh, jika jumlah toko bertambah satu unit, variabel target akan naik sebesar 1.369.

Selain faktor-faktor yang digunakan untuk memprediksi, ada juga nilai tetap atau intercept sebesar 0.355. Ini mengindikasikan bahwa jika semua faktor prediksi memiliki nilai nol, prediksi nilai target akan mendekati 0.355. Pengetahuan ini membantu dalam memahami sejauh mana kontribusi masing-masing faktor terhadap nilai target dalam model regresi linear yang telah dibuat.

Setelah selesai dilakukan tahapan *Correlation Matrix* selanjutnya adalah tahapan menyeleksi atribut yang tidak terpakai, pada kali ini, adalah penghapusan atribut bernama No.



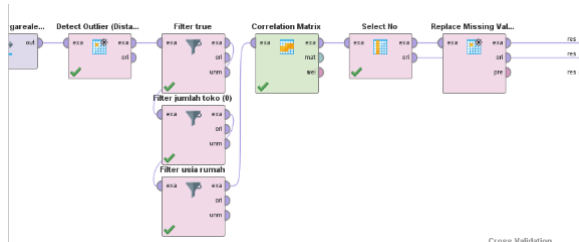
Gambar 14. Select Attribute

Salah satu tahapan *Data Reduction* yaitu adalah *Select Attributes*. *Select Attributes* berguna untuk memilih atau menyaring atribut atribut yang diperlukan untuk analisis atau permodelan lebih lanjut.

Row No.	Y_harga_unit	outlier	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_t...
1	0.398	false	0.727	0.010	1
2	0.462	false	0.455	0.044	0.900
3	0.538	false	0.295	0.083	0.500
4	0.650	false	0.295	0.083	0.500
5	0.475	false	0.114	0.057	0.500
6	0.311	false	0.455	0.333	0.300
7	0.434	false	0.795	0.093	0.700
8	0.529	false	0.455	0.041	0.600
9	0.162	false	0.136	0.272	0.300
10	0.012	false	0.682	0.059	0.100
11	0.756	false	0.841	0.010	0.900
12	0.419	false	0.091	0.073	0.500
13	0.188	false	0.159	0.378	0.400
14	0.548	false	0.341	0.177	0.400
15	0.280	false	0.227	0.086	0.200

ExampleSet (333 examples, 2 special attributes, 3 regular attributes)

Gambar 15. Hasil Select Attribute



Gambar 16. Replace Missing Value

Tahapan operator ini digunakan untuk menangani masalah yang disebabkan oleh nilai yang hilang atau kosong dalam dataset.

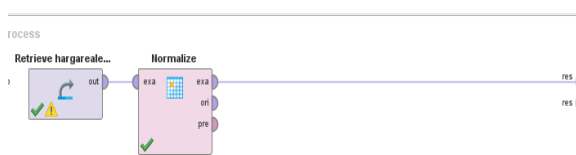
Label	Real	0	Min	11.500	Max	78.300
Y_harga_unit						
Outlier	Binominal	0	Negative	false	Positive	true
X1_Usia_Rumah	Integer	0	Min	1	Max	44
X2_jarak_MRT	Real	0	Min	23.383	Max	4510.359
X3_jumlah_toko	Integer	0	Min	1	Max	10

Gambar 17. Hasil Replace Missing Value

Pada hasil gambar diatas menunjukkan bahwa dalam dataset ini tidak menunjukkan adanya dataset yang kosong atau hilang.

4.5. Transformasi Data

Pada tahapan ini bertujuan untuk memanipulasi atau mengubah struktur, format, atau nilai-nilai yang terdapat dalam dataset agar sesuai dengan kebutuhan analisis atau permodelan data yang akan dilakukan. Dalam tahapan ini akan dilakukan proses normalisasi data .



Gambar 18. Normalisasi Data

Parameters

Normalize

create view

attribute filter type all

invert selection

include special attributes

method range transforma...

min 0.0

max 1.0

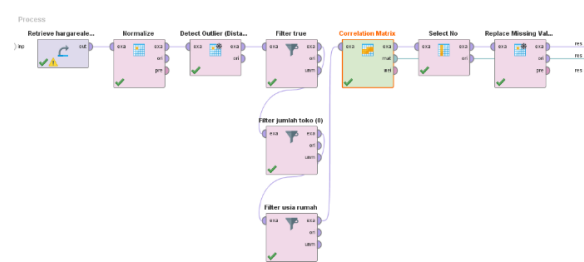
Gambar 19. Parameter Normalisasi

Pada parameter ini menentukan metode normalisasi range transformation min-max. Atau penggunaan nilai minimum dan nilai maksimum untuk melakukan konversi data secara linier. Pada parameter tersebut rentang nilai 0-1, minimum bernilai 0 dan maximum bernilai 1.

Row No.	Y_harga_unit	No	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_t...
1	0.398	0	0.727	0.010	1
2	0.462	0.002	0.455	0.044	0.900
3	0.538	0.005	0.295	0.083	0.500
4	0.650	0.007	0.295	0.083	0.500
5	0.475	0.010	0.114	0.057	0.500
6	0.311	0.012	0.455	0.333	0.300
7	0.434	0.015	0.795	0.093	0.700
8	0.529	0.017	0.455	0.041	0.600
9	0.113	0.019	0.795	0.849	0.100
10	0.162	0.022	0.136	0.272	0.300
11	0.012	0.024	0.682	0.059	0.100
12	0.756	0.027	0.841	0.010	0.900
13	0.419	0.029	0.091	0.073	0.500
14	0.188	0.031	0.159	0.378	0.400
15	0.548	0.034	0.341	0.177	0.400

ExampleSet (414 examples, 1 special attribute, 4 regular attributes)

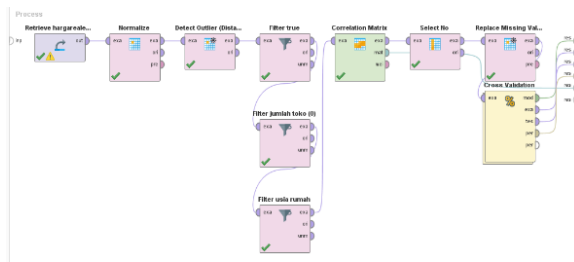
Gambar 20. Hasil Normalisasi



Gambar 21. Model Pre-Processing

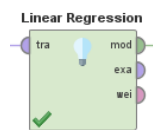
4.6. Data Mining

Tahapan ini adalah tahapan pengimplementasian metode atau tahapan utama dalam membangun sebuah model prediksi.



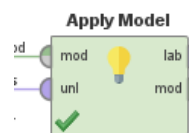
Gambar 22. Cross Validation

Operator ini digunakan untuk mengevaluasi kinerja model dengan meminimalkan risiko overfitting atau kesalahan evaluasi yang tidak akurat.



Gambar 23. Linear Regression

Selanjutnya tahapan implementasi algoritma Linear Regresi, penempatan operator ini berada didalam operator *cross validation*

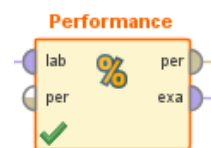


Gambar 24. Apply Model

Operator ini digunakan untuk menerapkan model machine learning yang sudah dilatih pada dataset harga *real estate*.

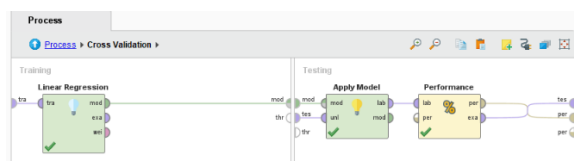
4.7. Evaluation

Tahapan ini adalah tahapan terakhir dalam implementasi metode linear regresi yaitu tahapan evaluation. Pada tahapan ini menggunakan operator:

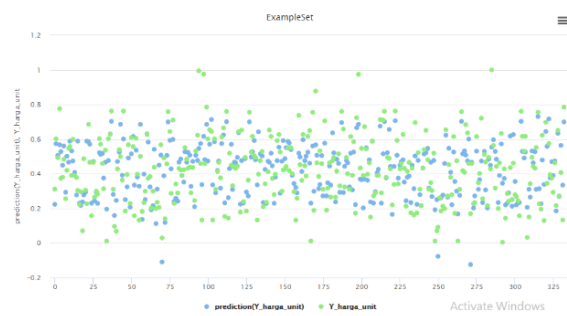


Gambar 25. Performance

Pada tahapan terakhir ini adalah untuk memberikan *insight* dan evaluasi terhadap kinerja model yang sudah dibangun.



Gambar 26. Model didalam cross validation



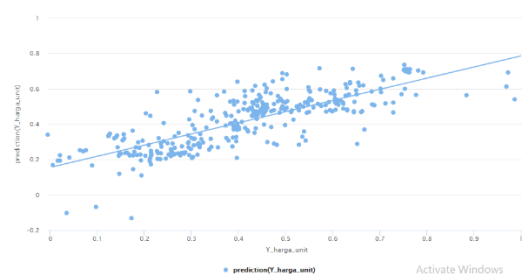
Gambar 27. Visualisasi Perbandingan Prediksi

Dari Gambar 27 visualisasi perbandingan hasil prediksi harga *real estate* dengan harga asli menggunakan metode linear regresi menghasilkan hasil prediksi yang cukup baik. Dimana letak titik prediksi dengan titik harga asli tidak memiliki rentang yang sangat jauh, kedua titik cukup berdekatan satu sama lain. Pada grafik tersebut titik berwarna biru melambangkan hasil prediksi harga unit sedangkan titik berwarna hijau melambangkan harga real..

Row No.	Y_harga_unit	prediction(Y...	outlier	X1_Usia_Ru...	X2_jarak_M...	X3_jumlah_L...
1	0.311	0.224	false	0.455	0.333	0.300
2	0.602	0.574	false	0.318	0.040	0.700
3	0.493	0.506	false	0.841	0.072	0.800
4	0.775	0.568	false	0.114	0.037	0.600
5	0.374	0.528	false	0.750	0.062	0.800
6	0.382	0.450	false	0.705	0.055	0.600
7	0.484	0.561	false	0.341	0.056	0.700
8	0.256	0.294	false	0.091	0.329	0.300
9	0.420	0.500	false	0.750	0.027	0.700
10	0.548	0.467	false	0.318	0.073	0.500
11	0.557	0.589	false	0.023	0.026	0.600
12	0.392	0.532	false	0.682	0.079	0.800
13	0.599	0.474	false	0.364	0.041	0.500
14	0.390	0.407	false	0.636	0.076	0.500

ExampleSet (333 examples, 3 special attributes, 3 regular attributes)

Gambar 28. Hasil Prediksi Implementasi Linear Regresi



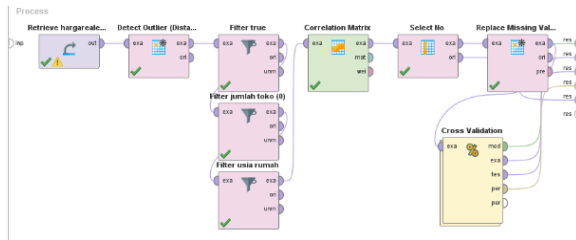
Gambar 29. Visualisasi hasil prediksi

Pada Gambar 29 menunjukkan hasil nilai prediksi, dapat dilihat bahwa hasil prediksi cukup akurat dimana titik titik berwarna biru mengelilingi garis linear. Garis ini dapat diinterpretasikan sebagai suatu representasi visual dari model regresi yang digunakan untuk memprediksi harga.

4.8 Hasil Performa Prediksi menggunakan metode Linear Regresi pada RapidMiner

Hasil dari performa prediksi menggunakan metode linear regresi adalah berupa nilai RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute*

Error) berikut ini adalah hasil permodelan dengan menggunakan metode Linear Regresi.



Gambar 30. Proses Uji Performa Model Linear Regresi

root_mean_squared_error

root_mean_squared_error: 0.116 +/- 0.018 (micro average: 0.117 +/- 0.000)

Gambar 31. Hasil RMSE

Hasil evaluasi performa model regresi linear ditunjukkan dalam Gambar 4.28. Perhitungan RMSE menghasilkan nilai yang sangat baik, yakni 0.116. Hasil RMSE tersebut menandakan jika hasil performa linear regresi sangat baik karena hasil sesuai dengan yang diharapkan, yaitu antara 0 dan 1. Nilai RMSE yang kecil menunjukkan bahwa prediksi model cenderung dekat dengan nilai sebenarnya. Semakin kecil RMSE, semakin kecil selisih antara nilai yang diprediksi dan nilai yang diamati.

absolute_error

absolute_error: 0.088 +/- 0.012 (micro average: 0.088 +/- 0.078)

Gambar 32. Hasil MAE

Hasil MAE (*Mean Absolute Error*) menghasilkan nilai 0.088 dimana nilai tersebut tergolong baik, karena pada umumnya semakin kecil nilai *absolute_error* semakin baik. Secara umum, ketika nilai *absolute_error* mencerminkan kesalahan atau deviasi dari nilai yang diharapkan, diinginkan untuk mencapai nilai yang seminim mungkin. Oleh karena itu, membandingkan nilai ini dengan standar atau ambang batas tertentu dapat bermanfaat dalam mengevaluasi seberapa baik performa system atau model.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan pada data harga *real estate* tahun 2021 dengan 414 *record* data, penggunaan metode Linear Regresi melalui aplikasi RapidMiner telah berhasil menghasilkan prediksi harga *real estate* yang cukup baik. Hasil analisis menunjukkan bahwa implementasi metode ini menghasilkan garis linear yang memadai untuk prediksi harga. Performa model dievaluasi dengan menggunakan nilai RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*) yang menghasilkan angka sebesar 0.116 dan 0.088. RMSE cenderung memberikan bobot lebih besar yaitu 0.116 pada

kesalahan besar karena mengkuadratkan setiap kesalahan. sehingga, jika terdapat nilai yang jauh dari prediksi, kontribusi terhadap RMSE akan lebih besar. Sedangkan hasil MAE sebesar 0.088 memberikan kontribusi yang setara dari setiap kesalahan tanpa memperhatikan ukuran atau arahnya. Kesalahan kecil dan besar memiliki dampak yang sama pada nilai MAE. Tujuan meminimalkan baik RMSE maupun MAE, Oleh karena itu, semakin kecil kedua nilai tersebut, semakin baik kualitas prediksi model.

Penelitian ini, yang bertujuan untuk melakukan prediksi pada data harga *real estate* dengan metode Linear Regresi, menyarankan peneliti selanjutnya untuk mengeksplorasi penggunaan metode lain, seperti algoritma ANN (*Artificial Neural Network*), yang terbukti memberikan nilai akurasi tinggi dalam beberapa literatur. Disarankan juga untuk membandingkan nilai akurasi tinggi dalam beberapa literatur.

DAFTAR PUSTAKA

- [1] G. Ayu Syafarina *et al.*, "Gita Ayu Syafarin et al, Prediksi Prospek Harga Real Estate 603 Prediksi Prospek Harga Real Estate di Masa Pandemi dengan 3 Atribut Berbasis Algoritma Linear Regression."
- [2] G. Ayu Syafarina *et al.*, "Gita Ayu Syafarin et al, Prediksi Prospek Harga Real Estate 603 Prediksi Prospek Harga Real Estate di Masa Pandemi dengan 3 Atribut Berbasis Algoritma Linear Regression."
- [3] G. Najla, A. #1, and D. Fitriah, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *Jurnal Telematika*, vol. 14, no. 2.
- [4] P. Sikap *et al.*, "Implementasi Data Mining Dalam Menentukan Korelasi Antara Association Rule."
- [5] L. Harianti Hasibuan, S. Musthofa, P. Studi Matematika, and U. Imam Bonjol Padang, "Journal of Science and Technology Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang," 2022.
- [6] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, May 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [7] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, May 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [8] "PERUSAHAAN PROPERTI DAN REAL ESTATE".