Penerapan Metode Regresi Komponen Utama Kernel untuk Prediksi Harga Rumah

Muhammad Arik Verdiansyah*, Suwanda

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*arikverdiansyah@gmail.com, idris1000358@gmail.com

Abstract. Linear regression analysis is a statistical method used to model the relationship between the dependent variable and one or more independent variables expressed in the form of a regression equation. The Ordinary Least Square (OLS) can be used to estimate regression parameters. The regression parameter estimator obtained will be good if the assumptions that apply to MKT are met. One of the assumptions that must be met is that there is a linear relationship between the dependent variable and each independent variable. If linearity is violated, kernels can be used. In addition, another assumption that must be met is the absence of multicollinearity between independent variables. An alternative method that can be used to overcome multicollinearity is using Principal Component Analysis. If the linearity and multicollinearity assumptions are violated, the Kernel Principal Component Regression method can be applied. This study will apply the Kernel Principal Component Regression (KPCR) method, where the principal component is a nonlinear combination of independent variables through a kernel function. This KPCR method is implemented to predict house prices in South Jakarta based on building area, land area, number of bedrooms, number of bathrooms, and number of car capacity in the garage. Based on the results of the study, 3 main components of the kernel were formed with a prediction model $\hat{Y} = 76,3612 + 409,5373 \text{ KUK1} + 328,4967 \text{ KUK2} - 291,7757 \text{ KUK3}$ and a coefficient of determination (R^2) of 70.4640%.

Keywords: *Linearity, Multicollinearity, Kernel Principal Component Regression (KPCR).*

Abstrak. Analisis regresi linear merupakan suatu metode statistika yang digunakan untuk memodelkan hubungan antara variabel dependen dengan satu atau lebih variabel independen yang dinyatakan dalam bentuk persamaan regresi. Dalam menaksir parameter regresi dapat menggunakan Metode Kuadrat Terkecil (MKT). Penaksir parameter regresi yang diperoleh akan baik jika asumsi-asumsi yang berlaku untuk MKT dipenuhi. Salah satu asumsi yang harus dipenuhi adalah terdapat hubungan linear antara variabel dependen dan masing-masing variabel independen. Apabila linearitas terlanggar, dapat menggunakan kernel. Selain itu, asumsi lainnya yang harus terpenuhi adalah tidak adanya multikolinearitas antar variabel independen. Metode alternatif yang dapat digunakan untuk mengatasi multikolinearitas adalah menggunakan Analisis Komponen Utama. Apabila asumsi linearitas dan multikolinearitas terlanggar, dapat diterapkan metode Regresi Komponen Utama Kernel. Dalam penelitian ini, akan diterapkan metode Regresi Komponen Utama Kernel (RKUK), dimana komponen utama merupakan kombinasi nonlinier dari variabel independen melalui fungsi kernel. Metode RKUK ini diimplementasikan untuk memprediksi harga rumah di Jakarta Selatan berdasarkan luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan jumlah kapasitas mobil dalam garasi. Berdasarkan hasil penelitian, terbentuk 3 buah komponen utama kernel dengan model prediksi $\hat{Y} = 76,3612 + 409,5373 \text{ KUK1} + 328,4967 \text{ KUK2} - 291,7757 \text{ KUK3}$ serta nilai koefisien determinasi (R^2) sebesar 70,4640%.

Kata Kunci: *Linearitas, Multikolinearitas, Regresi Komponen Utama Kernel (RKUK).*

A. Pendahuluan

Analisis regresi digunakan untuk mengetahui pengaruh satu atau lebih variabel independen terhadap sebuah variabel dependen. Suatu model regresi dikatakan baik apabila memenuhi asumsi-asumsi ideal, termasuk linearitas. Dalam pelaksanaannya, asumsi linearitas antara masing-masing variabel independen dengan variabel dependen sering terlanggar. Apabila terdapat masalah linearitas pada model regresi, akan menyebabkan model regresi menjadi tidak efektif dalam menjelaskan hubungan antara variabel independen dan dependen. Hal ini dapat menyebabkan hasil uji signifikansi menjadi tidak valid dan membuat kesimpulan yang diambil dari model menjadi salah (Kutner et al., 2004). Metode alternatif yang digunakan untuk mengatasi permasalahan linearitas dalam analisis regresi salah satunya, yaitu menggunakan kernel. Keuntungan menggunakan kernel adalah bisa digunakan untuk memodelkan data yang tidak memiliki pola tertentu (Härdle, 1994).

Selain itu, asumsi multikolinearitas dalam regresi sering terlanggar. Jika terdapat masalah multikolinearitas pada model regresi, maka hasil uji signifikansi menjadi tidak valid dan interpretasi hasil yang salah (Hair et al., 2006). Metode alternatif yang digunakan untuk mengatasi permasalahan multikolinearitas dalam analisis regresi salah satunya, yaitu Analisis Komponen Utama (AKU) (Montgomery & Peck, 1991).

Apabila asumsi linearitas dan multikolinearitas tidak terpenuhi, maka diperlukan sebuah metode yang dapat mengatasi kedua masalah tersebut. Metode alternatif untuk mengatasinya dengan menggunakan regresi komponen utama kernel. Regresi Komponen Utama Kernel (RKUK) merupakan pengembangan dari metode Regresi Komponen Utama. Secara umum, Regresi Komponen Utama Kernel (RKUK) dapat digunakan untuk masalah regresi di mana variabel independen memiliki hubungan nonlinear dengan variabel dependen dan terdapat banyak fitur input (Schölkopf et al., 1999). Kernel Gaussian merupakan salah satu metode kernel yang paling berguna di antara metode kernel lainnya dikarenakan memiliki komputasi yang lebih sederhana dan memiliki kemudahan yang lebih tinggi (Sun & Ma, 2020). Dengan mentransformasi data input ke dalam ruang fitur baru menggunakan fungsi kernel dan kemudian mengurangi dimensi data menggunakan AKU, RKUK dapat membantu meningkatkan akurasi dan efisiensi model regresi (Mika et al., 1999).

Seiring waktu, kebutuhan fisiologis seseorang akan selalu bertambah, termasuk membeli rumah. Menurut UU No. 4 Tahun 1992 tentang Perumahan dan Permukiman, rumah adalah bangunan yang berfungsi sebagai tempat tinggal atau hunian dan sarana pembinaan keluarga. Rumah harus memiliki nilai, ukuran, kondisi, dan yang terpenting sisi finansial yang memungkinkan untuk investasi jangka panjang. Harga setiap rumah tidak selalu pasti dan tidak dapat diprediksi secara akurat. Oleh karena itu, manusia memerlukan suatu sistem yang dapat memperkirakan harga rumah berdasarkan faktor pendukung (Rahayuningtyas et al., 2021).

Penelitian ini bertujuan untuk menentukan model prediksi harga rumah menggunakan metode regresi komponen utama kernel (Fadilah & Kudus, 2023). Analisis regresi digunakan untuk mengetahui pengaruh satu atau lebih variabel independen terhadap sebuah variabel dependen. Suatu model regresi dikatakan baik apabila memenuhi asumsi-asumsi ideal, termasuk linearitas. Dalam pelaksanaannya, asumsi linearitas antara masing-masing variabel independen dengan variabel dependen sering terlanggar. Apabila terdapat masalah linearitas pada model regresi, akan menyebabkan model regresi menjadi tidak efektif dalam menjelaskan hubungan antara variabel independen dan dependen. Hal ini dapat menyebabkan hasil uji signifikansi menjadi tidak valid dan membuat kesimpulan yang diambil dari model menjadi salah (Kutner et al., 2004). Metode alternatif yang digunakan untuk mengatasi permasalahan linearitas dalam analisis regresi salah satunya, yaitu menggunakan kernel. Keuntungan menggunakan kernel adalah bisa digunakan untuk memodelkan data yang tidak memiliki pola tertentu (Härdle, 1994).

Selain itu, asumsi multikolinearitas dalam regresi sering terlanggar. Jika terdapat masalah multikolinearitas pada model regresi, maka hasil uji signifikansi menjadi tidak valid dan interpretasi hasil yang salah (Hair et al., 2006). Metode alternatif yang digunakan untuk mengatasi permasalahan multikolinearitas dalam analisis regresi salah satunya, yaitu Analisis Komponen Utama (AKU) (Montgomery & Peck, 1991).

Apabila asumsi linearitas dan multikolinearitas tidak terpenuhi, maka diperlukan sebuah metode yang dapat mengatasi kedua masalah tersebut. Metode alternatif untuk mengatasinya dengan menggunakan regresi komponen utama kernel. Regresi Komponen Utama Kernel (RKUK) merupakan pengembangan dari metode Regresi Komponen Utama. Secara umum, Regresi Komponen Utama Kernel (RKUK) dapat digunakan untuk masalah regresi di mana variabel independen memiliki hubungan nonlinear dengan variabel dependen dan terdapat banyak fitur input (Schölkopf et al., 1999). Kernel Gaussian merupakan salah satu metode kernel yang paling berguna di antara metode kernel lainnya dikarenakan memiliki komputasi yang lebih sederhana dan memiliki kemudahan yang lebih tinggi (Sun & Ma, 2020). Dengan mentransformasi data input ke dalam ruang fitur baru menggunakan fungsi kernel dan kemudian mengurangi dimensi data menggunakan AKU, RKUK dapat membantu meningkatkan akurasi dan efisiensi model regresi (Mika et al., 1999).

Seiring waktu, kebutuhan fisiologis seseorang akan selalu bertambah, termasuk membeli rumah. Menurut UU No. 4 Tahun 1992 tentang Perumahan dan Permukiman, rumah adalah bangunan yang berfungsi sebagai tempat tinggal atau hunian dan sarana pembinaan keluarga. Rumah harus memiliki nilai, ukuran, kondisi, dan yang terpenting sisi finansial yang memungkinkan untuk investasi jangka panjang. Harga setiap rumah tidak selalu pasti dan tidak dapat diprediksi secara akurat. Oleh karena itu, manusia memerlukan suatu sistem yang dapat memperkirakan harga rumah berdasarkan faktor pendukung (Rahayuningtyas et al., 2021).

Penelitian ini bertujuan untuk menentukan model prediksi harga rumah menggunakan metode regresi komponen utama kernel.

B. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang berisi satu buah variabel dependen (Y), yaitu harga rumah dengan satuan ratusan juta rupiah dan lima buah variabel independen (x), yaitu luas bangunan dengan satuan m^2 (x_1), luas tanah dengan satuan m^2 (x_2), jumlah kamar tidur dengan satuan ruang (x_3), jumlah kamar mandi dengan satuan ruang (x_4), dan jumlah kapasitas mobil dalam garasi dengan satuan mobil (x_5).

Penelitian ini menggunakan metode Regresi Komponen Utama Kernel (RKUK) dengan Langkah-langkah sebagai berikut: (1) Menguji linearitas dan multikolinearitas; (2) Melakukan pembakuan data; (3) Membentuk matriks Kernel Gaussian; (4) Melakukan sentralisasi Kernel Gaussian; (5) Menghitung eigen kernel dan vektor eigen kernel; (6) Menentukan banyaknya komponen utama kernel; (7) Menghitung nilai komponen utama kernel (8) Membentuk model Regresi Komponen Utama Kernel (RKUK); (9) Menghitung nilai R^2 .

C. Hasil Penelitian dan Pembahasan

Linearitas

Berikut ini merupakan pengujian linearitas antara masing-masing variabel independen terhadap variabel dependen.

Tabel 1. Linearitas

Variabel	Signifikansi	Kesimpulan
Y dan x_1	0.000	Nonlinear
Y dan x_2	0.000	Nonlinear
Y dan x_3	0.004	Nonlinear
Y dan x_4	0.000	Nonlinear
Y dan x_5	0.000	Nonlinear

Berdasarkan Tabel 1 masing-masing nilai signifikansinya $< \alpha$, maka dapat disimpulkan bahwa masing-masing variabel independen, yaitu variabel x_1 , x_2 , x_3 , x_4 , dan x_5 memiliki hubungan nonlinear terhadap variabel dependen (Y).

Multikolinearitas

Langkah awal yang perlu dilakukan untuk mendeteksi multikolinearitas adalah membentuk variabel independen pada Tabel 3.1 menjadi sebuah matriks.

$$X = \begin{pmatrix} 220 & 220 & 3 & 3 & 0 \\ 180 & 137 & 4 & 3 & 2 \\ 267 & 250 & 4 & 4 & 4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 139 & 230 & 4 & 4 & 1 \\ 360 & 606 & 7 & 4 & 0 \\ 420 & 430 & 7 & 4 & 2 \end{pmatrix}$$

Setelah terbentuk matriks X , selanjutnya dilakukan perhitungan $X^T X$ dengan hasil sebagai berikut

$$X^T X = \begin{pmatrix} 109157817 & 90225292 & 1429008 & 1142674 & 669993 \\ 90225292 & 89613594 & 1240955 & 965331 & 593519 \\ 1429008 & 1240955 & 24503 & 18516 & 9729 \\ 1142674 & 965331 & 18516 & 15178 & 7747 \\ 66999360 & 593519 & 9729 & 7747 & 6030 \end{pmatrix}$$

Berikut ini merupakan hasil perhitungan yang dilakukan untuk mendapatkan nilai eigen dari matriks $X^T X$.

Tabel 1. Nilai Eigen $X^T X$

No	Nilai Eigen
1	190173459
2	8633106,4
3	8212,32814
4	1649,64043
5	694,251064

Untuk melakukan pengecekan multikolinearitas dapat dilakukan perhitungan menggunakan persamaan berikut.

$$CI = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = \sqrt{\frac{190173459}{694,251064}} = 523,3795$$

Diperoleh nilai CI sebesar 523,3795 dimana nilai tersebut > 30 , maka dapat disimpulkan bahwa terdapat multikolinearitas yang tinggi dalam data.

Pembakuan Data

Seluruh variabel independen, yaitu variabel luas bangunan (x_1), luas tanah (x_2), jumlah kamar tidur (x_3), jumlah kamar mandi (x_4), dan jumlah kapasitas mobil dalam garasi (x_5) akan dilakukan pembakuan atau pembakuan dengan menggunakan persamaan berikut.

$$\text{nilai baru} = \frac{\text{nilai lama} - \text{mean}}{\text{simpangan baku}}$$

Berikut ini merupakan hasil perhitungan pembakuan data.

Tabel 3. Hasil Pembakuan

No	x_1	x_2	x_3	x_4	x_5
1	-0,3179	-0,0969	-1,0607	-0,4281	-1,2712
2	-0,5428	-0,5581	-0,4249	-0,4281	0,0524
3	-0,0536	0,0698	-0,4249	0,2761	1,3760
4	0,6941	0,6533	0,8467	0,9803	0,7142
5	0,1319	-0,4636	0,2109	-0,4281	0,7142
⋮	⋮	⋮	⋮	⋮	⋮
1007	-0,7733	-0,0413	-0,4249	0,2761	-0,6094
1008	0,4692	2,0481	1,4825	0,2761	-1,2712
1009	0,8066	1,0701	1,4825	0,2761	0,0524

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Matriks Kernel

Fungsi kernel yang digunakan dalam penelitian ini merupakan Kernel Gaussian yang dapat dibentuk melalui persamaan berikut.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{c}\right)$$

Dalam menentukan nilai parameter c dilakukan melalui proses *trial and error* dengan nilai 1 hingga 1000. Nilai parameter yang terpilih dalam penelitian ini adalah dengan menggunakan nilai $c = 309$, matriks kernel yang terbentuk akan berukuran 1009×1009 yang tersaji berikut.

$$K = \begin{pmatrix} 1 & 0,9922 & 0,9744 & \dots & 0,9950 & 0,9613 & 0,9640 \\ 0,9922 & 1 & 0,9907 & \dots & 0,9959 & 0,9566 & 0,9725 \\ 0,9744 & 0,9907 & 1 & \dots & 0,9856 & 0,9531 & 0,9772 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0,9950 & 0,9959 & 0,9856 & \dots & 1 & 0,9682 & 0,9750 \\ 0,9613 & 0,9566 & 0,9531 & \dots & 0,9682 & 1 & 0,9909 \\ 0,9640 & 0,9725 & 0,9772 & \dots & 0,9750 & 0,9909 & 1 \end{pmatrix}$$

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Sentralisasi Matriks Kernel

Setelah didapat matriks kernel, dilakukan sentralisasi matriks kernel K menjadi matriks K_C dimana persamaan matriks K_C sebagai berikut.

$$K_C = K - UK - KU + UKU$$

dengan U dan K berukuran $n \times n$ dimana

$$U = \frac{1}{n} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

Berikut ini merupakan hasil perhitungan matriks U dengan ukuran 1009×1009 .

$$U = \begin{pmatrix} 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \\ 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \\ 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \\ 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \\ 0,00099 & 0,00099 & 0,00099 & \dots & 0,00099 & 0,00099 & 0,00099 \end{pmatrix}$$

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.
Berikut ini hasil perhitungan matriks K_C dengan ukuran 1009×1009 .

$$K_C = \begin{pmatrix} 0,0190 & 0,0048 & -0,0089 & \dots & 0,0085 & -0,0028 & -0,0133 \\ 0,0048 & 0,0062 & 0,0010 & \dots & 0,0030 & -0,0139 & -0,0111 \\ -0,0089 & 0,0010 & 0,0142 & \dots & -0,0033 & -0,0134 & -0,0025 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0,0085 & 0,0030 & -0,0033 & \dots & 0,0079 & -0,0014 & -0,0078 \\ -0,0028 & -0,0139 & -0,0134 & \dots & -0,0014 & 0,0528 & 0,0305 \\ -0,0133 & -0,0111 & -0,0025 & \dots & -0,0078 & 0,0305 & 0,0264 \end{pmatrix}$$

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Nilai Eigen dan Vektor Eigen

Langkah selanjutnya adalah menentukan nilai eigen dari matriks K_C, Hasil nilai eigen dapat dilihat pada Tabel 4 berikut.

Tabel 4. Nilai Eigen

No	Nilai Eigen
1	0,017317513
2	0,005491515
3	0,003483344
4	0,002228103
5	0,001342585
⋮	⋮
1007	4,99E-19
1008	3,07E-19
1009	9,56E-20

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Setelah didapat nilai eigen, dilakukan perhitungan untuk mendapatkan vektor eigen dari matriks K_C. Berikut ini merupakan hasil perhitungan vektor eigen.

$$V = \begin{pmatrix} -0,0060 & 0,0018 & -0,0212 & \dots & 1619672,5765 & 1889832,8962 & -1040974,1953 \\ -0,0041 & 0,0021 & 0,0090 & \dots & -2984746,3144 & 923401,0164 & -2783732,8896 \\ 0,0021 & 0,0107 & 0,0253 & \dots & 1552780,5019 & 2327192,3604 & -193940,7493 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -0,0032 & -0,0055 & -0,0032 & \dots & 326426,6215 & -443937,8041 & 1004981,1987 \\ 0,0067 & -0,0105 & -0,0455 & \dots & 65656,3105 & 193385,7255 & -352014,5614 \\ 0,0076 & -0,0058 & -0,0145 & \dots & 25369,4573 & 321204,9800 & -97768,2096 \end{pmatrix}$$

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Jumlah Komponen Utama Kernel

Pengambilan banyaknya komponen utama kernel dalam penelitian ini dilakukan berdasarkan proporsi varians kumulatif. Berikut ini merupakan hasil perhitungan proporsi varians kumulatif.

Tabel 5. Proporsi Varians Kumulatif

Banyak Komponen	Proporsi Varians Kumulatif
1	0,55878
2	0,73597
3	0,84837
4	0,92026
5	0,96358
⋮	⋮
1007	1
1008	1
1009	1

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Berdasarkan Tabel 5 dengan mengambil 3 buah komponen utama kernel sudah mencakup sekitar 84,837% dari keragaman data, dimana nilai tersebut sudah > 80%, maka untuk langkah selanjutnya akan digunakan 3 buah komponen utama kernel untuk melihat nilai komponen utama kernel yang selanjutnya akan dibentuk model regresinya.

Nilai Komponen Utama Kernel

Berikut ini merupakan persamaan untuk mendapatkan nilai komponen utama kernel.

$$\mathbf{B} = \mathbf{K}_C \cdot \mathbf{V}$$

Dengan menggunakan nilai vektor eigen yang sudah diperoleh, maka nilai komponen utama kernel yang terbentuk adalah sebagai berikut.

Tabel 6. Nilai Komponen Utama Kernel

No	KUK1	KUK2	KUK3
1	-0.1043	0.0099	-0.0746
2	-0.0709	0.0116	0.0315
3	0.0364	0.0593	0.0890
4	0.1368	-0.0207	0.0189
5	0.0005	0.0220	0.0528
⋮	⋮	⋮	⋮
1007	-0.0557	-0.0306	-0.0113
1008	0.1170	-0.0583	-0.1600
1009	0.1336	-0.0321	-0.0511

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Model Regresi Komponen Utama Kernel

Selanjutnya dilakukan perhitungan untuk mendapatkan model regresi menggunakan persamaan berikut.

$$\hat{w} = (B^T B)^{-1} B^T Y$$

Dengan matriks B sebagai berikut.

$$B = \begin{pmatrix} 1 & -0,1043 & 0,0099 & -0,0746 \\ 1 & -0,0709 & 0,0116 & 0,0315 \\ 1 & 0,0364 & 0,0593 & 0,0890 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -0,0557 & -0,0306 & -0,0113 \\ 1 & 0,1170 & -0,0583 & -0,1600 \\ 1 & 0,1336 & -0,0321 & -0,0511 \end{pmatrix}$$

Hasil selengkapnya terdapat pada <https://bit.ly/HasilAnalisisSkripsiArik>.

Berikut ini merupakan hasil perhitungan $\hat{w}$.

$$\hat{w} = \begin{pmatrix} 76,3612 \\ 409,5373 \\ 328,4967 \\ -291,7757 \end{pmatrix}$$

Setelah diperoleh nilai $\hat{w}$, langkah selanjutnya adalah membentuk model regresi komponen utama kernel. Berikut ini merupakan model prediksi regresi komponen utama kernel yang terbentuk:

$$\hat{Y} = 76,3612 + 409,5373 \text{ KUK1} + 328,4967 \text{ KUK2} - 291,7757 \text{ KUK3}$$

Nilai R²

Selanjutnya, akan dihitung nilai koefisien determinasi (R²). Berikut ini merupakan nilai perhitungan R².

$$R^2 = 0,704640$$

Nilai R² yang diperoleh adalah sebesar 0,704640 atau 70,4640%. Hal tersebut memiliki arti bahwa kemampuan nilai komponen utama kernel dalam penelitian ini mempengaruhi variabel dependen atau variabel harga rumah adalah sebesar 70,4640%, sedangkan sisanya (100% - 70,4640% = 29,536%) dijelaskan oleh variabel lain di luar penelitian.

D. Kesimpulan

Berdasarkan tahapan analisis pada penelitian ini dapat disimpulkan bahwa dengan metode Regresi Komponen Utama Kernel (RKUK) menggunakan data harga rumah di Jakarta Selatan berdasarkan luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan jumlah kapasitas mobil dalam garasi terbentuk 3 buah komponen utama kernel. Model prediksi yang diperoleh untuk memprediksi harga rumah adalah sebagai berikut $\hat{Y} = 76,3612 + 409,5373 \text{ KUK1} + 328,4967 \text{ KUK2} - 291,7757 \text{ KUK3}$ serta memperoleh nilai koefisien determinasi (R²) sebesar 70,4640%.

Acknowledge

Penulis ingin mengucapkan terima kasih kepada Allah SWT, kedua orang tua yang selalu memberikan doa serta dukungan dan Bapak Dr. Suwanda, M.S selaku dosen pembimbing atas arahan dan bimbingannya sehingga terselesaikannya penelitian ini. Serta kepada para dosen Statistika Unisba yang telah memberikan ilmu pengetahuannya, dan kepada teman-teman yang selalu membantu dan menemani penulis hingga terselesaikannya perkuliahan ini.

Daftar Pustaka

- [1] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis (Vol. 6)*.
- [2] Härdle, W. (1994). *Applied Nonparametric Regression* Berlin. Cambridge University Press.
- [3] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Wasserman, W. (2004). *Applied linear regression models (Vol. 4)*. McGraw-Hill/Irwin New York.
- [4] Mika, M., Hirano, & Tsumura. (1999). *Kernel Principal Component Regression in Welding Process Modeling*.
- [5] Montgomery, D. C., & Peck, E. A. (1991). *Introduction Linear Regression Analysis (Second Edition)*. John Wiley & Sons.
- [6] Rahayuningtyas, E. F., Rahayu, F. N., & Azhar, Y. (2021). Prediksi Harga Rumah Menggunakan General Regression Neural Network. *Jurnal Informatika*, 8(1), 59–66. <https://doi.org/10.31294/ji.v8i1.9036>.
- [7] Schölkopf, B., Burges, C. J. C., & Smola, A. J. (1999). *Advances in kernel methods: support vector learning*. MIT press.
- [8] Sun, C., & Ma, H. (2020). A Study of Complex Industrial Systems Based on Revised Kernel Principal Component Regression Method. *IFAC-PapersOnLine*, 53(2), 126–131. <https://doi.org/10.1016/j.ifacol.2020.12.108>.
- [9] Fadilah, T. S., & Kudus, A. (2023). Penerapan Metode Regresi Kernel Smoothing untuk Imputasi Data Lama Waktu Terinfeksi Covid-19. *Jurnal Riset Statistika*, 51–60. <https://doi.org/10.29313/jrs.v3i1.1802>