



K01 FI2151 Pemodelan, Generator Data, dan Analisisnya
Kelompok 3

Prediksi Harga Rumah di Jakarta Selatan

RESEARCH BASED LEARNING (RBL)

Anggota Kelompok 3

10223037 - Dietrich Pepalem Tarigan

10223064 - Muhammad Rasyid Ghifari

10223067 - Gwen Sevilen

10223075 - Muhammad Syamsuddiin

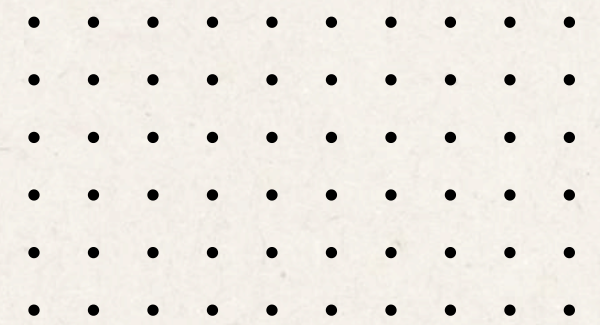
LATAR BELAKANG

Alasan memilih topik?

Harga rumah dipengaruhi banyak variabel seperti lokasi, luas bangunan, jumlah kamar, fasilitas umum di sekitar, dan kondisi pasar ekonomi. Dengan banyaknya variabel tersebut, proses analisis manual menjadi tidak efisien dan rawan kesalahan. Oleh karena itu, diperlukan pendekatan berbasis regresi multivariabel untuk memahami pola dan hubungan antarvariabel guna memberikan prediksi harga rumah yang akurat.



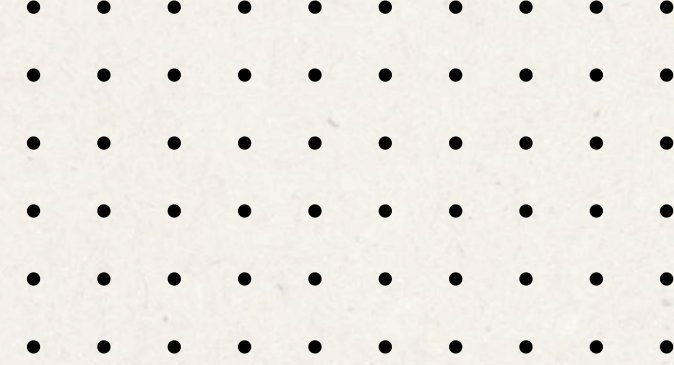
LATAR BELAKANG (2)



Tujuan

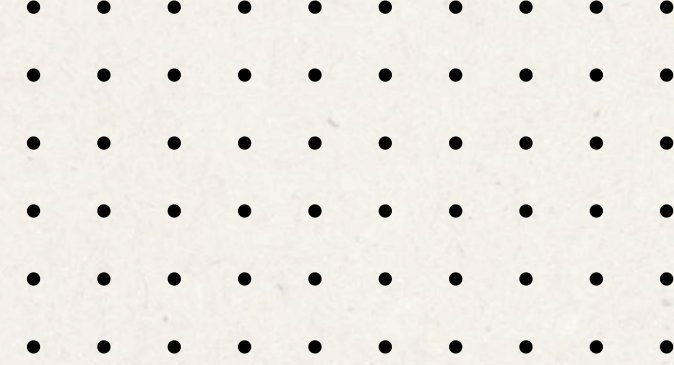
- Menganalisis hubungan luas tanah (LT) terhadap harga rumah di Jakarta Selatan
- Menganalisis hubungan luas bangunan (LB) terhadap harga rumah di Jakarta Selatan
- Menganalisis hubungan jumlah kamar mandi terhadap harga rumah di Jakarta Selatan.
- Menganalisis hubungan jumlah kamar tidur terhadap harga rumah di Jakarta Selatan.
- Memprediksi perkiraan harga rumah di Jakarta Selatan berdasarkan data terdahulu

METODE



- Metode ***Data Preprocessing***, yaitu dengan melakukan pengumpulan data, pembersihan data dari outliers (data cleaning), dan transformasi data.
- Analisis data menggunakan metode ***Gradient Descent***, yaitu teknik analisis yang bertujuan untuk meminimalkan fungsi harga rumah

METODE (2)



Data harga rumah yang digunakan dalam analisis ini diperoleh dari situs Kaggle. Pada tahap awal, data mentah tersebut dianalisis menggunakan metode statistik deskriptif untuk memperoleh karakteristik utama data, seperti nilai rata-rata, median, rentang, dan distribusi.

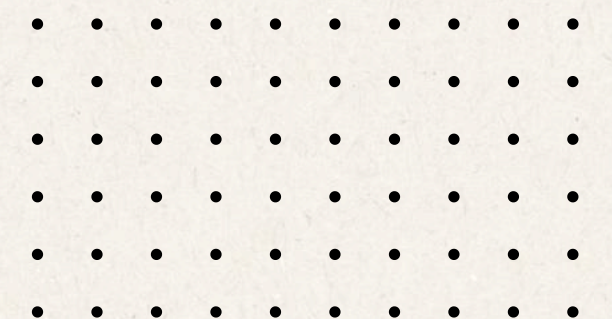
Hasil analisis menunjukkan bahwa terdapat banyak *outlier*, yaitu data yang menyimpang jauh dari pola umum. Kehadiran outlier ini dapat mengaburkan hubungan antarvariabel dalam dataset.

Oleh karena itu, dilakukan **pembersihan data** untuk menghilangkan atau menangani *outlier* sehingga data menjadi lebih konsisten dan dapat diinterpretasikan dengan lebih jelas.

METODE

Setelah proses pembersihan selesai, hubungan antarvariabel mulai terlihat dengan lebih jelas. Analisis korelasi kemudian dilakukan menggunakan metode ***gradient descent*** untuk mengidentifikasi pola atau hubungan antara variabel-variabel dalam data.

Berikut merupakan data yang diperoleh dari hasil analisis statistik deskriptif.



PENGOLAHAN DATA

Data Mula-Mula (yang belum dibersihkan)

HARGA	LT	LB	JKT	JKM	GRS	KOTA
28,000,000,000	1100	700	5	6	ADA	JAKSEL
19,000,000,000	824	800	4	4	ADA	JAKSEL
4,700,000,000	500	400	4	3	ADA	JAKSEL
4,900,000,000	251	300	5	4	ADA	JAKSEL
28,000,000,000	1340	575	4	5	ADA	JAKSEL
10,000,000,000	460	300	4	4	ADA	JAKSEL
7,600,000,000	278	350	4	4	ADA	JAKSEL
5,250,000,000	511	300	3	2	ADA	JAKSEL
670,000,000	70	69	3	2	TIDAK ADA	JAKSEL
480,000,000	66	42	2	1	TIDAK ADA	JAKSEL
10,700,000,000	449	500	6	7	ADA	JAKSEL
4.200.000.000	109	188	2	3	ADA	JAKSEL

.....
.....

Keterangan:
LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (2)

Statistik Data yang Sebelum Dibersihkan

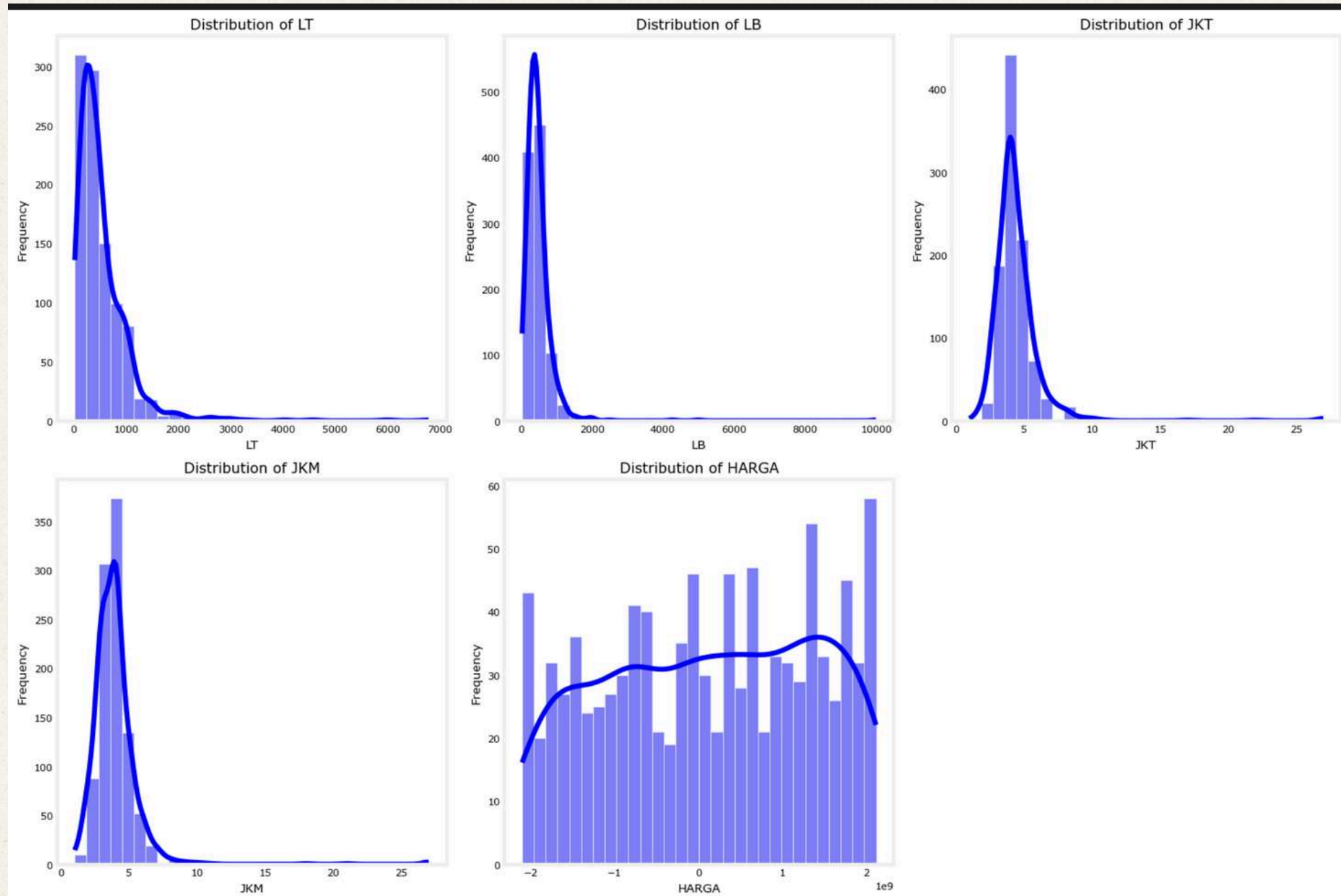
... Descriptive Statistics:						
	LT	LB	JKT	JKM	HARGA	
count	1001.000000	1001.000000	1001.000000	1001.000000	1.001000e+03	
mean	530.504496	487.275724	4.457542	3.940060	1.318044e+08	
std	531.069773	452.872262	2.004606	1.903261	1.257268e+09	
min	22.000000	38.000000	1.000000	1.000000	-2.104967e+09	
25%	220.000000	300.000000	4.000000	3.000000	-8.849019e+08	
50%	400.000000	411.000000	4.000000	4.000000	2.050327e+08	
75%	677.000000	600.000000	5.000000	4.000000	1.230196e+09	
max	6790.000000	10000.000000	27.000000	27.000000	2.120131e+09	

.....
.....
.....

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (3)



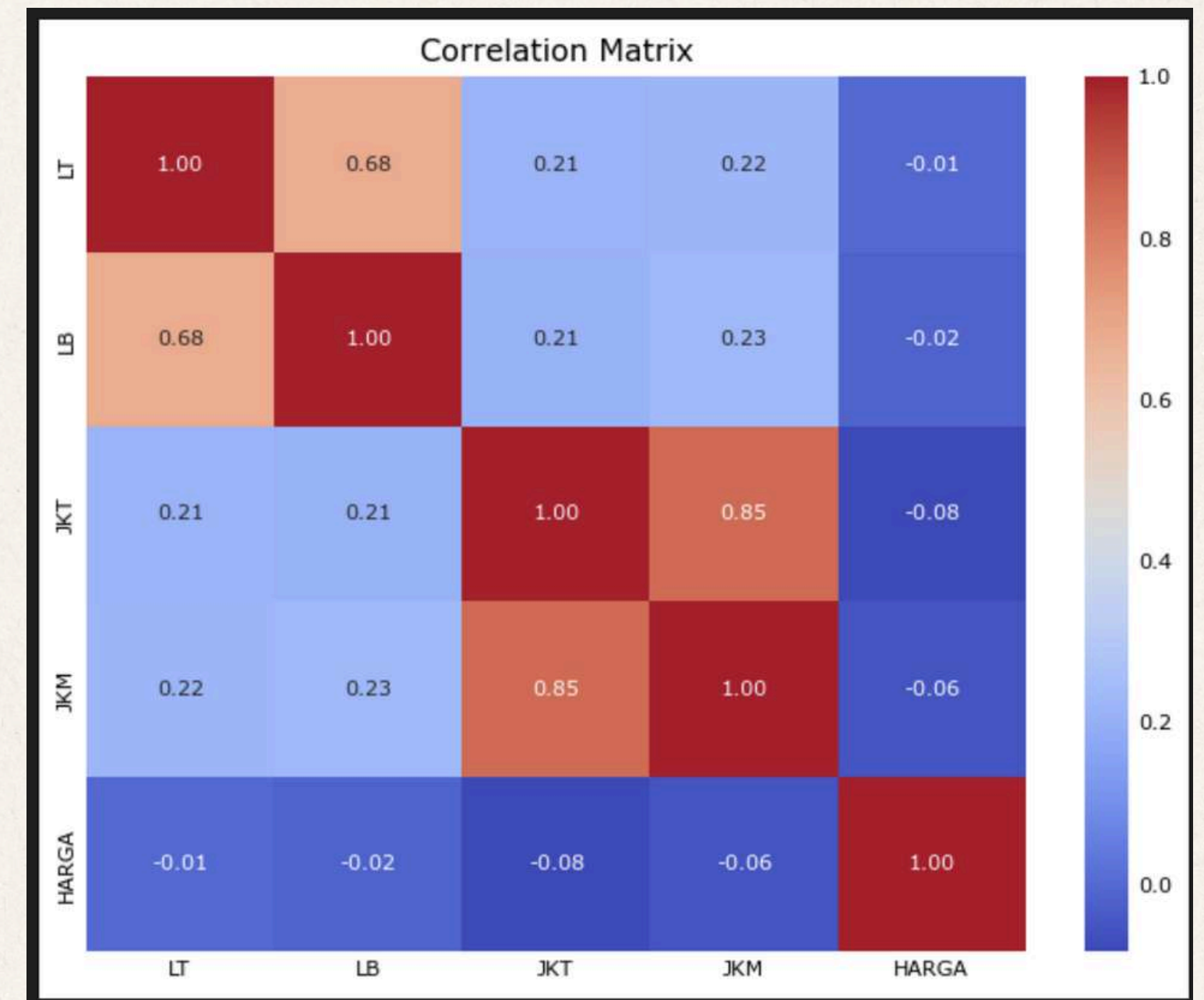
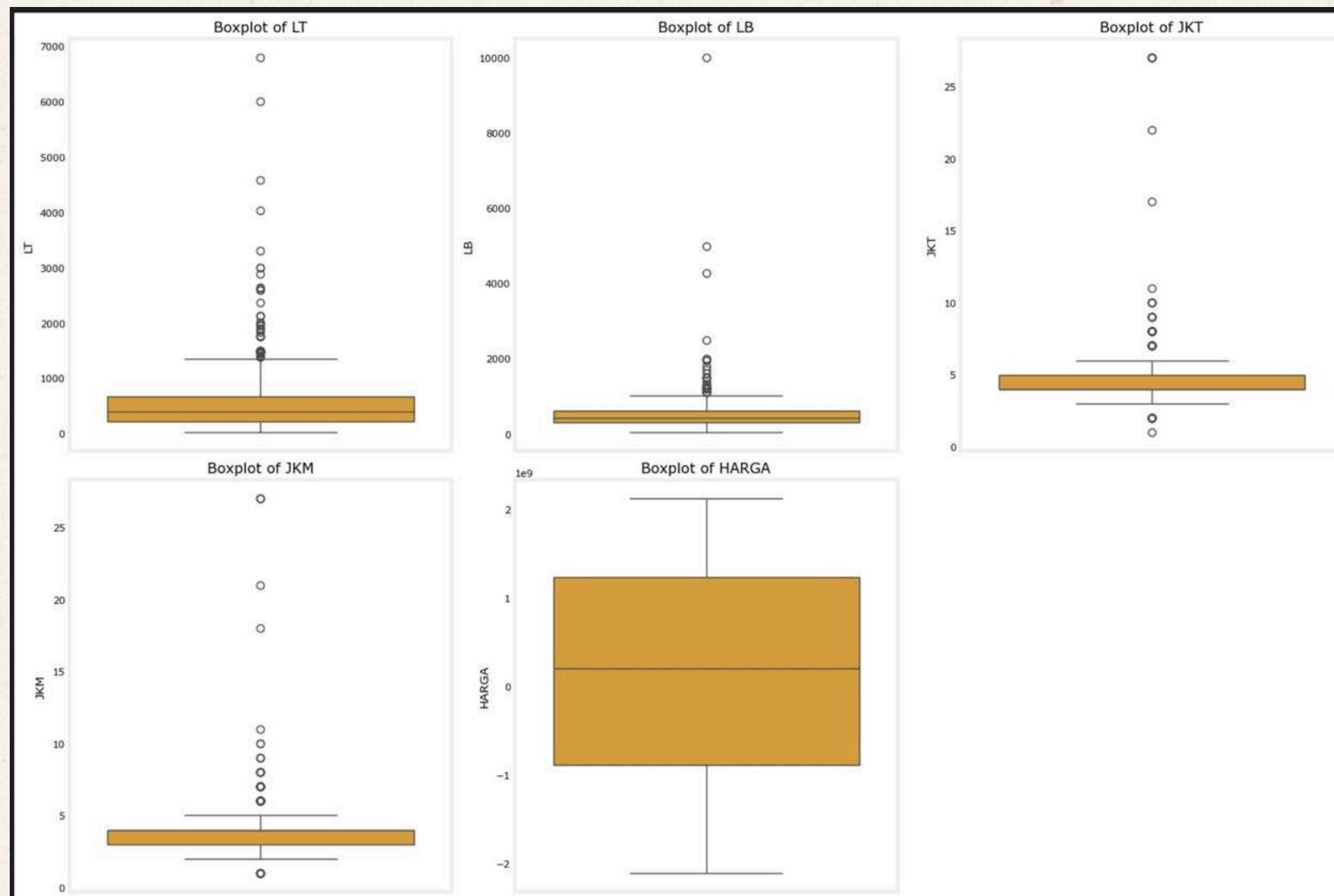
Grafik Distribusi Variabel

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan;
JKT = Jumlah Kamar Tidur; JKM =
Jumlah Kamar Mandi

PENGOLAHAN DATA (4)

Boxplot dan Matriks Korelasi

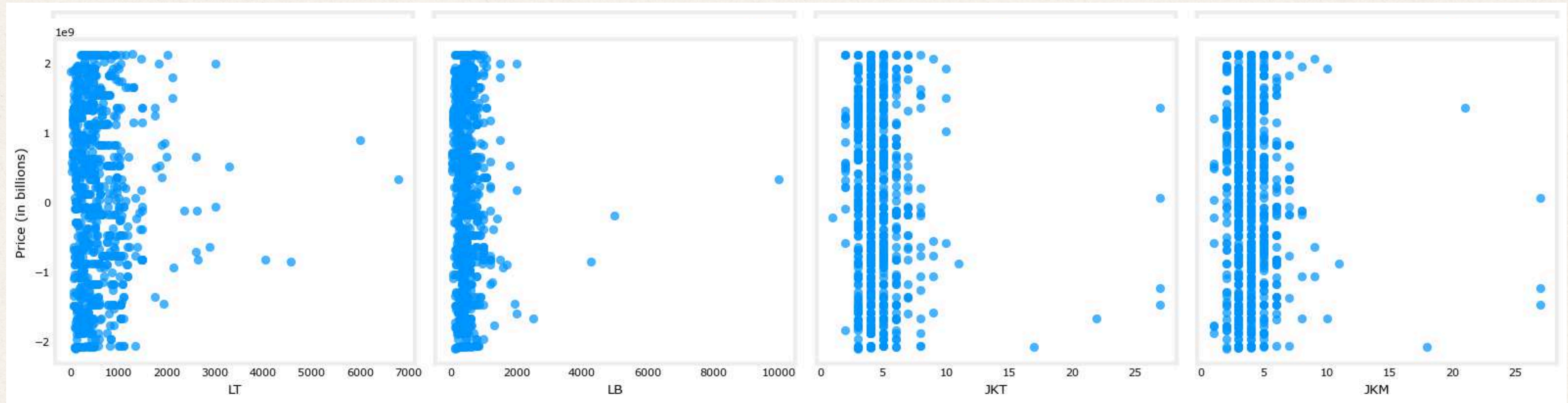


Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (5)

Scatter Plot Harga Rumah terhadap LT, LB, JKT, dan JKM



• • • • •
• • • • •
• • • • •
• • • • •
• • • • •
• • • • •

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (5)

Hasil plot fitur terhadap harga tidak menunjukkan korelasi yang kuat. Maka dari itu kami mencoba untuk menormalisasi dan mentransformasi logaritmik data

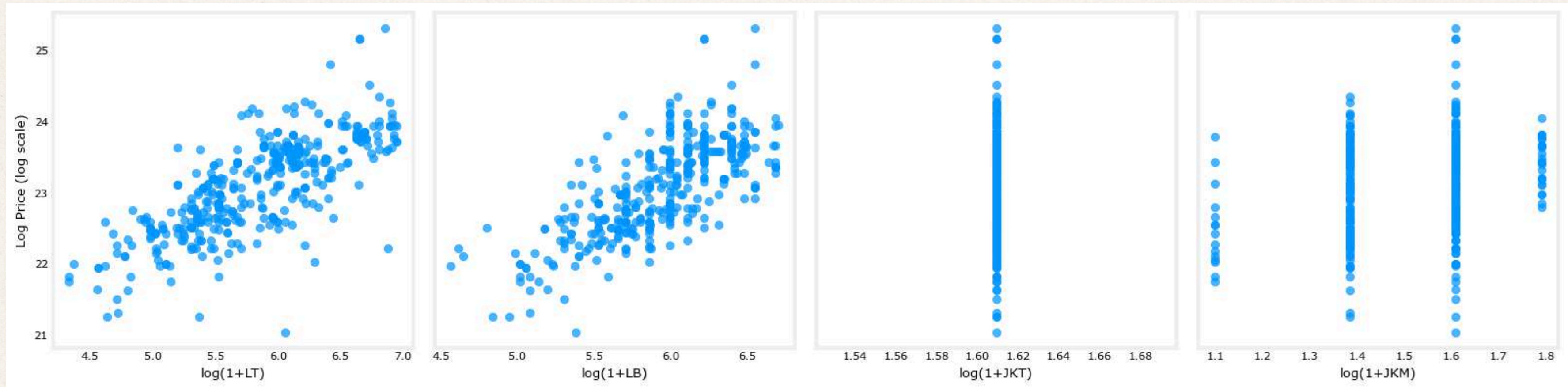
• • • • • • • • • •
• • • • • • • • • •
• • • • • • • • • •
• • • • • • • • • •
• • • • • • • • • •
• • • • • • • • • •

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (5)

Scatter Plot Harga Rumah terhadap LT, LB, JKT, dan JKM secara Logaritmik



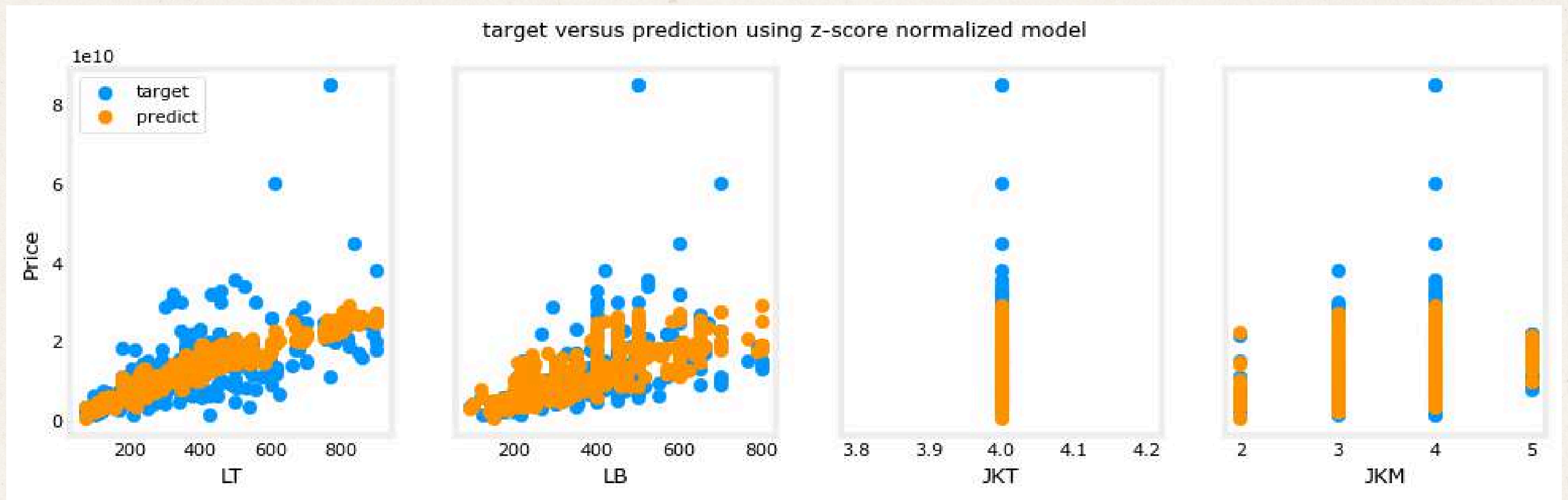
• • • • •
• • • • •
• • • • •
• • • • •
• • • • •
• • • • •

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PENGOLAHAN DATA (6)

Grafik Prediksi vs Target



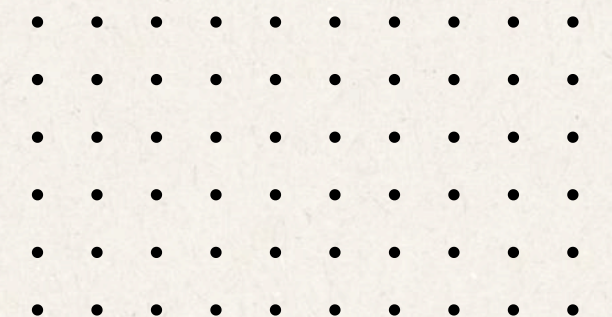
.....
.....
.....

Keterangan:

LT = Luas Tanah; LB = Luas Bangunan; JKT = Jumlah Kamar Tidur; JKM = Jumlah Kamar Mandi

PEMBAHASAN

- Model regresi secara efektif memprediksi harga rumah dengan akurasi yang cukup tepat ($R^2 = 0.72$)
- *Skewness* yang tinggi pada distribusi luas tanah (LT) dan luas bangunan (LB) memerlukan *preprocessing* yang signifikan
- Harga rumah yang bernilai negatif juga ditandai sebagai data tidak wajar
- Riak data jumlah kamar tidur dan jumlah kamar mandi mengurangi kemampuan prediksi

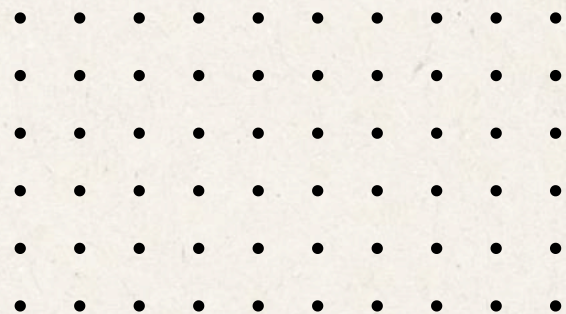


SIMPULAN

- Harga rumah di Jakarta Selatan berbanding lurus secara logaritmik dengan luas tanah
- Harga rumah di Jakarta Selatan berbanding lurus secara logaritmik dengan luas bangunan
- Harga rumah di Jakarta Selatan tidak berkorelasi dengan jumlah kamar mandi
- Harga rumah di Jakarta Selatan tidak berkorelasi dengan jumlah kamar tidur
- Model mampu memprediksi harga rumah di Jakarta Selatan dengan cukup tepat dengan nilai $R^2 = 0.72$

REFERENSI

1. [HTTPS://GITHUB.COM/GREYHATGUY007/MACHINE-LEARNING-SPECIALIZATION-COURSERA](https://github.com/Greyhatguy007/machine-learning-specialization-coursera)
2. [HTTPS://WWW.KAGGLE.COM/DATASETS/WISNUANGGARA/DAFTAR-HARGA-RUMAH](https://www.kaggle.com/datasets/wisnuanggara/daftar-harga-rumah)
3. BISHOP, C. M. (2006). PATTERN RECOGNITION AND MACHINE LEARNING. SPRINGER.



Terima Kasih

KELOMPOK 3

K01 FI2151