

REGRESSION MODEL

Dataset RegresiUTSTelkom

Ketut Satria Wibisana
1103213148



REGRESSION MODEL

Regression model dalam machine learning adalah proses di mana komputer dilatih untuk memprediksi nilai numerik berdasarkan data yang ada. Tujuannya adalah agar model dapat secara otomatis mengidentifikasi hubungan atau pola antara variabel input (fitur) dan output (target) sehingga mampu menghasilkan prediksi nilai kontinu. Contohnya termasuk memprediksi harga rumah berdasarkan ukuran dan lokasi, atau memperkirakan suhu berdasarkan data cuaca sebelumnya. Model regresi digunakan untuk masalah di mana hasilnya berupa angka, bukan kategori.



Tentang Dataset RegresiUTSTelkom

Dataset ini berisi data di mana setiap baris merepresentasikan satu observasi atau pengukuran. Dataset ini memiliki total 17 variabel, yang mencakup berbagai nilai numerik dengan berbagai skala (positif dan negatif). Berikut detail karakteristik dataset:

1. Struktur Dataset:

- Dataset terdiri dari 9 baris (observasi) dan 17 kolom (variabel).
- Kolom pertama adalah tahun, sementara kolom lainnya merupakan variabel numerik yang menggambarkan berbagai pengukuran atau indikator.

2. Jenis Data:

- Semua nilai dalam dataset berupa angka dengan presisi tinggi (desimal).
- Variabel ini mungkin mewakili pengukuran atau nilai dalam berbagai domain (misalnya, ekonomi, lingkungan, atau fisika), tetapi konteks spesifik belum diberikan.

3. Cakupan Nilai:

- Nilai variabel mencakup angka positif dan negatif, menunjukkan dataset ini berisi berbagai jenis pengukuran yang mungkin menggambarkan perubahan atau dinamika tertentu.
- Contohnya: Variabel di kolom ke-5 hingga ke-12 menunjukkan nilai negatif, yang bisa menggambarkan penurunan atau kehilangan dalam suatu fenomena.

4. Format Dataset:

- Data disimpan dalam format matriks 2D di mana baris mewakili observasi, dan kolom mewakili variabel.

5. Potensi Analisis:

- Dataset ini cocok untuk analisis regresi atau eksplorasi statistik untuk memahami hubungan antar variabel.
- Bisa digunakan untuk memprediksi salah satu variabel berdasarkan nilai variabel lainnya, atau menganalisis tren berdasarkan pengamatan.

Polynomial

Polynomial Regression adalah model regresi yang memperluas hubungan linier antara variabel independen (X) dan variabel dependen (Y) dengan menambahkan pangkat lebih tinggi dari X. Model ini digunakan ketika hubungan antara variabel tidak linier, namun masih dapat diwakili dengan persamaan polinomial.

Decision Tree

Decision Tree adalah model pembelajaran mesin yang digunakan untuk klasifikasi dan regresi dengan membagi dataset menjadi subset berdasarkan fitur yang paling relevan. Proses ini membentuk struktur pohon, di mana setiap cabang mewakili keputusan berdasarkan fitur, dan setiap daun mewakili hasil prediksi.

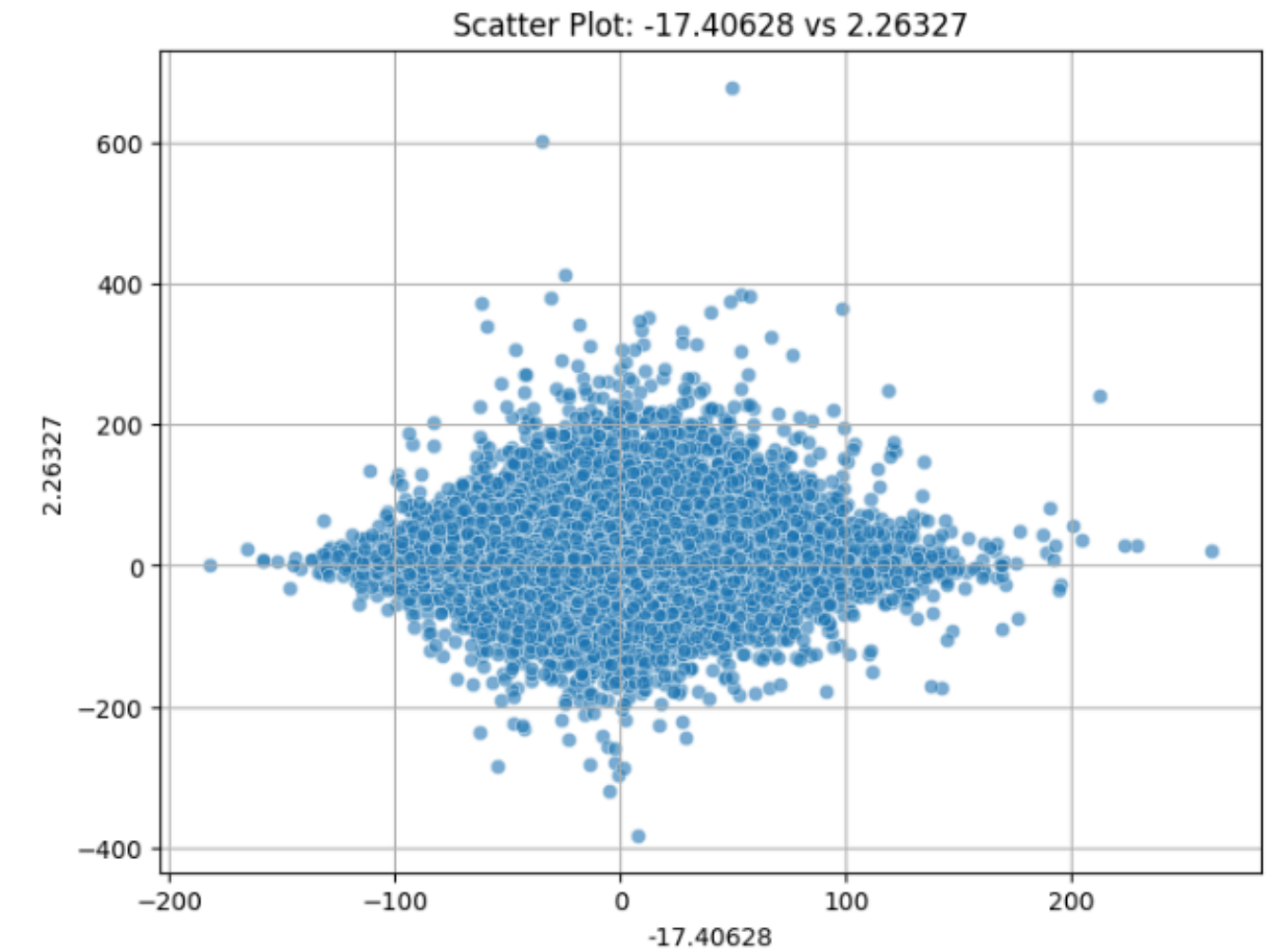
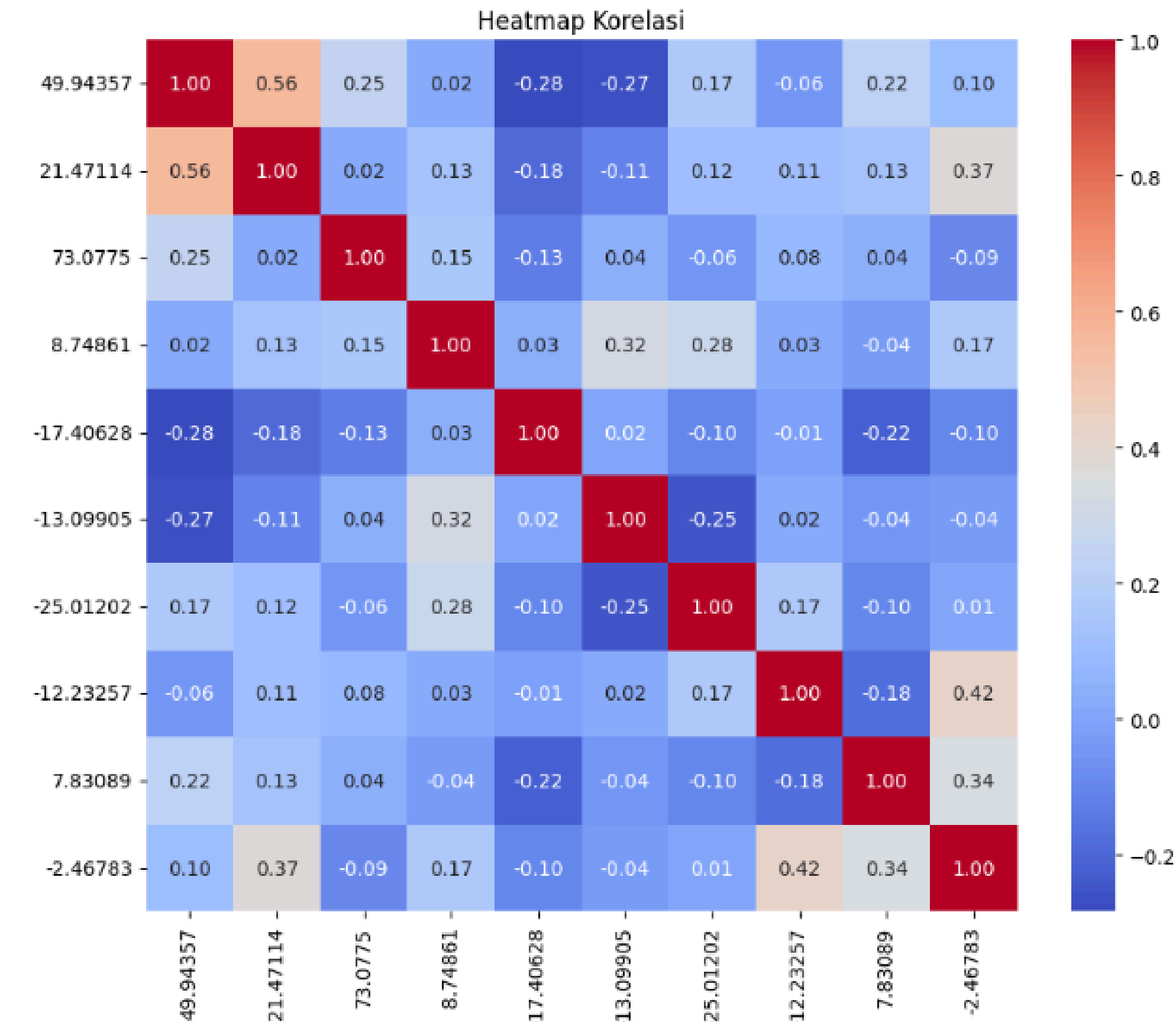
k-NN

k-NN adalah algoritma pembelajaran berbasis instance yang digunakan untuk klasifikasi dan regresi. Model ini bekerja dengan mengklasifikasikan data baru berdasarkan kedekatannya dengan titik data terdekat dalam dataset pelatihan.

XGBoost Regression

XGBoost (Extreme Gradient Boosting) adalah algoritma pembelajaran mesin berbasis pohon keputusan yang digunakan untuk tugas regresi dan klasifikasi. Model ini merupakan salah satu algoritma ensemble yang paling populer, yang menggabungkan beberapa pohon keputusan dalam satu model untuk meningkatkan akurasi prediksi.

Hasil Dan Kesimpulan



	Model	Best Params \
0	Polynomial Regression	{'poly_features__degree': 1, 'regressor__max_d...
1	Decision Tree	{'regressor__max_depth': 3, 'regressor__min_sa...
2	k-NN	{'regressor__n_neighbors': 10, 'regressor__wei...
3	XGBoost	{'regressor__learning_rate': 0.1, 'regressor__...

	Best R2 Score
0	0.004149
1	0.004149
2	-0.096378
3	0.004574

Thank You