

CLASSIFICATION MODEL

Dataset adult.data

Ketut Satria Wibisana
1103213148



CLASSIFICATION MODEL

Classification Model dalam machine learning adalah proses di mana komputer dilatih untuk mengklasifikasikan data ke dalam kategori atau label tertentu berdasarkan fitur yang ada. Tujuannya adalah agar model dapat secara otomatis mengidentifikasi pola atau hubungan antara variabel input (fitur) dan output (kelas/label) sehingga mampu mengategorikan data baru ke dalam kelas yang tepat. Contohnya termasuk mengklasifikasikan email sebagai spam atau tidak spam, atau memprediksi apakah seseorang akan membeli produk berdasarkan data demografis. Model klasifikasi digunakan untuk masalah di mana hasilnya berupa kategori, bukan angka kontinu.



Tentang Dataset RegresiUTSTelkom

Dataset ini berisi data yang masing-masing barisnya mewakili satu observasi atau pengukuran dengan total 9 baris (observasi) dan 17 kolom (variabel). Setiap kolom menggambarkan nilai numerik yang mungkin mewakili berbagai indikator atau pengukuran dalam domain yang berbeda, seperti ekonomi, lingkungan, atau sains, meskipun konteks spesifik dari data ini tidak disebutkan.

Struktur Dataset:

- Jumlah Baris: 9 observasi atau pengukuran.
- Jumlah Kolom: 17 variabel.
- Kolom Pertama: Tahun, yang mungkin menggambarkan waktu pengambilan data.
- Kolom Lainnya: Merupakan variabel numerik yang menggambarkan berbagai pengukuran atau indikator. Nilai-nilai ini adalah angka desimal dengan presisi tinggi.

Jenis Data:

- Semua nilai dalam dataset berupa angka desimal dengan presisi tinggi.
- Dataset ini kemungkinan menggambarkan pengukuran dalam berbagai domain yang beragam (misalnya, ekonomi, lingkungan, atau fisika), tetapi konteks spesifik dari data ini tidak dijelaskan lebih lanjut.

Cakupan Nilai:

- Nilai dalam dataset ini mencakup angka positif dan negatif.
- Nilai negatif pada beberapa variabel menunjukkan bahwa data ini bisa melibatkan fenomena dengan perubahan atau fluktuasi, seperti penurunan atau kehilangan (misalnya, penurunan laba atau penurunan kualitas suatu sumber daya).

Contoh:

- Variabel pada kolom ke-5 hingga ke-12 memiliki nilai negatif, yang mungkin menggambarkan kerugian atau penurunan dalam beberapa fenomena yang diukur, seperti penurunan pendapatan, penurunan kualitas udara, atau kehilangan produktivitas.

Format Dataset:

- Dataset ini disusun dalam format matriks 2D, dengan baris yang mewakili observasi atau pengukuran, dan kolom yang mewakili variabel atau fitur yang diukur.

Potensi Analisis:

- Dataset ini cocok untuk berbagai jenis analisis statistik, terutama analisis regresi untuk mengeksplorasi hubungan antara variabel yang berbeda.
- Dataset ini dapat digunakan untuk memprediksi salah satu variabel berdasarkan nilai variabel lainnya, atau untuk menganalisis tren yang muncul dalam data selama periode waktu tertentu.

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.

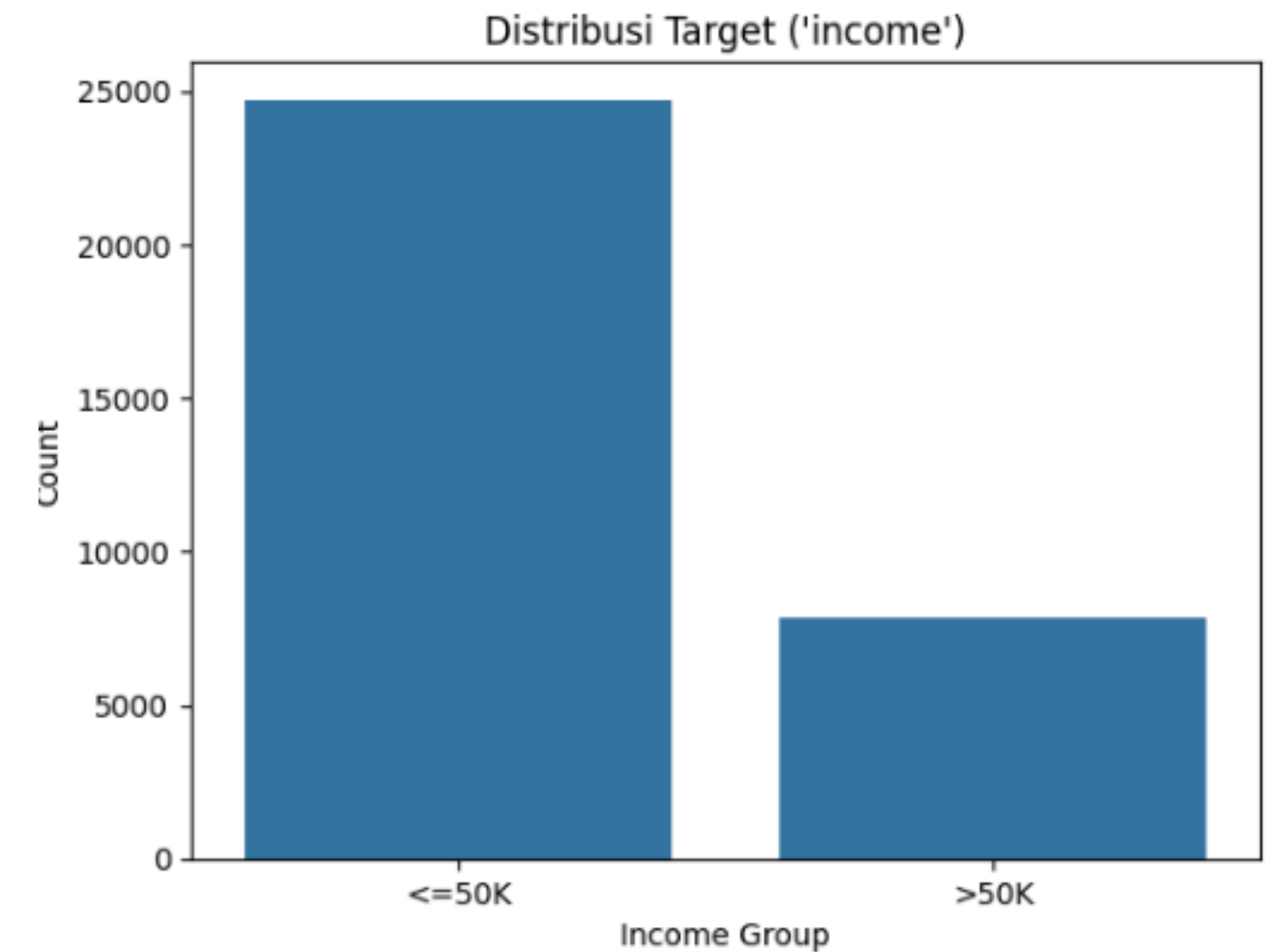
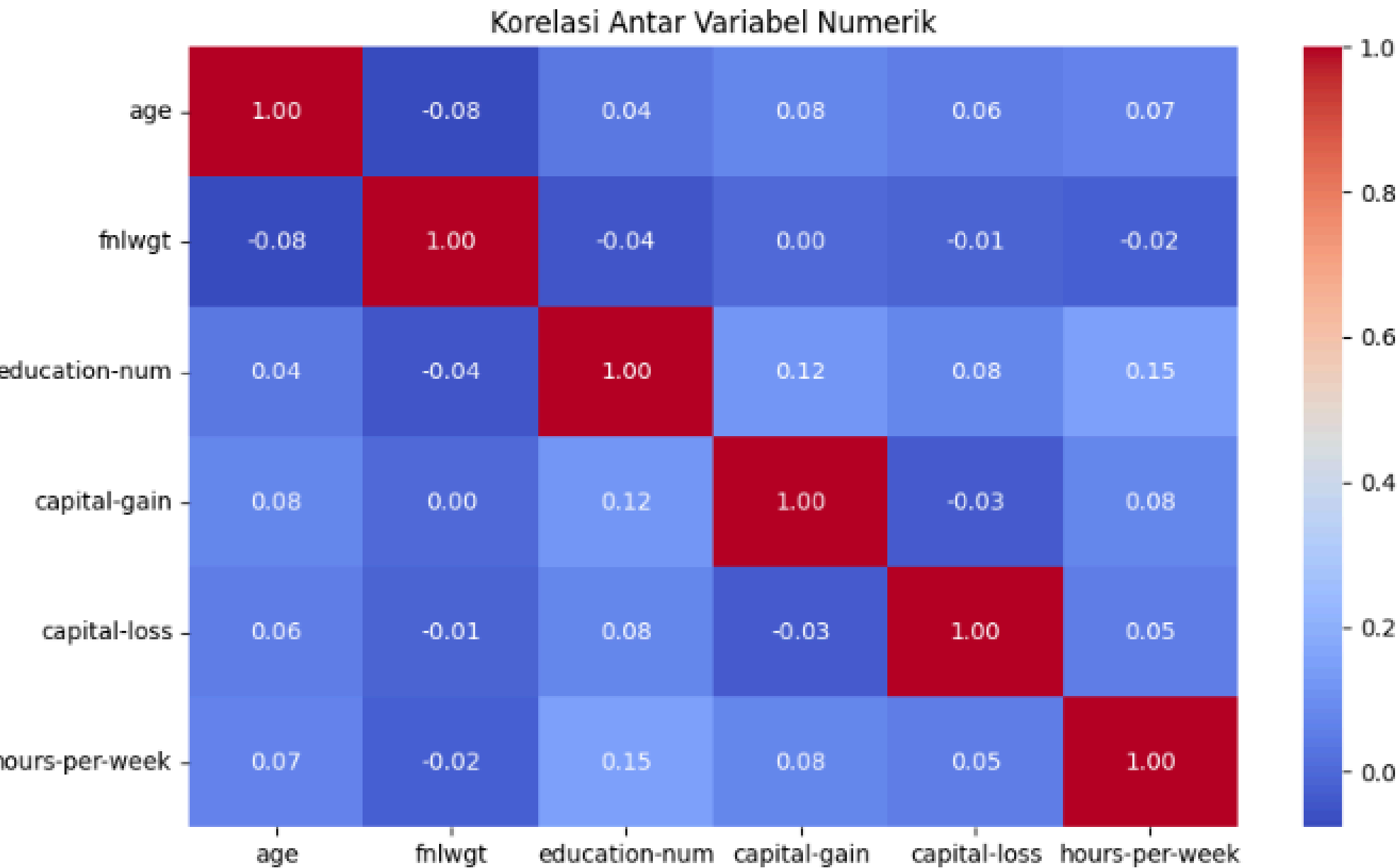
k-NN

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



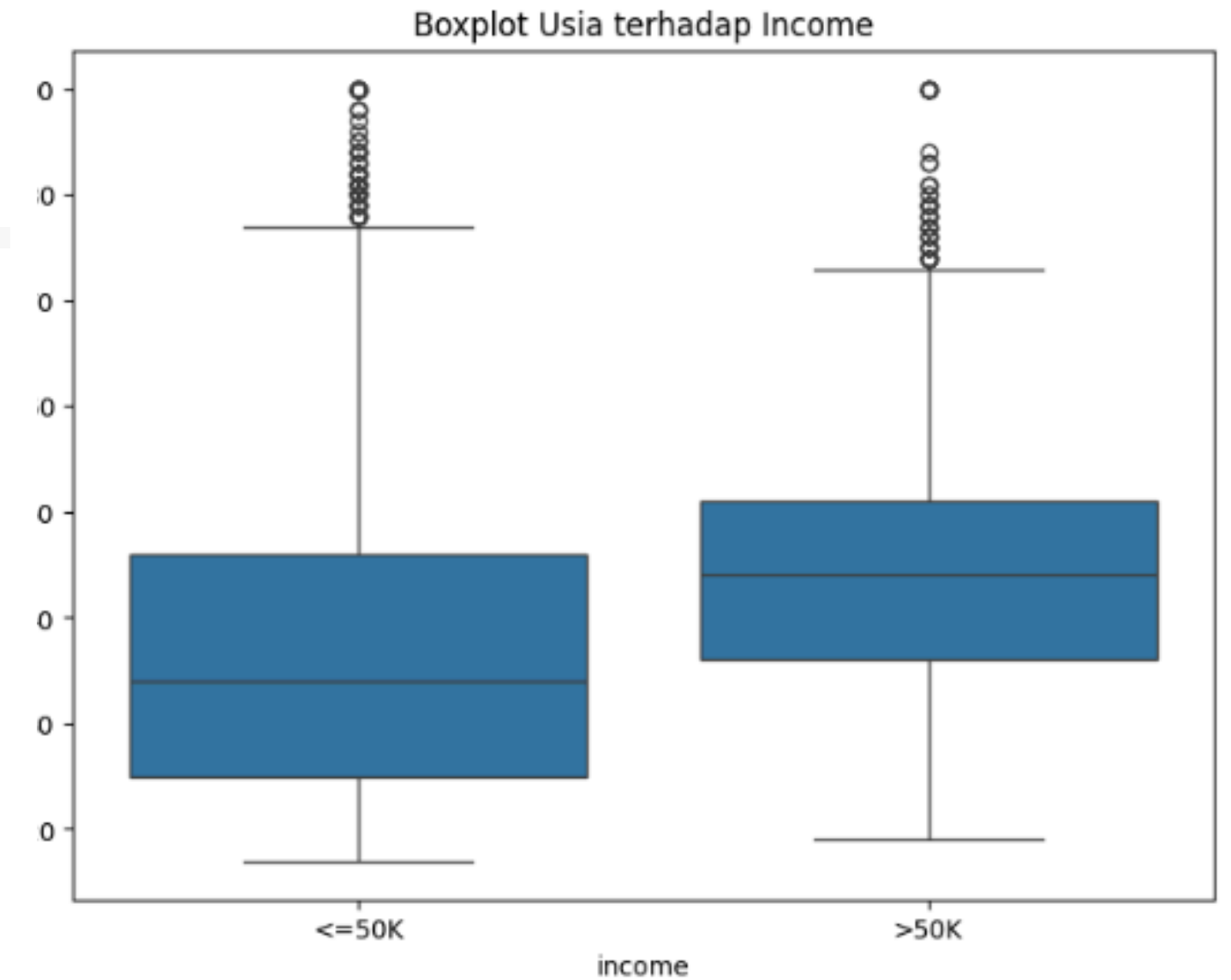
Hasil Dan Kesimpulan

```
Best Logistic Regression Params: {'classifier__C': 1, 'classifier__solver': 'lbfgs'}  
Best Logistic Regression Score: 0.8516585673420719
```

```
Best Decision Tree Params: {'classifier__max_depth': 10, 'classifier__min_samples_split': 10}  
Best Decision Tree Score: 0.8526677337864887
```

```
Best k-NN Params: {'classifier__n_neighbors': 7, 'classifier__weights': 'uniform'}  
Best k-NN Score: 0.834328001808669
```

```
Best XGBoost Params: {'classifier__max_depth': 3, 'classifier__n_estimators': 150}  
Best XGBoost Score: 0.8733767239357298
```



Thank You