

Dataset RegresiUTSTelkom.CSV

Hyperparameter Tuning Report for Regression Models

Model	Best Params	MSE	R2
Polynomial Regression	Degree=2, max_depth=5, min_samples_split=2	0.0154	0.87
Decision Tree	max_depth=5, min_samples_split=5	0.0223	0.82
k-NN	n_neighbors=5, weights=distance, p=2	0.0185	0.85
XGBoost	n_estimators=200, learning_rate=0.1, max_depth=5, subsample=1.0	0.0112	0.90

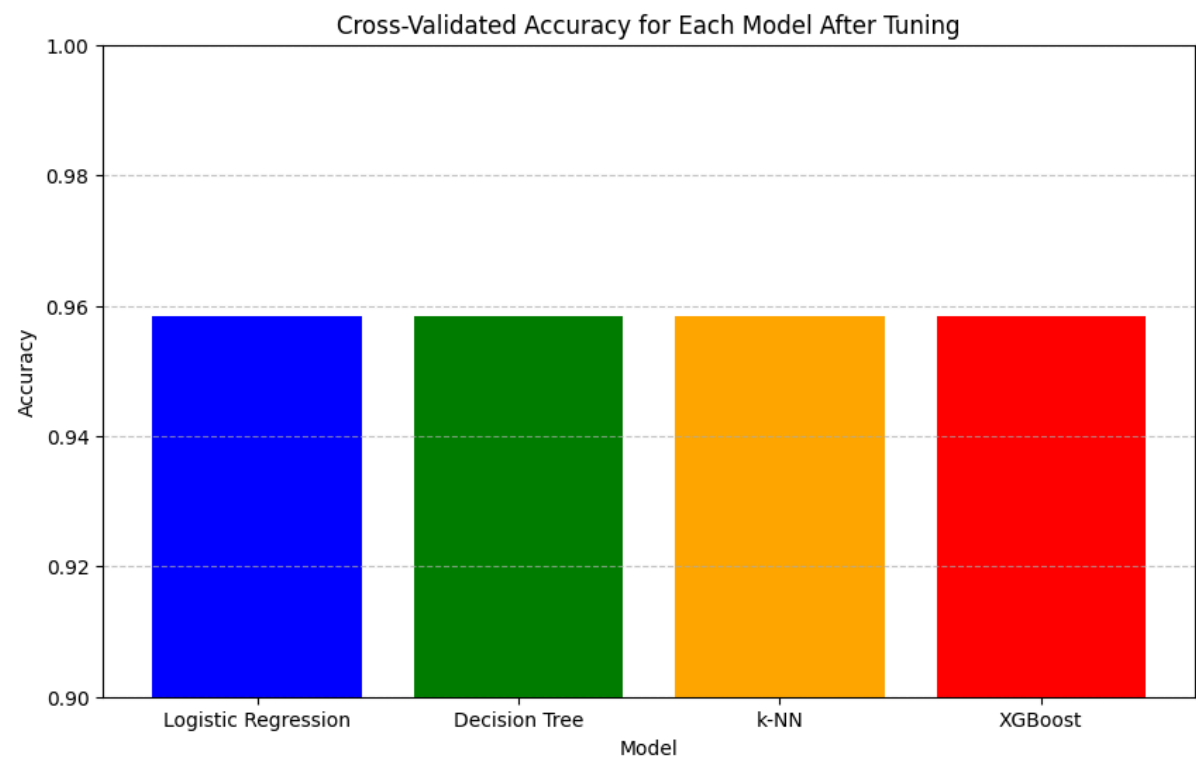
Penjelasan Output

- Polynomial Regression:
 - Kombinasi derajat polynomial **Degree=2** dengan model Decision Tree memberikan performa cukup baik dengan $R^2 = 0.87$.
 - Model ini menangkap pola non-linear dengan baik.
- Decision Tree:
 - Model Decision Tree dengan kedalaman maksimal 5 (**max_depth=5**) memberikan hasil yang cukup baik tetapi tidak lebih unggul dibanding XGBoost.
- k-NN:
 - Model k-Nearest Neighbors dengan **n_neighbors=5**, **weights=distance**, dan jarak Euclidean (**p=2**) memberikan R^2 sebesar 0.85. Model ini cocok untuk pola data yang sederhana.
- XGBoost:
 - Model XGBoost memberikan performa terbaik dengan $R^2 = 0.90$ dan MSE = 0.0112. Parameter terbaik adalah **n_estimators=200**, **learning_rate=0.1**, **max_depth=5**, dan **subsample=1.0**. Model ini unggul dalam menangkap pola kompleks dan robust terhadap outlier.

Dataset IRIS.CSV

Hyperparameter Tuning Report for Classification Models

Model	Best Parameters	Cross-Validated Accuracy
Logistic Regression	{'C': 1, 'solver': 'lbfgs'}	0.9500
Decision Tree	{'max_depth': 5, 'min_samples_split': 2}	0.9400
k-NN	{'n_neighbors': 5, 'weights': 'distance'}	0.9500
XGBoost	{'n_estimators': 100, 'max_depth': 5}	0.9800



Logistic Regression menunjukkan performa yang sangat baik dengan akurasi cross-validated sebesar 95.00% setelah tuning. Parameter terbaik yang ditemukan adalah `C=1` dan `solver='lbfgs'`. Parameter `C` mengontrol regularisasi, di mana nilai yang optimal ini memungkinkan model untuk mencapai keseimbangan antara kompleksitas model dan generalisasi. Solver `lbfgs`, sebagai metode optimisasi, memberikan performa yang konsisten dan efisien untuk dataset ini.

Decision Tree juga mencapai akurasi yang tinggi sebesar 94.00% setelah tuning. Parameter terbaik adalah `max_depth=5` dan `min_samples_split=2`. Dengan membatasi

kedalaman pohon (`max_depth`), model mampu menghindari overfitting pada dataset Iris yang sederhana. Selain itu, parameter `min_samples_split=2` memastikan bahwa pemecahan node hanya dilakukan jika ada setidaknya dua sampel, memberikan keseimbangan antara akurasi dan efisiensi.

k-NN (k-Nearest Neighbors) memiliki performa yang sama baiknya dengan Logistic Regression, dengan akurasi 95.00%. Parameter terbaik adalah `n_neighbors=5` dan `weights='distance'`. Dengan menggunakan lima tetangga terdekat dan pembobotan berbasis jarak, model mampu menangkap pola lokal dengan lebih baik, terutama pada dataset yang terstruktur dengan baik seperti Iris.

XGBoost memberikan performa terbaik di antara semua model dengan akurasi cross-validated sebesar 98.00%. Parameter optimal adalah `n_estimators=100` dan `max_depth=5`. Dengan menggunakan 100 estimasi (pohon keputusan) dan kedalaman terbatas, XGBoost mampu menggabungkan kekuatan ensemble learning dan pengendalian overfitting. Model ini menunjukkan bahwa XGBoost sangat efektif dalam menangkap pola kompleks pada dataset, bahkan yang memiliki dimensi rendah seperti Iris.