

Laporan Analisis UTS Bagian 2

Nama : Ricky Junianto Wijaya

NPM : 223400002

UTS Pembelajaran Mesin

1. Tugas 1 - Regresi Harga Rumah

Berdasarkan hasil evaluasi model, Multiple Linear Regression menunjukkan performa yang sedikit lebih baik daripada Polynomial Regression (Degree 2) dalam memprediksi harga rumah. Alasannya:

- R-squared: Multiple Linear Regression memiliki nilai R-squared sebesar 0.49, yang lebih tinggi daripada Polynomial Regression (0.41). Hal ini menunjukkan bahwa model Multiple Linear Regression mampu menjelaskan 49% variasi dalam data harga rumah, lebih baik daripada Polynomial Regression yang hanya mampu menjelaskan 41%.
- RMSE: Multiple Linear Regression memiliki nilai RMSE sebesar 8757622760.71, yang sedikit lebih rendah daripada Polynomial Regression 9352097086.46. RMSE yang lebih rendah menunjukkan bahwa model Multiple Linear Regression memiliki rata-rata error prediksi yang lebih kecil.

Justifikasi Pemilihan Model Terbaik: Berdasarkan hasil dari R-squared dan RMSE tersebut, Multiple Linear Regression saya pilih menjadi model terbaik untuk memprediksi harga rumah dalam konteks pada dataset ini. Model ini memiliki kemampuan yang lebih baik dalam menjelaskan variasi dalam data dan memiliki error prediksi yang lebih rendah.

Meskipun Multiple Linear Regression dipilih sebagai model terbaik, nilai R-squared yang masih relatif masih rendah (0.49) menunjukkan bahwa model masih punya potensi untuk ditingkatkan lagi. Ada faktor-faktor seperti pemilihan fitur yang lebih relevan dan penggunaan model yang lebih kompleks atau detail dapat dipertimbangkan untuk meningkatkan akurasi prediksi.

2. Tugas 2

- Gender Submission

Dari hasil evaluasi model yang sudah dilakukan pada dataset gender_submission, dilakukan perbandingan tiga model klasifikasi yaitu dengan Logistic Regression, kNN, dan SVM. Logistic Regression menunjukkan performa yang baik dengan akurasi, presisi, recall, dan F1 score yang cukup tinggi. kNN belum sebaik Logistic Regression, nilai metrik yang didapat lebih rendah. SVM juga menunjukkan performa yang baik, tetapi cenderung lebih rendah dibandingkan dengan Logistic Regression.

Berdasarkan hasil ini, Logistic Regression menjadi model terbaik karena performa yang paling baik dan konsisten, sederhana dan mudah di implementasikan, efisien dalam waktu pelatihan dan prediksi.

- Test dan Train

Berdasarkan dari hasil evaluasi tersebut, ketiga model menunjukkan akurasi yang cukup baik dalam memprediksi kelangsungan hidup penumpang. Logistic Regression menunjukkan akurasi yang baik dengan nilai akurasi, presisi, recall, dan F1 score yang relatif cukup tinggi. kNN disini memiliki performa yang sedikit lebih rendah dibanding dengan Logistic Regression, mulai dari akurasi, presisi, recall, dan F1 score mendapatkan nilai yang lebih rendah. SVM dengan kernel linear memiliki kinerja yang relatif baik terutama dalam akurasi, namun jika dibandingkan dengan Logistic Regression masih sedikit lebih rendah.

Dari hasil ini, Logistic Regression menjadi model terbaik dikarenakan memiliki akurasi paling baik dan keseimbangan metrik yang cukup seimbang, dan merupakan model dengan kompleksitas yang relatif sederhana.