

UTP RESPONSI AI 2024

Ketentuan:

1. Cari dan eksplorasi datasets sesuai dengan model yang ingin kalian kembangkan, dengan minimal jumlah data adalah 500 data. (setiap mahasiswa memiliki dataset yang berbeda). Referensi sumber dataset : <https://archive.ics.uci.edu/datasets> , <https://www.kaggle.com/> , <https://data.go.id/home>, dll.
2. Berikan penjelasan terkait dataset yang telah kalian pilih dan tujuan dibuatnya model untuk dataset tersebut.
3. Lakukan EDA dan preprocessing pada dataset yang kalian pilih. Preprocessing dapat meliputi **namun tidak terbatas** pada :
 - a. Data Cleaning
 - b. Feature Selection
 - c. Scaling
 - d. Fill NaN Value / Balancing
 - e. Data Type Converting
4. Bagilah data kalian menjadi train dan test. Rasio pembagiannya dapat mengikuti aturan berikut, ataupun menyesuaikan dengan kondisi data kalian :
 - a. Untuk data yang jumlahnya > 800 , bagi dengan rasio 4:1
 - b. Untuk data yang jumlahnya < 800 , bagi dengan rasio 3:2
5. Buat model klasifikasi atau regresi dengan menggunakan algoritma yang kalian inginkan yang sesuai dengan kebutuhan data kalian, model dapat berupa namun tidak terbatas pada :
 - a. Linear Regression
 - b. Logistic Regression
 - c. SVM
 - d. Decision Tree
 - e. KNN
 - f. DBSCAN.
6. Evaluasi model yang kalian buat, dengan akurasi minimum sebesar 60%.
7. Buat 5 (lima) data dummy baru untuk melakukan pengujian (prediksi class untuk model klasifikasi atau value untuk model regresi) dari model yang telah kalian buat.
8. Uji akurasi data dummy, dan buatlah kesimpulan dengan membandingkan hasil akurasi data dummy dengan akurasi data train sebelumnya.
9. Kumpulkan dalam bentuk link Google Colaboratory. Pastikan semua kode sudah di running dan menampilkan output.
10. Dilarang untuk melakukan plagiarisme terhadap *assignment* orang lain. Apabila ditemukan 2 (dua) atau lebih *assignment* berbeda dengan konten yang sama (kesamaan $> 80\%$) maka nilai akan dibagi sesuai dengan jumlah *assignment*.

Skema Penilaian

Range Nilai : 50 - 100 per poin penilaian

Poin Penilaian :

1. Pemilihan Dataset (10%)
2. Preprocessing (25%)
3. Model (30%)
4. Evaluasi (15%)
5. Dummy Data Test (10%)
6. Kesimpulan (10%)

Link Pengumpulan : <https://forms.gle/jdGpXdwBQaAM9XB49>

Deadline : 28 April 2024 pukul 23.59 WIB

***Dilarang melakukan perubahan kode atau running melebihi tanggal deadline.**

Terima kasih,

Selamat Mengerjakan^^.