

Homework

Data Pre-Processing

Final Project - Stage 2



Estimasi Waktu Pengerjaan



3 - 5 jam

Jumlah Soal



2 Soal

Total Point



100 poin

Teknis Pengerjaan

1. Pekerjaan dilakukan secara **berkelompok, sesuai kelompok Final Project**
2. Masing-masing anggota kelompok tetap perlu submit ke LMS (jadi bukan perwakilan)
3. File yang perlu dikumpulkan:
 - File **jupyter notebook** (.ipynb) yang berisi source code.
 - File **laporan homework** (.pdf) yang berisi rangkuman dari apa saja yang telah dilakukan.
4. Upload hasil pengerjaanmu melalui LMS.
 - Masukkan semua file ke dalam **1 file** dengan format **ZIP**.
 - Nama File:
Preprocessing - <Nama Kelompok>.zip

1. Data Cleansing (50 poin)

Lakukan pembersihan data, sesuai yang diajarkan di kelas, seperti:

- A. Handle missing values
- B. Handle duplicated data
- C. Handle outliers
- D. Feature transformation
- E. Feature encoding
- F. Handle class imbalance

Di laporan homework, tuliskan apa saja yang telah dilakukan dan metode yang digunakan.

* Tetap tuliskan jika memang ada tidak yang perlu di-handle (contoh: “Tidak perlu feature encoding karena semua feature sudah numerical” atau “Outlier tidak di-handle karena akan fokus menggunakan model yang robust terhadap outlier”).

2. Feature Engineering (35 poin)

Cek feature yang ada sekarang, lalu lakukan:

- A. Feature selection (membuang feature yang kurang relevan atau redundan)
- B. Feature extraction (membuat feature baru dari feature yang sudah ada)
- C. Tuliskan minimal 4 feature tambahan (selain yang sudah tersedia di dataset) yang mungkin akan sangat membantu membuat performansi model semakin bagus (ini hanya ide saja, untuk menguji kreativitas teman-teman, tidak perlu benar-benar dicari datanya dan tidak perlu diimplementasikan)

* Untuk 2A & 2B, tetap tuliskan jika memang tidak bisa dilakukan (contoh: “Semua feature digunakan untuk modelling (tidak ada yang dihapus), karena semua feature relevan”)

3. Git (15 poin)

Upload project teman-teman di sebuah repository git. Berkolaborasi di Git jika ada perubahan version dari waktu ke waktu.

- A. Buat Repository Git
- B. Upload file notebook atau file pengerjaan lainnya pada repository tersebut

Untuk file README, dapat merupakan summary dari proses data preproses yang telah dilakukan. Boleh menggunakan repositori yang sama atau membuat baru.

Selamat Mengerjakan!