

## Tugas Praktikum: Praproses Data

### Deskripsi Tugas

Pada praktikum ini, mahasiswa akan melakukan berbagai tahapan praproses data untuk memastikan dataset siap digunakan dalam analisis atau model machine learning. Mahasiswa akan mengatasi data yang hilang, mengidentifikasi dan menangani data outlier, serta melakukan transformasi data.

### Tujuan Praktikum

1. Memahami dan menerapkan langkah-langkah praproses data seperti pembersihan, transformasi, dan reduksi data.
2. Meningkatkan kualitas data melalui penanganan nilai yang hilang dan data yang tidak konsisten.
3. Menerapkan teknik transformasi data seperti normalisasi dan encoding.

### Langkah-Langkah Praktikum

#### 1. \*\*Unduh Dataset\*\*

Gunakan dataset House Prices - Advanced Regression Techniques dari Kaggle. Dataset ini berisi informasi tentang harga rumah dengan berbagai atribut seperti ukuran, tahun dibangun, dan kondisi rumah.

Link dataset: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

#### 2. \*\*Pembersihan Data\*\*

- Periksa jumlah nilai yang hilang pada setiap atribut menggunakan Python (`pandas`).
- Terapkan metode untuk menangani nilai yang hilang:
  - Isi dengan rata-rata (untuk atribut numerik) atau modus (untuk atribut kategoris).
  - Hapus kolom/atribut dengan proporsi nilai yang hilang lebih dari 50%.

#### 3. \*\*Deteksi dan Penanganan Outlier\*\*

- Gunakan boxplot untuk mendeteksi outlier pada atribut numerik seperti `'LotArea'` atau `'SalePrice'`.
- Tentukan apakah outlier perlu dihapus atau disesuaikan.

#### 4. \*\*Transformasi Data\*\*

- Lakukan normalisasi atribut numerik menggunakan skala 0-1 (min-max normalization).
- Lakukan encoding untuk atribut kategoris seperti `'Neighborhood'` menggunakan `*one-hot encoding*`.

#### 5. \*\*Reduksi Dimensi (Opsional)\*\*

- Gunakan analisis korelasi untuk mengidentifikasi atribut yang sangat berkorelasi dan buang salah satu jika perlu.
- Jika dataset besar, gunakan metode seperti PCA (\*Principal Component Analysis\*) untuk mengurangi dimensi data.

#### 6. **\*\*Laporan Praktikum\*\***

- Buat laporan singkat (1-2 halaman) yang mencakup langkah-langkah yang dilakukan, hasil dari setiap tahapan praproses, dan dataset akhir setelah praproses.

### **Hasil yang Diharapkan**

1. Dataset yang sudah bersih dan siap digunakan, lengkap dengan penanganan nilai yang hilang dan outlier.
2. Transformasi data yang telah dilakukan (normalisasi dan encoding).
3. Laporan singkat yang mencakup hasil tiap langkah praproses.

### **Penilaian**

1. **\*\*Pembersihan Data (30%)\*\***: Kemampuan mahasiswa menangani data yang hilang dan data yang tidak konsisten.
2. **\*\*Transformasi Data (30%)\*\***: Kesesuaian penerapan teknik transformasi seperti normalisasi dan encoding.
3. **\*\*Reduksi Dimensi (20%)\*\***: Pemahaman dalam memilih atribut yang relevan (opsional).
4. **\*\*Laporan Praktikum (20%)\*\***: Kerapihan dan kelengkapan laporan.