

## Tugas Praktikum: Eksplorasi Data

### Deskripsi Tugas

Pada praktikum ini, mahasiswa akan melakukan eksplorasi data untuk memahami karakteristik dataset yang diberikan. Mahasiswa akan mengidentifikasi atribut, mendeteksi nilai yang hilang, menganalisis distribusi data, serta memvisualisasikan hubungan antar atribut.

### Tujuan Praktikum

1. Memahami langkah-langkah eksplorasi data.
2. Mengidentifikasi atribut dalam dataset (nominal, ordinal, numerik).
3. Menganalisis nilai yang hilang dan outlier.
4. Membuat visualisasi data untuk mendukung analisis.

### Langkah-Langkah Praktikum

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

Gunakan dataset Titanic - Machine Learning from Disaster dari Kaggle. Dataset ini berisi informasi tentang penumpang kapal Titanic dengan berbagai atribut seperti usia, jenis kelamin, kelas tiket, dan status kelangsungan hidup.

Link dataset: <https://www.kaggle.com/c/titanic/data>

#### 2. **Import dan Pemahaman Dataset**

- Impor dataset menggunakan Python (`pandas`).
- Tampilkan 5 baris pertama dataset untuk memahami struktur data.
- Identifikasi tipe atribut: nominal, ordinal, numerik.

#### 3. **Analisis Statistik**

- Hitung statistik deskriptif (rata-rata, median, modus, standar deviasi) untuk atribut numerik.
- Identifikasi jumlah nilai yang hilang pada setiap atribut.

#### 4. **Deteksi Outlier**

- Gunakan visualisasi seperti boxplot untuk mendeteksi outlier pada atribut numerik.

#### 5. **Visualisasi Data**

- Buat histogram untuk melihat distribusi usia penumpang.
- Buat scatter plot untuk menganalisis hubungan antara atribut `Fare` (harga tiket) dan `Age` (usia).
- Gunakan diagram batang untuk menunjukkan jumlah penumpang berdasarkan kelas tiket (`Pclass`).

#### 6. **\*\*Laporan Analisis\*\***

- Tulis hasil analisis dalam bentuk laporan singkat (1-2 halaman).
- Sertakan visualisasi yang dibuat dan interpretasi terhadap data.

#### **Hasil yang Diharapkan**

1. Statistik deskriptif untuk atribut numerik.
2. Visualisasi data berupa histogram, scatter plot, dan diagram batang.
3. Penanganan nilai yang hilang (jika ada) dan identifikasi outlier.

#### **Penilaian**

1. **\*\*Pemahaman Dataset (20%)\*\***: Mahasiswa mampu mengidentifikasi atribut dengan benar.
2. **\*\*Analisis Statistik (30%)\*\***: Mahasiswa dapat menghitung statistik deskriptif dan mengidentifikasi nilai yang hilang.
3. **\*\*Visualisasi Data (30%)\*\***: Kualitas visualisasi dan interpretasinya.
4. **\*\*Laporan (20%)\*\***: Kerapihan, kelengkapan, dan kejelasan laporan.