**EKSPLORASI DAN PENGOLAHAN DATA KATEGORI PADA DATA KORBAN KAPAL TITANIC**

**PRAKTIKUM PEMROSESAN DATA**

**Oleh**

**Nama**

**NIM**

**(Program Studi Informatika)**

**A black background with a black square

Description automatically generated with medium confidence**

**UNIVERSITAS TRISAKTI  
Oktober 2024**

**I. Pendahuluan**

Pada praktikum ini, peserta akan mempelajari fundamental pemrosesan data, dengan fokus khusus pada Exploratory Data Analysis (EDA) dan teknik encoding untuk data kategorikal. EDA bertujuan untuk memahami karakteristik dataset yang diberikan, membantu menentukan langkah-langkah yang diperlukan, seperti membersihkan data atau melakukan transformasi dan scaling. Melalui EDA, peserta akan mengeksplorasi dataset dengan tujuan mengidentifikasi fitur, target, serta memperbaiki kualitas data dengan penanganan missing values, outlier, dan inkonsistensi data.

Selain itu, praktikum ini juga membahas dua teknik encoding yang umum digunakan dalam pemrosesan data kategorikal, yaitu One-Hot Encoding dan Ordinal Encoding. One-Hot Encoding digunakan untuk merepresentasikan kategori tanpa urutan dalam bentuk biner, di mana setiap kategori akan diubah menjadi kolom biner yang menunjukkan kehadiran atau ketiadaan kategori tersebut. Teknik ini cocok untuk data kategorikal seperti jenis kelamin atau warna. Sedangkan, Ordinal Encoding digunakan untuk data kategorikal yang memiliki urutan bermakna, seperti level prioritas "Low", "Medium", dan "High", yang akan di-encode menjadi 0, 1, dan 2 sesuai dengan urutannya.

Melalui praktikum ini, saya diharapkan tidak hanya memahami langkah-langkah penting dalam EDA, tetapi juga menguasai proses pembersihan dan transformasi data. Semua keterampilan ini sangat berharga untuk mempersiapkan data agar siap digunakan dalam analisis yang lebih mendalam atau pemodelan yang lebih kompleks.

**II. Hasil Praktikum**

**II.1 *Import Library***

[Penjelasan fungsi dari setiap library dan sertakan screenshot]

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.2 Menyiapkan *Dataset***

|  |
| --- |

**Penjelasan :** Jelaskan tipe data dari fitur yang ada!

|  |
| --- |

**II.3 Menampilkan kolom dataset**

|  |
| --- |

**Penjelasan :** Apa saja kolom yang tersedia didataset tersebut?

|  |
| --- |

**II.4 Menghapus data beberapa kolom yang tidak diperlukan**

|  |
| --- |

**Penjelasan :** Mengapa pada kolom PassengerId, Cabin dan Ticket harus kita hapus/ drop?

|  |
| --- |

**II.5 Mengecek Jumlah nilai unik pada setiap kolom**

|  |
| --- |

**Penjelasan :** Apa tujuan dari mengecek jumlah nilai unik pada setiap kolom dalam dataframe?

|  |
| --- |

**II.6 Melihat nilai unik pada kolom Sex dan Embarked**

|  |
| --- |

**Penjelasan :** Apakah semua kolom di dataframe harus memiliki nilai unik seperti Sex dan Embarked? Mengapa?

|  |
| --- |

**II.7 Mengecek nilai untuk imputasi kolom Embarked**

|  |
| --- |

**Penjelasan :** Mengapa penting untuk melakukan imputasi pada kolom Embarked?

|  |
| --- |

**II.8 Mengisi dan mengecek nilai kosong pada kolom embarked**

|  |
| --- |

**Penjelasan :** Pada code tersebut nilai NaN pada kolom Embarked diisi oleh nilai apa?

|  |
| --- |

**II.9 Melakukan one hot encoding**

|  |
| --- |

**Penjelasan :** Mengapa memilih kolom Sex dan Embarked untuk dilakukan ***one-hot encoding?***

|  |
| --- |

**II.10 Melakukan One-Hot Encoding K-1**

|  |
| --- |

**Penjelasan :** perbedaan apa yang anda lihat dari penggunaan drop\_first?

|  |
| --- |

**II.11 Mengabungkan kolom data df dengan data df\_ohe\_kmin1**

|  |
| --- |

**Penjelasan :** Apa tujuan dari menggabungkan dataframe df dengan df\_ohe\_kmin1?

|  |
| --- |

**II.12 Mengisi nilai Nilai Kosong dengan Median**

|  |
| --- |

**Penjelasan :** Apa tujuan dari mengisi nilai kosong di kolom Age dengan median?

|  |
| --- |

**II.13 Memvisualisasi Distribusi Kolom Age dan Fare Menggunakan Histogram**

|  |
| --- |

**Penjelasan :** Apa informasi yang bisa diperoleh dari histogram kolom Age dan Fare?

|  |
| --- |

**II.14 Melakukan Pelabelan untuk Binning Data**

|  |
| --- |

**Penjelasan :** Untuk apa dilakukan Binning data?

|  |
| --- |

**II.15 Membagi Kolom Age Menjadi 8 Kategori dengan Label**

|  |
| --- |

**Penjelasan :** Apa tujuan dari menggunakan fungsi pd.cut dalam konteks kolom Age?

|  |
| --- |

**II.16 Melakukan diskretisasi Usia Menggunakan Fungsi pd.cut**

|  |
| --- |

**Penjelasan :** Mengapa penting untuk melakukan diskretisasi pada kolom numerik seperti Age?

|  |
| --- |

**II.17 Melakukan visualisasi Jumlah Observasi Per Kategori Usia (Age)**

|  |
| --- |

**Penjelasan :** Apa tujuan dari visualisasi jumlah data per bin menggunakan bar chart?

|  |
| --- |

**II.18 Membagi Data Usia Menjadi 10 Kategori Berdasarkan kuantil**

|  |
| --- |

**Penjelasan :** Setelah mendiskretkan kolom Age dengan kedua metode ini, langkah analisis apa yang dapat dilakukan?

|  |
| --- |

**II.19 Memvisualisasi Jumlah Observasi Per Kategori Usia Berdasarkan Kuantil**

|  |
| --- |

**Penjelasan :** Apa yang dapat Anda simpulkan jika terdapat ketidakseimbangan dalam jumlah data di setiap bin?

|  |
| --- |

**Kesimpulan : Minimal 5 Baris**

|  |
| --- |