**EXPLORATORY DATA ANALYSIS (EDA) DAN ENCODING**

**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  
September 2024**

**I. Pendahuluan**

Pada praktikum ini, peserta akan mempelajari fundamental pemrosesan data, Exploratory Data Analysis (EDA), dan Encoding. 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.

Praktikum ini diharapkan dapat memberikan pemahaman mendalam tentang langkah-langkah penting dalam pemrosesan data, mulai dari eksplorasi hingga transformasi yang diperlukan agar data siap digunakan dalam analisis atau pemodelan lebih lanjut.

**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!

|  |
| --- |

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.3 Menampilkan data missing value**

|  |
| --- |

**Penjelasan :** Fitur mana saja yang terdapat missing value dan apa penyelesaian langkah untuk fitur yang memiliki nilai missing value?

|  |
| --- |

**II.4 Mengecek distribusi Age menggunakan fungsi histogram**

|  |
| --- |

**Penjelasan :** Jelaskan berapa range umur penumpang yang paling banyak!

|  |
| --- |

**II.5 Mengisi data dengan median pada kolom Age**

|  |
| --- |

**Penjelasan :** Apa yang dimaksud imputasi dengan median pada kolom Age dan untuk apa digunakannya?

|  |
| --- |

**II.6 Menampilkan histogram dari kolom-kolom numerik**

|  |
| --- |

**Penjelasan :** Jelaskan informasi apa yang kalian dapatkan dari tampilan ketiga histogram tersebut!

|  |
| --- |

**II.7 Melakukan Scaling pada beberapa fitur**

|  |
| --- |

**Penjelasan :** Mengapa harus dilakukan standarisasi?

|  |
| --- |

**II.8 Membandingkan rata-rata pada beberapa fitur**

|  |
| --- |

**Penjelasan :** Jelaskan hubungan antara fitur Age, Fare, Parch, SibSp dengan Survived!

|  |
| --- |

**II.9 Menampilkan nilai unik pada beberapa fitur**

|  |
| --- |

**Penjelasan :** Jelaskan apa yang dimaksud dengan nilai unik dalam dataframe tersebut!

|  |
| --- |

**II.10 Menghapus beberapa nilai pada kolom embarked**

|  |
| --- |

**Penjelasan :** Mengapa kita harus menghapus beberapa nilai pada kolom embarked?

|  |
| --- |

**II.11 Menampilkan hasil kolom setelah didrop**

|  |
| --- |

**Penjelasan :** Mengapa pada fitur cabin, ticket dan passengerId di drop?

|  |
| --- |

**II.12 Melihat kolom fitur setelah didrop**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.13 Mengubah tipe data kolom Survived menjadi kategori**

|  |
| --- |

**Penjelasan :** Apa tipe data fitur Survived sebelumnya? Dan kenapa tipe datanya harus diubah menjadi kategori?

|  |
| --- |

**II.14 Melakukan one hot encoding menggunakan library pandas dan nilai unik dari kolom Embarked**

|  |
| --- |

**Penjelasan :** Apa yang dimaksud dengan One-Hot Encoding da apa yang direpresentasikan One-Hot Encoding terhadap fitur Embarked?

|  |
| --- |

**II.15 Menghapus kolom Embarked dari dataframe dan menambahkan hasil One-hot encoding ke dalam dataframe**

|  |
| --- |

**Penjelasan :** Mengapa kolom Embarked\_C, Embarked\_Q, Embarked\_S, bisa masuk ke dalam dataframe ?

|  |
| --- |

**II.16 Melihat data unik pada kolom Sex, menghapus baris yang bernilai NaN dan mengubah tipe data Sex menjadi Kategori**

|  |
| --- |

**Tampilkan hasil kolom dataframe setelah diubah tipedatanya!**

|  |
| --- |

**Penjelasan : Mengapa harus melihat data unik, menghapus baris yang bernilai NaN sebelum mengubah tipe data tersebut?**

|  |
| --- |

**II.17 Mengubah kolom Sex dari Kategori menjadi biner menggunakan One-hot Encoding (One-hot Encoder)**

|  |
| --- |

**Penjelasan :** Jelaskan apa hubungan mengganti tipe data Sex dari kategori menjadi biner dengan mengecek distribusi menggunakan fungsi histogram!

|  |
| --- |

**II.18 Cek Distribusi Sex\_Female menggunakan Fungsi Histogram**

|  |
| --- |

**II.19 Tampilkan Screenshoot jika Anda menghapus kolom Sex dari dataframe dan menambahkan hasil One-hot encoding ke dalam dataframe**

|  |
| --- |

**Hasil tampilan Setelah ditambahkan**

|  |
| --- |

**Penjelasan :** Mengapa kolom Sex\_female, Sex\_male bisa masuk ke dalam dataframe ?

|  |
| --- |

**II.20 Menghapus baris yang memiliki nilai NaN pada kolom hasil One-hot Encoding untuk kolom Embarked**

|  |
| --- |

**Hasil tampilan setelah dimenghapus baris yang memiliki nilai NaN**

|  |
| --- |

**Penjelasan :** Apa yang dimaksud dengan df.dropna dan mengapa nilai Non-Null Countnya untuk semua fitur sama = 887?

|  |
| --- |

**II.21 Mengubah tipe data kolom Pclass dari Kategori menjadi Biner menggunakan OrdinalEncoding**

|  |
| --- |

**Penjelasan :** Apa yang dimaksud dengan Ordinal Encoder? Dan apa yang terjadi setelah dilakukannya Encoding menggunakan Ordinal Encoder?

|  |
| --- |

**II.22 Menghapus kolom Pclass dari dataframe dan menambahkan hasil One-hot encoding ke dalam dataframe**

|  |
| --- |

**Hasil Tampilan setelah ditambahkan hasil One-Hot Encoding kedalam data frame**

|  |
| --- |

**Penjelasan :** Mengapa kolom sex\_female dan sex\_male bisa masuk ke dalam data frame?

|  |
| --- |

**II.23 Melakukan standarisasi pada fitur yang bertipe data Numerik**

|  |
| --- |

**Penjelasan :** Mengapa melakukan standarisasi kepada fitur yang tipe data yang numerik dan tidak pada tipe data yang kategori?

|  |
| --- |

**Kesimpulan :** Minimal 5 Baris

|  |
| --- |