**EXPLORATORY DATA ANALYSIS (EDA) PADA DATA**

**UKURAN BAJU**

**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 akan dilakukan pembelajaran mengenai fundamental pada pemrosesan data dan *exploratory data analysis* (EDA). Hal ini berfungsi untuk mengetahui tindakan yang harus dilakukan pada saat diberikan sebuah *dataset*, baik untuk membersihkan data atau melakukan transformasi/*scaling* terhadap data yang diberikan.

Pada praktikum ini dilakukan untuk menjawab pertanyaan:

1. Meng-*load* data yang diberikan pada komputer.
2. Melakukan pengelompokkan data yang diberikan merupakan Fitur atau target.
3. Melakukan pembersihan dan penanganan data jika diperlukan, seperti penanganan *missing value*, penanganan data *outlier*, penanganan data tidak konsisten, dsb.

**II. Hasil Praktikum**

**II.1 Import Library**

[Penjelasan fungsi dari setiap library dan sertakan screenshot]

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.2 Menyiapkan Dataset**

|  |
| --- |

**Penjelasan :** Jelaskan tipe data pada setiap kolom di dataset tersebut?

|  |
| --- |

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.3 Menampilkan statistik deskriptif berdasarkan data**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.4 Menampilkan jumlah kemunculan setiap nilai unik berdasarkan fitur Size**

|  |
| --- |

**Penjelasan :** informasi apa yang kalian dapat berdasarkan data yang muncul diatas?

|  |
| --- |

**II.5 Menampilkan grafik count plot berdasarkan jumlah kemunculan setiap nilai unik fitur Size**

|  |
| --- |

**Penjelasan :** Jelaskan informasi apa saja yang bisa didapatkan dari tampilan grafik diatas?

|  |
| --- |

**II.6 Menampilkan histogram berdasarkan beberapa fitur pada data set**

|  |
| --- |

**Penjelasan :** Jelaskan informasi apa yang bisa didapatkan dari tampilan grafik diatas?

|  |
| --- |

|  |
| --- |

**Penjelasan :** Jelaskan informasi apa yang bisa didapatkan dari tampilan grafik diatas?

|  |
| --- |

|  |
| --- |

**Penjelasan :** Jelaskan informasi apa yang bisa didapatkan dari tampilan grafik diatas?

|  |
| --- |

**II.7 Menghapus Outlier**

|  |
| --- |

**Penjelasan :** Apa yang terjadi jika tidak menghapus outlier dalam dataset?

|  |
| --- |

**II.8 Menampilkan dan mengisi nilai yang hilang**

|  |
| --- |

**Penjelasan :** Apa yang dimaksud dengan missing values, dan bagaimana nilai ini bisa muncul dalam data? Dan bagaimana cara menangani missing values setelah mengetahui jumlahnya?

|  |
| --- |

|  |
| --- |

**Penjelasan :** Berdasarkan penjelasan sebelumnya, pada code tersebut menggunakan Langkah apa untuk mengisi missing value? Jelaskan!

|  |
| --- |

**II.9 Mengubah ukuran pakaian berdasarkan fitur Size**

|  |
| --- |

**Penjelasan :** Apa tujuan dari kode **map()** yang digunakan pada kolom **size** dalam DataFrame? Mengapa perlu mengubah kategori ukuran pakaian dari string menjadi angka?

|  |
| --- |

**II.10 Menampilkan nilai yang hilang**

|  |
| --- |

**Penjelasan :** Berdasarkan tampilan diatas, apakah masih terdapat missing value pada setiap kolom?

|  |
| --- |

**Tugas :** Buktikan tampilan diatas dengan jumlah outlier dalam bentuk histogram dari fitur Age sampai Size!

|  |
| --- |

**II.11 Melakukan Feature Engineering**

|  |
| --- |

**Penjelasan :** Apa itu feature Engineering? Jelaskan kolom baru yang dibuat tersebut diisi berdasarkan nilai apa? Dan apa Tujuan membuat kolom baru?

|  |
| --- |

**II.12 Menampilkan Heatmap**

|  |
| --- |

**Penjelasan :** Berdasarkan Heatmap diatas informasi apa yang anda dapat? Jelaskan!

|  |
| --- |

**II.13 Membagi data fitur dan target**

|  |
| --- |

**Penjelasan :** Berdasarkan tampilan diatas, tentukan mana saja yang menjadi data fitur dan targetnya!

|  |
| --- |

**II.14 Menampilkan data X atau data fitur**

|  |
| --- |

**Penjelasan :** berdasarkan tampilan diatas, data tersebut menampilkan apa? Jelaskan!

|  |
| --- |

**II.15 Menyimpan df\_raw ke dalam file CSV**

|  |
| --- |

**Penjelasan :** Apa fungsi dari compression\_opts dalam kode tersebut?

|  |
| --- |

**II.16 Mensplit data menjadi data train dan test**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.17 Menampilkan jumlah baris pada data train dan data test**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.18 Melakukan training dengan model Naïve Bayes**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**II.19 Membandingkan performa model dalam bentuk grafik**

|  |
| --- |

**Penjelasan :**

|  |
| --- |

**Kesimpulan : Minimal 5 Baris**

|  |
| --- |