|  |  |  |
| --- | --- | --- |
| **Nama:**  **(Adrian Maulana Yusup)**  **NIM: (065002100036)** | C:\Users\RPL-SI 02\Pictures\288px-Trisakti_Logo.svg.png | **MODUL 7**  **Nama Dosen:**  **Dedy Sugiarto** |
| **Hari/Tanggal:**  **Kamis, 3 November 2022** | **Praktikum Statistika** | **Nama Asisten Labratorium:**   1. **Azhar Rizki Zulma 065001900001** 2. **Arfa Maulana 064001900039** |

**Data Preprocessing Menggunakan Python**

1. **Teori Singkat**

Data Preprocessing adalah sebuah tahapan awal dalam sebuah pengolahan data sebelum data diaplikasikan dengan algoritma machine learning. Data yang biasanya kita gunakan dalam kehidupan sehari — hari entah itu dari database, data excel dan sumber lainnya, merupakan data unstruktur (datanya tidak sempurna). Misalkan dalam sebuah dataset (kumpulan data) terdapat data yang kosong, tipe data yang berbeda dengan yang lain, dan sebagainya. Masalah tersebut harus bisa kita selesaikan terlebih dahulu agar data yang kita kelola lebih mudah dan outputnya sesuai dengan yang kita harapkan.

Terdapat beberapa case yang akan kita pelajari satu per satu, antara lain seperti:

* Mengimport libraries
* Mengimport dataset
* Menangani data kosong di dataset
* Mengolah data string menjadi kategori
* Membagi dataset menjadi training dan test set
* Feature Scaling

**Informasi Dataset**

Sumber Data: Kaggle

Deskripsi: Memberikan informasi dari penumpang Titanic yang selamat dan tidak.

Jumlah data: 1309

Jumlah atribut: 12 (termasuk class)

Terdiri dari:

* PassengerId urutan nomor data dari penumpang
* Survived: status selamat (0:meninggal, 1:selamat)
* Pclass: kelas kamar dari penumpang (1: highclass, 2:midclass, 3:lowclass)
* Name: nama penumpang
* Sex: jenis kelamin penumpang (male, female)
* Age: umur penumpang
* SibSp: jumlah saudara kandung dan pasangan dari penumpang yang ada di kapal
* Parch: jumlah orangtua dan anak dari penumpang
* Ticket: kode tiket penumpang
* Fare: ongkos tiket yang dibeli penumpang
* Cabin: Kode kabin
* Embarked: Kota keberangkatan penumpang (C:Cherbourg, Q:Queenstown, S:Southampton)

1. **Alat dan Bahan**

Hardware : Laptop/PC

Software : R Studio

1. **Elemen Kompetensi**
   1. Latihan pertama – Materi

1. Buka Jupyter Notebook atau gunakan Google Colab

|  |
| --- |
|  |

2. Script

|  |
| --- |
| import pandas as pd  #memanggil dan menampilkan dataset  data\_nama = pd.read\_csv('D:/dll/titanic.csv')  print(data\_nama) |

Output:

|  |
| --- |
|  |

3. Script

|  |
| --- |
| #mengambil data pada kolom tertentu  data1 = data\_nama.loc[:,['Age','Pclass','Survived']]  print(data1) |

Output:

|  |
| --- |
|  |

4. Script

|  |
| --- |
| #memvisualisasikan data titanic  data2 = data\_nama[['Age', 'Pclass', 'Survived']]  data2.plot(title='Persebaran Data', x='Age', y='Pclass', kind='scatter', c='Survived', colormap='Paired') |

Output:

|  |
| --- |
|  |

5. Script

|  |
| --- |
| #memanipulasi data jumlah penumpang berdasarkan group Pclass  data3 = data\_nama[['Name', 'Sex', 'Age', 'Pclass', 'Fare']]  penumpang=data3.groupby('Pclass')['Name'].nunique()  print('Jumlah Penumpang:\n', penumpang) |

Output:

|  |
| --- |
|  |

6. Script

|  |
| --- |
| #memfilter data penumpang yang selamat berdasarkan pclass  data4 = data\_nama[['Name', 'Sex', 'Age', 'Pclass', 'Fare']]  notsurvivedpassanger=data4['Pclass'].loc[data\_nama['Survived']==0]  print('Penumpang yang tidak survived:\n', notsurvivedpassanger.value\_counts())  survivedpassanger=data4['Pclass'].loc[data\_nama['Survived']==1]  print('\nPenumpang yang survived:\n', survivedpassanger.value\_counts()) |

Output:

|  |
| --- |
|  |

* 1. Latihan Kedua – Tugas

1. Buatlah manipulasi data jumlah penumpang berdasarkan group sex

Script:

|  |
| --- |
| #memanipulasi data jumlah penumpang berdasarkan group Sex data5 = data\_adrian[['Name', 'Sex', 'Age', 'Pclass', 'Fare']] penumpang=data5.groupby('Sex')['Name'].nunique() print('Jumlah Penumpang:\n', penumpang) |

Output:

|  |
| --- |
|  |

Penjelasan: Screenshot di atas merupakan pemanipulasian data di Jupyter Notebook dari data titanic jumlah penumpang berdasarkan Sex(Gender).

2. Buatlah filter data penumpang yang selamat berdasarkan sex

Script:

|  |
| --- |
| #memfilter data penumpang yang selamat berdasarkan pclass data6 = data\_adrian[['Name', 'Sex', 'Age', 'Pclass', 'Fare']] notsurvivedpassanger=data4['Sex'].loc[data\_adrian['Survived']==0] print('Penumpang yang tidak survived:\n', notsurvivedpassanger.value\_counts()) survivedpassanger=data4['Pclass'].loc[data\_adrian['Survived']==1] print('\nPenumpang yang survived:\n', survivedpassanger.value\_counts()) |

Output:

|  |
| --- |
|  |

Penjelasan: Screenshot di atas merupakan filter data penumpang yang selamat berdasarkan Sex/Gender, dari data titanic. Data tersebut merupakan tampilan dari penumpang survived dan not survived.

1. **File Praktikum**

Github Repository:

|  |
| --- |
|  |

1. **Soal Latihan**

Soal:

1. Perintah apa yang digunakan untuk mengimport data CSV pada bahasa pemrograman python?
2. Apa kegunaan dari filter data?

Jawaban:  
1. Perintah yang digunakan untuk import data CSV adalah import pandas as pd, kemudian definisikan variabel, lalu gunakan method read\_csv untuk mengimport data  
2. Kegunaan filter data adalah untuk menyaring atau mengkategorikan data yang berjumlah banyak berdasarkan informasi data yang ingin diketahui.

1. **Kesimpulan**
   1. Dalam pengerjaan praktikum Statistika, Mahasiswa mampu menemukan solved problem dari data yang tersedia bukan hanya dengan Rstudio, namun juga dapat menggunakan Python dengan IDE Jupyter Notebook.
   2. Kita juga dapat mengetahui Teknik filterisasi data yang dapat membantu mahasiswa dalam menampilkan data yang diinginkan dari data yang berjumlah banyak.
2. **Cek List (✓)**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| **Selesai** | **Tidak Selesai** |
| **1.** | Latihan Pertama | **√** |  |
| **2.** | Latihan Kedua | **√** |  |

1. **Formulir Umpan Balik**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| **1.** | Latihan Pertama | 20 Menit | 1 |
| **2.** | Latihan Kedua | 20 Menit | 1 |

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang