

# PRAKTIKUM MACHINE LEARNING

## UNIT 4 : DEFINE DATA OBJECT



`adult_klasifikasi.ipynb`

Eka Praja Wiyata Mandala, S.Kom, M,Kom, CADS

# Unit 4 : Define Data Object

## ✓ Unit 4: Menentukan Objek Data

### Memutuskan kriteria dan teknik pemilihan data

Kuantitas dataset **Adult Income** yang digunakan adalah:

- Memiliki 14 Features (variabel)
- Terdiri dari 48842 Instance (record)
- Ukuran data sebesar 605.7 KB

Kualitas dataset **Adult Income** yang digunakan adalah:

- Digunakan untuk Task: Classification
- Feature Type : Categorical dan Integer
- Memiliki Missing Values

# Unit 4 : Define Data Object

3

**Menentukan attributes (columns) dan records (row) data**

**Menentukan attributes (columns)**

- age : Feature : Integer
  - workclass : Feature : Categorical
  - fnlwgt : Feature : Integer
  - education : Feature : Categorical
  - education-num : Feature : Integer
  - marital-status : Feature : Categorical
  - occupation : Feature : Categorical
  - relationship : Feature : Categorical
-

# Unit 4 : Define Data Object

4

- race : Feature : Categorical
- sex : Feature : Binary
- capital-gain : Feature : Integer
- capital-loss : Feature : Integer
- hours-per-week : Feature : Integer
- native-country : Feature : Categorical
- income : Target : Binary



```
df.head()
```

# Unit 4 : Define Data Object

5

## Menentukan records (row)

Menampilkan baris dataset terbatas :

- age : 39
  - workclass : State-gov
  - fnlwgt : 77516
  - education : Bachelors
  - education-num : 13
  - marital-status : Never-married
  - occupation : Adm-clerical
-

# Unit 4 : Define Data Object

- relationship : Not-in-family
  - race : White
  - sex : Male
  - capital-gain : 2174
  - capital-loss : 0
  - hours-per-week : 40
  - native-country : United-States
  - income : <=50K
-