

Lab 5 AI / ML

IT20038496

Ranasinghe M.H.

1. Choose a data set from the given supervised learning datasets.

- Adult Census Income Binary Classification Dataset

2. Explore the dataset using the exploration method shown above.

a. For each feature in the dataset visualize and check the statistical data.

b. Based on the Analysis and the statistical data,

i. What features do you think will be the most appropriate for the Model Training?

- Sex
- Race
- Native Country
- Income
- Work class
- Education
- Marital status
- Income
- Relationship

ii. What features should be cleaned?

- Work class
- Occupation
- Native Country

Apply a suitable technique (Drop row, drop column, replace missing values)

- Replace the missing values.

iii. What features should be encoded?

- Sex
- Occupation
- Race
- Marital Status

iv. What features should be scaled?

- Age
- Income