Assignment #06

Hope To Skills

Free Artificial Intelligence Course

Instructor: Irfan Malik, Dr. Sheraz

Submission:

- Make a Google Collab notebook to implement this assignment.
- In case you face difficulty in creating the Google Collab Notebook Follow these **Steps**
- Submit a **.ipynb** file names as **HTS_Assignment_06.ipynb**
- Deadline for this Assignment is **Tuesday 11-07-2023**
- Also mention your name in the Assignment.
- Make Submission in the **Assignment-06** Google Form and press the submit button.
- To visit submission portal click here
- To download the dataset, click here

Solve the Following Task

Task 1: Exploratory Data Analysis (EDA)

Objective: Perform initial data exploration to gain insights and understand the Titanic dataset.

1. Load the Titanic dataset into a Google Collab notebook.
2. Display the first 5 rows of the dataset to examine the structure and features.
3. Display the last 5 rows of the dataset to examine the structure and features
4. Check the dimensions of the dataset (number of rows and columns).
5. Identify the data types of each feature.
6. Explore categorical features by counting unique values in each category.
7. Investigate missing values in the dataset and determine the percentage of missing values for each feature.

Task 2: Handling Missing Values

Objective: Implement strategies to handle missing values in the Titanic dataset.

1. Check the dimension of dataset using the **.shape()** function.
2. Determine the percentage of missing values for each feature.
3. Remove the missing values from the dataset.
4. Again, check the number of missing values in dataset.

5. Check for duplicate values in the dataset and display the number of duplicate rows
6. Remove the duplicate rows from the dataset
7. Again, display the number of rows of data frame to verify the duplicate values are removed

Task 3: Descriptive Statistics

Objective: Calculate and interpret descriptive statistics to gain insights into the Titanic dataset.

1. Calculate basic summary statistics (mean, median, min, max, etc.) for the numerical features. Using **.describe()**
2. Calculate the basic summary statistics (mean, median, min, max, etc.) for the numerical features. Using **.quantile()**
3. Calculate the correlation matrix for all numerical features.
4. Visualize the correlation matrix using a heatmap.