AI Assignment 2

TITANIC PASSENGER SURVIVAL PREDICTION SYSTEM – MACHINE LEARNING CYCLE

By

Team IOIT

AI Assignment 2

AI 2024 Online Summer Internship
Make Your Place in Top 20% AI Engineers of the World

Titanic Passenger Survival Prediction System
(Machine Learning Cycle)

Quick Recap – AI Assignment 1

AI Assignment 1 focussed on **Left Hand Side (LHS)**, i.e., Data, of the following Equation:

Data = Model + Error

The aim of Assignment 1 was to understand the basics of Python Programming Language and Pandas Library, so that you should become familiar with the basic techniques needed to read, understand, and pre-process different types of Data for building Machine Learning Models

Aim

The focus of this Assignment is on **Right Hand Side (RHS)** of the Equation i.e., **Model + Error**

The aims of this Assignment are:

- To introduce how to develop Machine Learning based Real-world Applications using Python, Pandas and Scikit-learn
- To introduce how we can develop Machine Learning Models by using the Data in which **FEATURES** are **ALREADY EXTRACTED**

- **To introduce how we can execute the Machine Learning Cycle (Training Phase, Testing Phase, Application Phase and Feedback Phase) to develop Real-world Applications**

Learning Outcomes

After reading, understanding, analyzing and doing this Assignment, In Sha Allah, you will learn:

- **How to use Python, Pandas and Scikit-learn to build Machine Learning based Real-world Application?**
- **How to execute Machine Learning Cycle (Training Phase, Testing Phase, Application Phase and Feedback Phase) to develop Real-world Applications using the Data in which FEATURES are ALREADY EXTRACTED?**

Supporting Material

Please read the following two Lectures, before you run the code files

- **Lecture 1 - Treating Titanic Passenger Survival Prediction Problem as a Machine Learning Problem using Train-Test Split Approach**
- **Lecture 2 - Developing a System for Titanic Passenger Survival Prediction using Train-Test Split Approach**
- **See Supporting Material Folders**
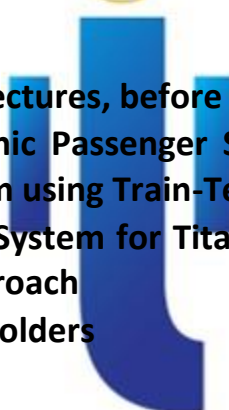
Important Note – Update Libraries

- **Python and other Libraries such as Pandas, Scikit-learn, numpy, scipy etc. are frequently updated. It is important to know how to install, uninstall and upgrade these Libraries to run your Code without any Error** 😊

Tasks

Note that this Assignment is planned, designed, and developed using a HalfCooked Approach i.e., some work is done by the instructor and the remaining

will be done by you. Since the Assignment is to develop Machine Learning based Realworld Applications using Python, Pandas and Scikit-learn. Following the HalfCooked Approach, Instructor has provided with the Code, Data, and Documentation, for the Project titled: Titanic Passenger Survival Prediction System. Note that the Titanic Project is developed using Train-Test Split Approach. Your job is to carry out the Tasks given below and answer the associated questions.

Your main job is to Code, Document and Develop Prototype System for the TODO Task, i.e., Heart Disease Prediction System, given in the documentation of Titanic Passenger Survival Prediction Project by performing the following Tasks.

Important Note: There are total 14 Input Attributes in Heart Disease Dataset. You must select 4 most discriminating Input Attributes for this Assignment.

- Task 1
  - Read the following Lecturs to understand the Titanic Passenger Survival Prediction System
    - Lecture 1 - Treating Titanic Passenger Survival Prediction Problem as a Machine Learning Problem using Train-Test Split Approach
    - Lecture 2 - Developing a System for Titanic Passenger Survival Prediction using Train-Test Split Approach
- Task 2
  - Run the Code for Titanic Passenger Survival Prediction System using Train-Test Split Approach
- Task 3
  - Edit the Code given for Titanic Passenger Survival Prediction System so that it works for Heart Disease Prediction System

Guidelines for Assignment Submission

Deadline: 21-Jul-2024 (23:59 PST)

How to Submit?
Upload your Tutorial (Assignment Solution) on Microsoft Teams. Upload the following files:
1. Jupyter Notebook File
2. Jupyter Notebook in HTML Format
3. Dataset(s)