

# Assignment #05

## Hope to Skills

### Free Artificial Intelligence Advance 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 detailing all the information. No other format will be accepted
- Submission file should be named as **Assignment\_05\_StudentName.ipynb**
- Deadline for this Assignment is **Friday 07-06-2024**.
- Strictly follow the submission deadline.
- Make Submission in the **Assignment-05** Google Form and press the submit button.
- Click [here](#) to submit the Assignment

#### What you will learn

- How to create the google Collab NoteBook from Scratch.
- Using the basic built in functions and apply the following.

#### Solve the Following Task

**Question 1: Use the House Prices dataset from Kaggle to build a linear regression model to predict house prices. (Marks 20)**

- Load the dataset
- Preprocess the dataset
- Build and train linear regression model
- Evaluate its performance using Mean Square Error (MSE)

**Instructions:** You can download the dataset from [here](#)

**Question 2: Use the Iris dataset from Kaggle to build a K-Nearest Neighbors (KNN) classifier to classify iris species. (Marks 20)**

- Load the dataset
- Preprocess the dataset
- Build and train KNN classifier
- Evaluate its performance using Accuracy and Confusion Matrix

**Instructions:** You can download the dataset from [here](#)

**Question 3: Use the Diabetes dataset from Kaggle to build a decision tree regressor to predict diabetes progression. Train the model and evaluate its performance using Mean Squared Error (MSE). (Marks 20)**

- Load the dataset
- Preprocess the dataset
- Build and train Decision Tree Regressor
- Evaluate its performance using Mean Square Error (MSE)

**Instructions:** You can download the dataset from [here](#)

**Question 4: Use the Titanic dataset from Kaggle to build a logistic regression model to predict whether a passenger survived or not. Train the model and evaluate its performance using accuracy and confusion matrix. (Marks 20)**

- Load the dataset
- Preprocess the dataset
- Build and train Logistic Regression Model
- Evaluate its performance using Accuracy and Confusion Matrix

**Instructions:** You download the dataset from [here](#)

**Question 5: Use "Heart Disease Dataset | Cleaned" dataset to build a classification model to predict heart disease. (Marks 20)**

**Load the dataset:**

Preprocess the dataset Include necessary data preprocessing steps, such as handling missing values, encoding categorical variables, and scaling features.

Identify and separate the target variable from the dataset. The target variable indicates the presence or absence of heart disease.

Split the dataset into training and testing sets (e.g., 80% training, 20% testing).

**Train and evaluate the following algorithms:**

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Gradient Boosting Classifier
- Support Vector Machine (SVM)

**For each algorithm, calculate and compare the following metrics:**

- Gini Coefficient (use AUC score for binary classification)
- Confusion Matrix
- Accuracy
- Precision
- Recall
- F1 Score
- AUC-ROC Curve

**Report Findings:**

Summarize the performance of each model. Discuss which model performed best based on the evaluation metrics.

**Instructions:** You can download the dataset from [here](#)