Challenge 2 (Summary Report)

Name: **Anwai Archit**

Date of Submission**: 29th March 2021**

Subject: **Machine and Deep Learning (Medical Imaging and Applications)**

Requisite Points of Concern:

1. Dataset – Titanic Disaster Dataset.
2. Using Scikit-Learn Library.
3. Training and Testing the Classifier Model and Studying the Features and Applications of:
    a. Feature Engineering (Handling Missing Values, Finding Meaningful Features, Encoding Categorical Features).
    b. Stratified 10-Fold Cross Validation
4. Highest Accuracy Reached – **89.88%**

❖ *Summary of the Techniques Used*:

1. Removing features without meaningful values (*dataset.drop*).
2. Imputation for 'Embarked' Column's missing values (*SimpleImputer*); Encoding Categorical Features for 'Embarked' Column (*OrdinalEncoder*).
3. Encoding Categorical Features for 'Sex' Column (*OrdinalEncoder*).
4. Imputation for entire "Features" dataset's missing values – secret target: 'Age' Column (*KNNImputer*).
5. Parameters for Classifier with Stratified 10-Fold Cross Validation (*StratifiedKFold*).
6. Standardization (*StandardScaler*) – scaling the data to obtain better accuracy (for some classifiers).
7. Classifiers Used:
    - KNN (*KNeighborsClassifier*) – 74.15% (without scaling); 86.51% (with scaling).
    - Naïve Bayes (*GaussianNB*) – 83.14%
    - Decision Tree (*DecisionTreeClassifier*) – 85.39%
    - Logistic Regression (*LogisticRegression*) – 84.26%

Model Presented: - Feature Engineering for Handling the Missing Values using SimpleImputer, OrdinalEncoder, & KNNImputer, Stratified 10-Fold Cross Validation and training a RandomForestClassifier and getting an accuracy of *89.88%* on the test set.

Silly Note: Importing the dataset gave me quite bizarre 'Fare' values (e.g. 138.15.00), that I initially handled whilst dallying, but when skimming the actual dataset, I came across the blunder (due to unidentified mistakes). Nevertheless, learnt to find strings in a column, essentials of reshaping, and how imputers fail when the dataset itself is wrong.