|  |  |
| --- | --- |
| Name Of The Student | Vaishnavi G |
| Internship Project Topic | Build a Classification Model for Drug Trials Dataset |
| Name of the Organization | TCS iON |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | SRM Institute of Science and Technology |

|  |  |  |
| --- | --- | --- |
| Date | Day # | Hours Spent |
| 23/11/2022 | 43 | 5 hours |
| Activities done during the day:  **Project Hands on - Split data into train and test**  **Split your data into train and test:**  The train-test split is a technique for evaluating the performance of a machine learning algorithm.  It can be used for classification or regression problems and can be used for any supervised learning algorithm.   * The procedure involves taking a dataset and dividing it into two subsets. * The first subset is used to fit the model and is referred to as the training dataset. * The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. * This second dataset is referred to as the test dataset.   **Train Dataset**: Used to fit the machine learning model.  **Test Dataset:** Used to evaluate the fit machine learning model.  The objective is to estimate the performance of the machine learning model on new data: data not used to train the model.  The train-test procedure is appropriate when there is a sufficiently large dataset available.   |  | | --- | | scaler = StandardScaler()  df = scaler.fit\_transform(df)  df |   **StandardScaler() function:**  Python sklearn library offers us with StandardScaler() function to standardize the data values into a standard format.  **fit\_transform()**  fit\_transform() is used on the training data so that we can scale the training data and also learn the scaling parameters of that data. Here, the model built by us will learn the mean and variance of the features of the training set. These learned parameters are then used to scale our test data.  **df** - A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns.  **Code:**   |  | | --- | |  |   **Output:**   |  | | --- | |  | | | |