|  |  |
| --- | --- |
| 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 |
| 28/10/2022 | 17 | 5 hours |
| Activities done during the day:  Create training and testing dataset in python  **What is train test split ?**   * The train-test split is used to estimate the performance of machine learning algorithms that are applicable for prediction-based Algorithms/Applications. * This method is a fast and easy procedure to perform such that we can compare our own machine learning model results to machine results. * By default, the Test set is split into 30 % of actual data and the training set is split into 70% of the actual data.   **Prerequisites for Train and Test Data**   * We will need the following Python libraries for this tutorial- pandas and sklearn.We can install these with pip- commands  |  | | --- | | pip install pandas  pip install sklearn |  * We use pandas to import the dataset and sklearn to perform the splitting. You can import these packages as-  |  | | --- | | import pandas as pd  from sklearn.model\_selection import train\_test\_split  from sklearn.datasets import load\_iris |   **How to Split Train and Test Set in Python Machine Learning?**   1. Loading the Dataset   Let’s load the csv dataset using pandas.   |  | | --- | | data=pd.read\_csv('forestfires.csv')  data.head() |  1. Splitting   Let’s split this data into labels and features. Now, what’s that? Using features, we predict labels. I mean using features (the data we use to predict labels), we predict labels (the data we want to predict).   |  | | --- | | y=data.temp  x=data.drop('temp',axis=1) |   Temp is a label to predict temperatures in y; we use the drop() function to take all other data in x. Then, we split the data.   |  | | --- | | x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)  x\_train.head() |  1. Plotting of Train and Test Set in Python   We fit our model on the train data to make predictions on it. Let’s import the linear\_model from sklearn, apply linear regression to the dataset, and plot the results.   |  | | --- | | from sklearn.linear\_model import LinearRegression as lm  model=lm().fit(x\_train,y\_train)  predictions=model.predict(x\_test)  import matplotlib.pyplot as plt  plt.scatter(y\_test,predictions) |   Reference: <https://www.geeksforgeeks.org/how-to-split-a-dataset-into-train-and-test-sets-using-python/>  <https://machinelearningmastery.com/train-to-the-test-set-in-machine-learning/> | | |