|  |  |
| --- | --- |
| 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 |
| 17/11/2022 | 37 | 5 hours |
| Activities done during the day:  **Project Handson - Import related modules and dataset file**  Upload the dataset in colab.    **Import required libraries in python:**   |  | | --- | | **import pandas as pd** |   pandas (all lowercase) is a popular Python-based data analysis toolkit which can be imported using import pandas as pd.  It presents a diverse range of utilities, ranging from parsing multiple file formats to converting an entire data table into a NumPy matrix array.  This makes pandas a trusted ally in data science and machine learning.  **Impor**t = “Bring this functionality or library to my python script”  **Pandas** = The library you want to import, in this case, it’s pandas  **As** = The python nomenclature for creating as alias. This is a fancy way of taking a long word and referencing it as a short word  **pd** = The standard short name for referencing pandas   |  | | --- | | **import numpy as np** |   NumPy, which stands for Numerical Python, is a scientific computing library built on top of the Python programming language.  The import numpy portion of the code tells Python to bring the NumPy library into your current environment.  The as np portion of the code then tells Python to give NumPy the alias of np. This allows you to use NumPy functions by simply typing np.function\_name rather than numpy.function\_name.   |  | | --- | | **from sklearn.svm import SVC** |   Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.  It is C-support vector classification whose implementation is based on libsvm. The module used by scikit-learn is sklearn.svm.SVC. This class handles the multiclass support according to one-vs-one scheme.   |  | | --- | | **from sklearn.ensemble import RandomForestClassifier** |   A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting.   |  | | --- | | **from sklearn.model\_selection import train\_test\_split** |   The train\_test\_split function of the sklearn.model\_selection package in Python splits arrays or matrices into random subsets for train and test data, respectively.  To use the train\_test\_split function, we’ll import it into our program   |  | | --- | | **from sklearn.preprocessing import StandardScaler** |   The StandardScaler function of sklearn is based on the theory that the dataset's variables whose values lie in different ranges do not have an equal contribution to the model's fit parameters and training function and may even lead to bias in the predictions made with that model.  Therefore, before including the features in the machine learning model.   |  | | --- | | **from sklearn.metrics import classification\_report** |   A classification report is a performance evaluation metric in machine learning. It is used to show the precision, recall, F1 Score, and support of your trained classification model.   |  | | --- | | **import matplotlib.pyplot as plt** |   matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB.  Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.  In matplotlib.pyplot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area, and the plotting functions are directed to the current axes  **Code:**   |  | | --- | |  |   **Output:**   |  | | --- | |  | | | |