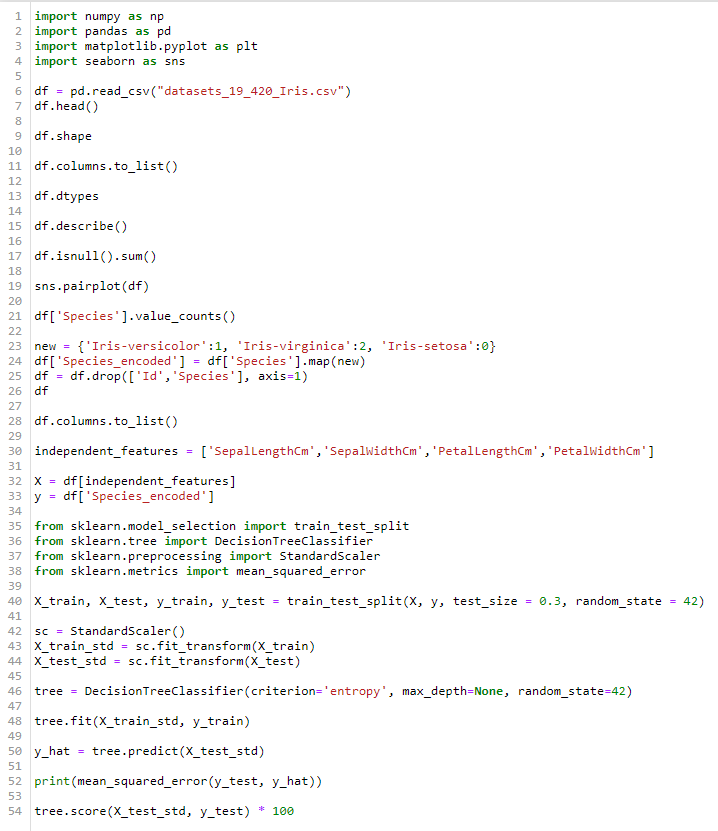
**Assignment-4 Documentation**

Submitted by : **Anurag Singh**

**Code:**

****

**Explanation:**

1. Imported all the required libraries (numpy, pandas, matplotlib and seaborn).
2. Read the dataset using ‘read\_csv()’ function of pandas library and stored it in dataframe ‘df’.
3. Displayed the first 5 rows of the dataset using ‘head()’ function of pandas library.
4. Checked the insights of the dataset using several functions and attributes like ‘shape’, ‘columns’, ‘dtypes’, ‘describe()’.
5. Checked for any null values using ‘isnull()’ function and found that there were no null values in the dataset.
6. Visualized the dataset using ‘pairplot()’ function present in seaborn library.
7. Now, Machine Learning models only take numerical values for model creation, so I did ordinal encoding in ‘Species’ column, which was dependent column and was of type ‘object’ which is basically a string.
8. After this, I dropped unwanted columns like ‘Id’ and ‘Species’.
9. Divided the dataset into Independent(‘X’) and Dependent(‘y’) features for model making.
10. Imported all the libraries for Machine Learning algorithms.
11. Here ‘train\_test\_split’ is for splitting the dataset into training and testing part.
12. ‘DecisionTreeClassifier’ is the algorithm which we will use for classification as it is a classification problem.
13. ‘StandardScalar’ is used because when I plotted ‘pairplot’, I found that independent features were not in the same scale, so I did feature scaling using this ‘StandardScalar’.
14. After this, I splitted the dataset into training part(70%) and testing part(30%).
15. Performed feature scaling using StandardScalar.
16. Created an object of DecisionTreeClassifier.
17. Trained the dataset using ’fit()’ function.
18. Predicted the output by the model using ‘predict()’ function.
19. Finally, checked for the error percentage and accuracy using ‘mean\_squared\_error()’ and ‘score()’ functions respectively.

**Result:**

**Error Percentage : 0.04**

**Accuracy Percentage : 95.5**