**Regression Assignment:**

**1.Identify your problem statement :**

This problem comes under Supervised-Regression because Input and Output is clear.

Single Linerar Regression is not Applicable because we have more the one inputs.

**2.Tell basic info about the dataset (Total number of rows, columns)**

Total number of rows = 1338

Total number of columns = 6

Input Variable = age, sex, bmi,children, smoker

Output Variable = charges

**3. Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)**

**Two column have nominal data sex and smoker**

**So we change the categorical data into numerical value**

**4.Select the model:**

**1. Multi Linear Regression:**

**R2 Score = 0.789479**

**2. Support Vector Machine:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.NO** | **C Value** | **Linear** | **Rbf** | **poly** | **sigmoid** |
| 1. | 10 | 0.462468 | -0.03227 | 0.03871 | 0.03930 |
| 2. | 100 | 0.628879 | 0.320031 | 0.61795 | 0.52761 |
| 3. | 1000 | 0.764931 | 0.810206 | 0.85663 | 0.28747 |
| 4. | 2000 | 0.744041 | 0.854776 | 0.860559 | -0.593950 |
| 5. | 3000 | 0.741423 | 0.866339 | 0.859893 | -2.124419 |
| 6. | 5000 | 0.741417 | 0.874781 | 0.85956 |  |
| 7. | 7000 | 0.741422 | 0.877692 | 0.85966 |  |
| 8. | 10000 | 0.7414230 | 0.877995 | 0.85917 |  |

**The highest R2 Score in support vector Machine**

**Is 0.877995 to the parameter ‘rbf’,C=10000**

**3. DecisTreeRegression:**

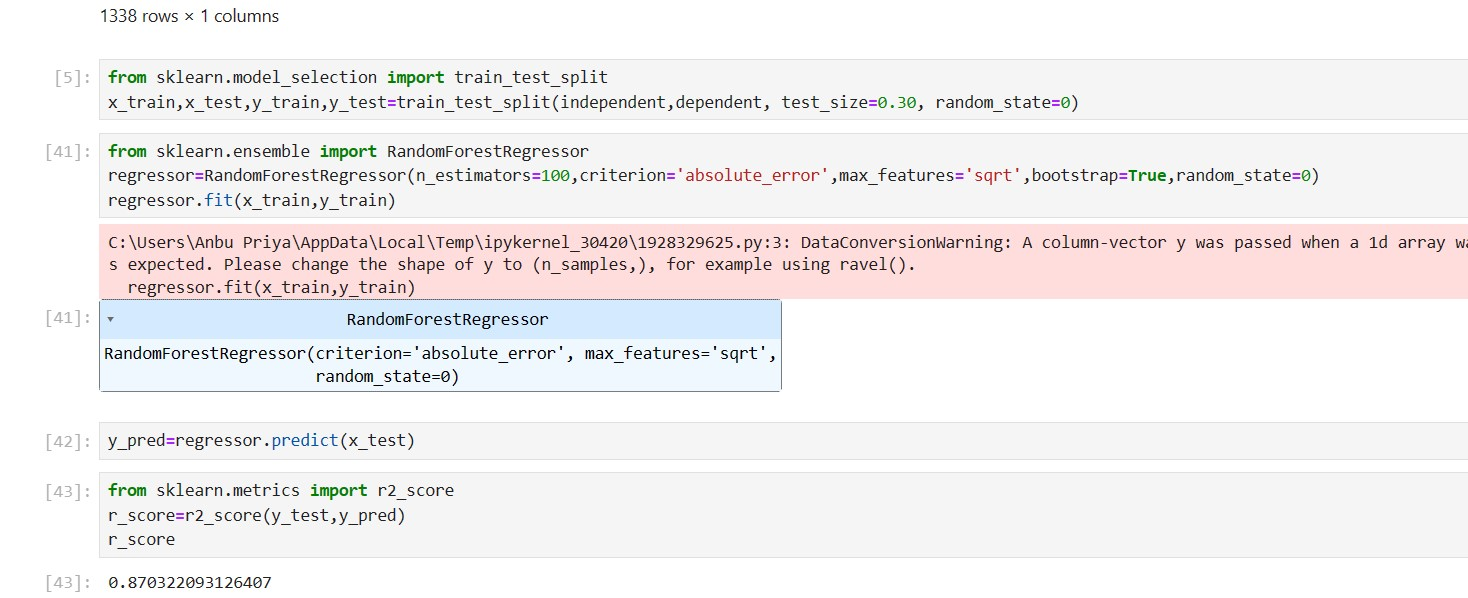
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Criterion** | **splitter** | **Max-features** | **R2 value** |
| **1.** | **MSE** | **Best** | **Auto** | **0.695330** |
| **2.** | **MSE** | **Best** | **Sqrt** | **0.574614** |
| **3.** | **MSE** | **Best** | **Log2** | **0.715050** |
| **4.** | **MSE** | **Random** | **Auto** | **0.656921** |
| **5.** | **MSE** | **Random** | **Sqrt** | **0.641544** |
| **6.** | **MSE** | **Random** | **Log2** | **0.672530** |
| **7.** | **MAE** | **Best** | **Auto** | **0.655440** |
| **8.** | **MAE** | **Best** | **Sqrt** | **0.639987** |
| **9.** | **MAE** | **Best** | **Log2** | **0.769374** |
| **10.** | **MAE** | **Random** | **Auto** | **0.744232** |
| **11.** | **MAE** | **Random** | **Sqrt** | **0.646634** |
| **12.** | **MAE** | **Random** | **Log2** | **0.746351** |
| **13.** | ***Friedman\_mse*** | **Best** | **Auto** | **0.699437** |
| **14.** | ***Friedman\_mse*** | **Best** | **Sqrt** | **0.700844** |
| **15.** | ***Friedman\_mse*** | **Best** | **Log2** | **0.720291** |
| **16.** | ***Friedman\_mse*** | **Random** | **Auto** | **0.723718** |
| **17.** | ***Friedman\_mse*** | **Random** | **Sqrt** | **0.665577** |
| **18.** | ***Friedman\_mse*** | **Random** | **Log2** | **0.707787** |

**The highest R2 Score in Decision Tree Regression**

**Is 0.769374 to the parameter absolute\_error,**

**’Best’,‘Log2’.**

**4. RainForestRegression:**



**Conclusion:**

**Among these types of regresstion R2 value highest in both SupportVectorMachine the parameters are ‘rbf’, C=10000 and Rain Forest regression gave the same result.**