To find following the machine learning regression method using in r2 value

**1.DecisionTreeRegressor :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.no** | **Criterion** | **Max Features** | **Splitter** | **R2\_Value** |
| 1 | friedman\_mse | Sqrt | best | 0.6735 |
| 2 | friedman\_mse | sqrt | random | 0.4633 |
| 3 | friedman\_mse | Log2 | best | 0.4801 |
| 4 | friedman\_mse | Log2 | random | 0.8886 |
| 5 | friedman\_mse | none | best | 0.9076 |
| 6 | friedman\_mse | none | random | 0.9467 |
| 7 | squared\_error | sqrt | best | 0.9385 |
| 8 | squared\_error | sqrt | random | 0.4933 |
| 9 | squared\_error | Log2 | best | 0.7767 |
| 10 | squared\_error | Log2 | random | 0.4268 |
| 11 | squared\_error | none | best | 0.9072 |
| 12 | squared\_error | none | random | 0.8343 |
| 13 | absolute\_error | sqrt | best | 0.5876 |
| 14 | absolute\_error | sqrt | random | 0.3069 |
| 15 | absolute\_error | Log2 | best | 0.5727 |
| 16 | absolute\_error | Log2 | random | 0.3937 |
| 17 | absolute\_error | none | best | 0.9686 |
| 18 | absolute\_error | none | random | 0.8119 |
| 19 | Poisson | sqrt | best | 0.5582 |
| 20 | Poisson | sqrt | random | 0.6593 |
| 21 | Poisson | Log2 | best | 0.2618 |
| 22 | Poisson | Log2 | random | 0.5964 |
| 23 | Poisson | none | best | 0.9268 |
| 24 | Poisson | none | random | 0.9248 |

The **DecisionTreeRegressor R2\_Value** by using following hyper parameters

**Criterion** : absolute\_error , **Max Features** : None ,**Splitter** : best ,**R2\_Value : 0.9689.**

**2.Support Vector Regression :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.no** | **Hyper**  **parameters** | **Linear**  **(r2\_value)** | **RBF - non linear**  **(r2\_value)** | **Poly**  **(r2\_value)** | **Sigmoid**  **(r2\_value)** |
| 1 | C=10 | -0.0396 | -0.0568 | -0.0536 | -0.0547 |
| 2 | C=100 | 0.1064 | -0.0507 | -0.0198 | -0.0304 |
| 3 | C=500 | 0.5928 | -0.0243 | 0.1146 | 0.0705 |
| 4 | C=1000 | 0.7802 | 0.0067 | 0.2661 | 0.1850 |
| 5 | C=2000 | 0.8767 | 0.0675 | 0.4810 | 0.3970 |
| 6 | C=3000 | 0.8956 | 0.1232 | 0.6370 | 0.5913 |

The **Support Vector Regression R2\_Value** by using following hyper parameters

**C=3000 , kernel = linear ,r2\_value=0.8956.**

**3. Regression Assignment R2\_Value :**

**1)Identity the problem statement :**

Here we need to predict “Insurance Charges” ,it comes under numbers so next step need to proceed with Regression.

**2)Basic info about the “Dataset” :**

Total no. of columns = 6, Total no. of rows =1338.

**3)Mention preprocessing method if you used to the model – nominal data.**

Here I used “one hot encoding” because in the dataset has sex column and smoker column was present categorical data.

So I used one hot encoding to convert categorical data into nominal data.(it helps read the data from table i.e.dataset)

**4)Develope the good model with r2\_value by using any machine learning algorithms.**

I got r2\_value is 0.8720 by using machine learning RandomForestAlgorithm with the hyper parameter are following

N\_estimators = 100,criterion = absolute\_error,max\_features = log2 .

**5)All the research r\_score values are showing below tabulation.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.no** | **N\_estimators** | **Criterion** | **Max\_features** | **R2\_Value** |
| 1 | 50 | squared\_error | Sqrt | 0.8699 |
| 2 | 100 | squared\_error | sqrt | 0.8710 |
| 3 | 50 | squared\_error | log2 | 0.8708 |
| 4 | 100 | squared\_error | log2 | 0.8706 |
| 5 | 50 | squared\_error | None | 0.8528 |
| 6 | 100 | squared\_error | None | 0.8552 |
| 7 | 50 | friedman\_mse | Sqrt | 0.8675 |
| 8 | 100 | friedman\_mse | sqrt | 0.8688 |
| 9 | 50 | friedman\_mse | log2 | 0.8688 |
| 10 | 100 | friedman\_mse | log2 | 0.8688 |
| 11 | 50 | friedman\_mse | None | 0.8534 |
| 12 | 100 | friedman\_mse | None | 0.8490 |
| 13 | 50 | absolute\_error | Sqrt | 0.8728 |
| 14 | 100 | absolute\_error | sqrt | 0.8712 |
| 15 | 50 | absolute\_error | log2 | 0.8670 |
| 16 | 100 | absolute\_error | log2 | 0.8720 |
| 17 | 50 | absolute\_error | None | 0.8565 |
| 18 | 100 | absolute\_error | None | 0.8538 |
| Th19 | 50 | poisson | Sqrt | 0.8684 |
| 20 | 100 | poisson | sqrt | 0.8698 |
| 21 | 50 | poisson | log2 | 0.8701 |
| 22 | 100 | poisson | log2 | 0.8711 |
| 23 | 50 | poisson | None | 0.8498 |
| 24 | 100 | poisson | None | 0.8494 |

The **RandomForestRegression(Regression Assignment) r2\_value = 0.8720** by using hyper parameters are following

N\_estimators = 100,criterion = absolute\_error,max\_features = log2

**6)Mention** **your final model and justify.**

In this RandomForest model I researched with all hyper parameters and finally get the r\_score value was 0.8720.

It is also not much better model because I got 0.87 but while comparing with other r\_score value it is good model.

**4.RandomForest R2\_Value :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.no** | **n\_estimators** | **Criterion** | **Max\_features** | **R2\_Value** |
| 1 | 50 | squared\_error | Sqrt | 0.7656 |
| 2 | 100 | squared\_error | sqrt | 0.8115 |
| 3 | 50 | squared\_error | Log2 | 0.8023 |
| 4 | 100 | squared\_error | Log2 | 0.7972 |
| 5 | 50 | squared\_error | None | 0.9427 |
| 6 | 100 | squared\_error | None | 0.9376 |
| 7 | 50 | absolute\_error | Sqrt | 0.8143 |
| 8 | 100 | absolute\_error | sqrt | 0.7575 |
| 9 | 50 | absolute\_error | Log2 | 0.7298 |
| 10 | 100 | absolute\_error | Log2 | 0.8066 |
| 11 | 50 | absolute\_error | None | 0.9439 |
| 12 | 100 | absolute\_error | None | 0.9426 |
| 13 | 50 | friedman\_mse | Sqrt | 0.7844 |
| 14 | 100 | friedman\_mse | sqrt | 0.8447 |
| 15 | 50 | friedman\_mse | Log2 | 0.7944 |
| 16 | 100 | friedman\_mse | Log2 | 0.8319 |
| 17 | 50 | friedman\_mse | None | 0.9377 |
| 18 | 100 | friedman\_mse | None | 0.9303 |
| 19 | 50 | poisson | Sqrt | 0.8091 |
| 20 | 100 | poisson | sqrt | 0.7385 |
| 21 | 50 | poisson | Log2 | 0.7585 |
| 22 | 100 | poisson | Log2 | 0.8230 |
| 23 | 50 | poisson | None | 0.9354 |
| 24 | 100 | poisson | None | 0.9320 |

The **RandomForestRegression r2\_value = 0.9439** by using hyper parameters are following

N\_estimators = 50,criterion = absolute\_error,max\_features=None.

**5.Multiple Linear Regression :**

r\_score value is 0.9358,here r\_score is much better so we can proceed next step of

deployment phase. Hence our model will perform well then the end users feels good vibe.

In this multiple linear regression model has used “one hot encoding” to convert the categorical data into nominal data.