**Problem Statement or Requirement:**

A client’s requirement is, he wants to predict the insurance charges based on

the several parameters. The Client has provided the dataset of the same.

**Data set Analysis:**

1.It contains 6 columns : Age, Sex, bmi, children, smoker and charges

2. Dataset contains 1338 rows

3. I could see 2 nominal columns : sex and smoker which has to be converted with the help of one hot encoding

4. Domain Selection : Machine Learning

5. Learning Selection : Supervised Learning

6. Method : Regression

Since the data involves multiple inputs, here we can’t use Simple Linear Regression. So lets start with other available algorithms.

**1.MLR – Multiple Linear Regression**

**R2\_score value : 0.7894**

****

**2. Support Vector Machine – SVM:**

**R2 score value : -0.0765**

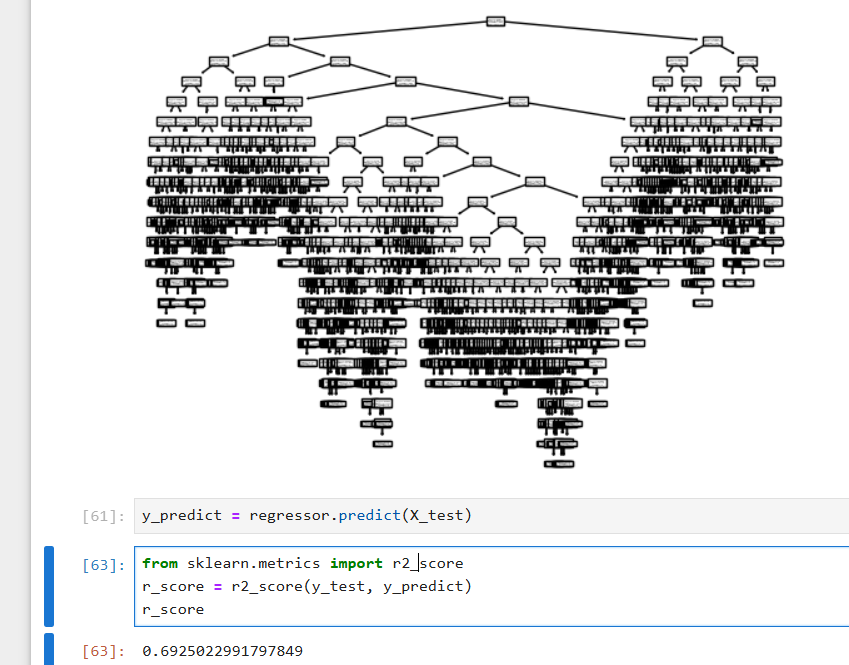
**3. Decision Tree:**

**R2 score value (without parameter tuning) : 0.6882844096334517**

**With Parameter turning:**

|  |  |
| --- | --- |
| **criterion='squared\_error', splitter='best'** | **0.6950003678214174** |
| **criterion='friedman\_mse', splitter='best'** | **0.6819201895420186** |
| **criterion='absolute\_error', splitter='best'** | **0.69172581764679** |
| **criterion='poisson', splitter='best'** | **0.7318777565761565** |
| **criterion='poisson', splitter='best'** | **0.6925022991797849** |

**So we can take that one as the best score of this model**

****

**4. Random Forest:**

|  |  |
| --- | --- |
| ***n\_estimators=50, criterion='squared\_error', random\_state=0*** | **0.8498329315421834** |
| **n\_estimators=100, criterion='squared\_error', random\_state=0** | **0.8538307913484513** |
| **n\_estimators=50, criterion='absolute\_error', random\_state=0** | **0.8526655993519747** |
| **n\_estimators=100, criterion='absolute\_error', random\_state=0** | **0.8520093621081837** |
| **n\_estimators=50, criterion='friedman\_mse', random\_state=0** | **0.8500716139332296** |
| **n\_estimators=100, criterion='friedman\_mse', random\_state=0** | **0.8540518935149612** |
| **n\_estimators=50, criterion='poisson', random\_state=0** | **0.8491075958392151** |
| **n\_estimators=100, criterion='poisson', random\_state=0** | **0.8526334258892607** |

****