**Assignment-Regression Algorithm  
  
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. As a data scientist, you must develop a model which will predict the insurance charges.  
  
1.) Identify your problem statement  
 3 stages of problem identification method is**

**Machine Learning - ( so far It’s a Number data)  
Supervised Learning -( Both input and Output is very clear)  
Regression – (Output is numerical data)**

**2)Tell basic info about the dataset: (Total number of rows, columns)  
 total number of rows -1338  
 total number of columns -6  
  
3.) Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)  
 one hot encoding  
  
4.)Develop a good model with r2\_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.  
  
Multiple Linear Regression:  
  
 so far Multiple Linear Regression method we found the R2 value is =0.78  
 its not a good module we need to check out another process**

**Support Vector Machine:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl.No** | **Hyper parameter (c)** | **Linear**  **R2** | **Poly**  **R2** | **Rbf**  **R2** | **Sigmoid**  **R2** |
| **1** | **C=10** | **0.46** | **0.03** | **-0.03** | **0.03** |
| **2** | **C=100** | **0.62** | **0.61** | **0.32** | **0.52** |
| **3** | **C=500** | **0.76** | **0.82** | **0.66** | **0.44** |
| **4** | **C=1000** | **0.76** | **0.85** | **0.81** | **0.28** |
| **5** | **C=2000** | **0.74** | **0.86** | **0.85** | **-0.59** |
| **6** | **C=3000** |  | **0.85** | **0.86** |  |
| **7** | **C=4000** |  |  | **0.87** |  |

**The SVM Regression best R2 value is 0.87 using Rbf parameter(C=4000)  
  
Decision Tree Method:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.No** | **criterion** | ***splitter*** | **R2** |
| **1** | **mse** | **best** | **0.70** |
| **2** | **mse** | **random** | **0.71** |
| **3** | ***friedman\_mse*** | **best** | **0.69** |
| **4** | ***friedman\_mse*** | **random** | **0.68** |
| **5** | **mae** | **best** | **0.67** |
| **6** | **mae** | **random** | **0.74** |

**The Decision Forest Regression best R2 value is 0.74 using**

**Criterion = mae(*absolute\_error),* Splitter = Random parameter**

**Random Forest:**

|  |  |  |
| --- | --- | --- |
| **Sl.No** | **n\_estimators** | **R2** |
| **1** | **50** | **0.84** |
| **2** | **100** | **0.85** |
| **3** | **200** | **0.85** |
| **4** | **250** | **0.85** |

**The Decision Forest Regression best R2 value is 0.85 using n\_estimators=200  
  
6.) Mention your final model, justify why u have chosen the same**.

**So far Analys the all regression algorithm we got a best R2 value is 0.87 using Support Vector Machine Algorithm..So we choose the final model is SVM**.