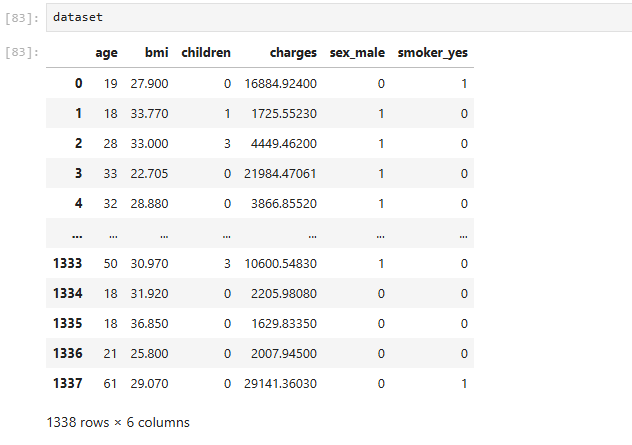
**1.) Identify your problem statement**

Develop a model to predict the insurance charges based on these inputs age, sex, bmi, children& smoker parameters.

**2.) Tell basic info about the dataset (Total number of rows, columns)** = 1338 Rows and 6 Columns



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

As per the dataset, we could see Sex and smoker is categorial column but smoker we can convert from string to int datatype.  
For sex column, we have to use One Hot encoding method to convert string to number (0- Female, Male -1)

Sex in Nominal data because we cant compare.

**4. Model creation using different Algorithms:**

**1. Multiple Linear Regression Algorithm:**

R\_score value =0.7894

**2.Support Vector Machine:**

SVM R2 score = 0.8779

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SNO | Hyper Parameter | Linear (r value) | Rbf(non-linear) | POLY | SIGMOID |
| 1 | C=1.0 | -0.0101 | -0.0833 | -0.0756 | -0.0754 |
| 2 | C=100 | 0.6288 | 0.3200 | 0.6179 | 0.5276 |
| 3 | C=1000 | 0.7649 | 0.8102 | 0.8566 | 0.2874 |
| 4 | C=10000 | 0.7414 | 0.8779 | 0.8591 | -34.1515 |
| 5 | C=100000 | 0.7414 | 0.8724 | 0.8577 | -3465.9535 |

**3.DECISION TREE:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SNO | CRITERION | SPLITTER | R2 Value | Max\_Features | Max\_R2 |
| 1 | ***squared\_error*** | best | 0.6832 |  |  |
| 2 | ***squared\_error*** | random | 0.7601 |  |  |
| 3 | ***squared\_error*** | random |  | sqrt | 0.6119 |
| 4 | ***squared\_error*** | random |  | Log2 | 0.6856 |
| 5 | ***squared\_error*** | best |  | sqrt | 0.6757 |
| 6 | ***squared\_error*** | best |  | Log2 | 0.7136 |
| 7 | ***friedman\_mse*** | best | 0.6944 |  |  |
| 8 | ***friedman\_mse*** | random | 0.6944 |  |  |
| 9 | ***friedman\_mse*** | best |  | Log2 | 0.5647 |
| 10 | ***friedman\_mse*** | random |  | Log2 | 0.6408 |
| 11 | ***friedman\_mse*** | best |  | sqrt | 0.7050 |
| 12 | ***friedman\_mse*** | random |  | sqrt | 0.7126 |
| 13 | ***absolute\_error*** | best | 0.6597 |  |  |
| 14 | ***absolute\_error*** | random | 0.7425 |  |  |
| 15 | ***absolute\_error*** | best |  | sqrt | 0.7352 |
| 16 | ***absolute\_error*** | best |  | Log2 | 0.6754 |
| 17 | ***absolute\_error*** | random |  | sqrt | 0.6943 |
| 18 | ***absolute\_error*** | random |  | Log2 | 0.7744 |
| 19 | ***poisson*** | best | 0.7227 |  |  |
| 20 | ***poisson*** | random | 0.7582 |  |  |
| 21 | ***poisson*** | best |  | sqrt | 0.7352 |
| 22 | ***poisson*** | best |  | Log2 | 0.7015 |
| 23 | ***poisson*** | random |  | sqrt | 0.6556 |
| 24 | ***poisson*** | random |  | Log2 | 0.6401 |

Decision Tree R2 Value (absolute\_error,random,log2) = 0.7744

**5. Random Forest:**

R2 Random Forest value = 0.8498

**6.) Based on the R2 score best model for this dataset is**

SVM R2 score = 0.8779