1.Insurance Premium Data set having 6 Columns(ie’Age’,’Sex’,’Bmi’,’Childeren’,’Smoker’,’Charges’) and 1339 Rows.

2.This problem Statement comes under Regression in Supervised learning.

3.So we have to calculate the Insurance Premium.

4.Age,Sex,Bmi,Children,Smoker Columns are Independent values.

5.Charges Column is a Dependent value

6.For this problem statement we have to use one hot encoding method for Sex and Smoker columns.

Multiple Linear Regression for Insurance Premium:

r\_score=0.78794790349867009

Support Vector Machine:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No | Hyper  Parameter | Linear(r value) | Rbf(non linear)  (r value) | Poly  (r value) | Sigmoid  (r value) |
| 1 | C=2000 | 0.74404183081 | 0.854776642539298 | 0.8605579275773966 | -0.59395097 |
| 2 | C=10 | 0.46246841423398 | -0.03227329390671 | 0.038716222760231456 | 0.093307143 |
| 3 | C=100000 | 0.74141889773252 | 0.872498442949260 | 0.8577881142478652 | -0.34320318 |
| 4 | C=100 | 0.62887928573203 | 0.323003178205083 | 0.617956962405795 | 0.52761035 |
| 5 | C=5000 | 0.74147193017040 | 0.874777817494097 | 0.8595656394397817 | -7.53004323 |
| 6 | C=4000 | 0.74141988030669 | 0.871740786976204 | 0.8600049580588055 | -5.51033354 |

Support Vector Machine Regression for Insurance Premium:

r\_2value=0.874777817494097

Decision Tree:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Criterion | Max\_  Features | Splitters | R\_Value |
| 1 | Mse | Auto | Best | 0.6892290740708633 |
| 2 | Mse | Auto | Random | 0.7312029920693102 |
| 3 | Mse | Sqrt | Best | 0.7544298592026704 |
| 4 | Mse | Sqrt | Random | 0.7345274807370004 |
| 5 | Mse | Log2 | Best | 0.6289290599906876 |
| 6 | Mse | Log2 | Random | 0.6492755230617556 |
| 7 | Mae | Auto | Best | 0.6743439637665865 |
| 8 | Mae | Auto | Random | 0.74933034485887850 |
| 9 | Mae | Sqrt | Best | 0.0.66961864705833243 |
| 10 | Mae | Sqrt | Random | 0.712869737823193667 |
| 11 | Mae | Log2 | Best | 0.77340219656595574305 |
| 12 | Mae | Log2 | Random | 0.671727671418883974 |
| 13 | Friedman\_  Mse | Auto | Best | 0.687166396731769398 |
| 14 | Friedman\_  Mse | Auto | Random | 0.724011369094970645 |
| 15 | Friedman\_  Mse | Sqrt | Best | 0.727485841439698712 |
| 16 | Friedman\_  Mse | Sqrt | Random | 0.695085685346750498 |
| 17 | Friedman\_  Mse | Log2 | Best | 0.746669455185433278 |
| 18 | Friedman\_  Mse | Log2 | Random | 0.676948995681251104 |

Decision Tree Regression for Insurance Premium:

r\_2value=0.77340219656595574

Random Forest:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Criterion | Max\_  Features | N\_Estimator | R Value |
| 1 | Mse | Auto | 10 | 0.8331512465124 |
| 2 | Mse | Auto | 100 | 0.8539235792990 |
| 3 | Mse | Sqrt | 10 | 0.8520006346688 |
| 4 | Mse | Sqrt | 100 | 0.8709953890446 |
| 5 | Mse | Log2 | 10 | 0.8520006346682 |
| 6 | Mse | Log2 | 100 | 0.8709953890446 |
| 7 | Mae | Auto | 10 | 0.8355343533965 |
| 8 | Mae | Auto | 100 | 0.8521468946589 |
| 9 | Mae | Sqrt | 10 | 0.8574913110512 |
| 10 | Mae | Sqrt | 100 | 0.87177215500385 |
| 11 | Mae | Log2 | 10 | 0.85749131012527 |
| 12 | Mae | Log2 | 100 | 0.8747155003857 |

Random Forest Regression for Insurance Premium:

r\_Score value=0.8747155003857

Conclusion:

From the above Algorithms , we can conclude that there is no r\_Score value nearer to 1. So, I think this is not a good model.Tho, I could say Support Vector Machine Algorithm gives the maximum r\_Score value comparing to other Algorithms.So, I saved SVM Algorithm and also I created Deployment Phase for the same.