Insurance Assignment:

Problem Statement:

There is a vast set of useful data from which insurance charge has to be derived for the policy holder. Insurance premium for each policy holder is the output that need to be derived from the input data of age, sex, bmi, smoker and # of children the policy holder has.

Dataset:

The dataset contains 1338 rows and 5 columns. 4 columns as independent variable and 1 dependent variable.

Preprocessing

2 Columns (Smoker and sex) required to changed to number

**2 SUPPORT VECTOR MACHINE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No | HYPER PARAMETER | R value | | | |
|  |  | LINEAR | RBF | POLY | SIGMOID |
| 1 | C10 | 0.11360274238065438 | -0.08234397724561848 | -0.08425151511956219 | -0.09063424892586847 |
| 2 | C100 | 0.5950891903531594 | -0.12324686690902853 | -0.09720622584536942 | -0.11387107273998276 |
| 3 | C500 | 0.6340618038625447 | -0.11949878541835779 | -0.07664981478294308 | -0.34555016141987727 |
| 4 | C1000 | 0.6893270405259424 | -0.10885171648275094 | -0.04359363831727636 | -1.1085765060679802 |
| 5 | C2000 | 0.7650782728760414 | -0.09054647841605568 | 0.021918774776422567 | -3.665322974973816 |
| 6 | C3000 | 0.7649476546367047 | -0.06926034019721072 | 0.08500683507974771 | -7.507776980676335 |

3 **DECISION TREE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S# | CRITERION | MAX FEATURES | SPLITTER | R VALUE |
| 1 | Squared error | Sqrt | Best | 0.6760976763669552 |
| 2 | Squared error | Log2 | Best | 0.7314367016496074 |
| 3 | Squared error | None | Best | 0.698576498892554 |
| 4 | Squared error | None | Random | 0.7049248233408005 |
| 5 | Squared error | Sqrt | Random | 0.6146387408156663 |
| 6 | Squared error | Log2 | Random | 0.6818956019234983 |
| 7 | *friedman\_mse* | Log2 | Random | 0.7313486536947675 |
| 8 | *friedman\_mse* | Sqrt | Random | 0.7058491859317344 |
| 9 | *friedman\_mse* | Sqrt | best | 0.6514641839703965 |
| 10 | *friedman\_mse* | Log2 | best | 0.7725109560260952 |
| 11 | *friedman\_mse* | None | best | 0.6767922600150549 |
| 12 | *friedman\_mse* | None | Random | 0.7094668881744135 |
| 13 | *absolute\_error* | None | random | 0.7623625979962945 |
| 14 | *absolute\_error* | None | Best | 0.6896707392706691 |
| 15 | *absolute\_error* | Sqrt | best | 0.7002863577430392 |
| 16 | *absolute\_error* | Log2 | Best | 0.6945231621562639 |
| 17 | *absolute\_error* | Log2 | random | 0.7372981375699483 |
| 18 | *absolute\_error* | Sqrt | random | 0.7512226340089692 |
| 19 | ***Poisson*** | Sqrt | random | 0.7060585096141188 |
| 20 | ***Poisson*** | Log2 | Random | 0.6722546390781965 |
| 21 | ***Poisson*** | Log2 | Best | 0.7231018673225995 |
| 22 | ***Poisson*** | None | Best | 0.7370705230796333 |
| 23 | ***Poisson*** | None | Random | 0.7209700429653658 |

3 **RANDOM FOREST**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S# | n\_estimator | CRITERION | MAX FEATURES | R VALUE |
| 1 | 100 | Squared error | Sqrt | 0.8567855553013841 |
| 2 | 100 | Squared error | Log2 | 0.8567855553013841 |
| 3 | 100 | Squared error | None | 0.8384200372358014 |
| 4 | 100 | *friedman\_mse* | Log2 | 0.85937719395222 |
| 5 | 100 | *friedman\_mse* | Sqrt | 0.85937719395222 |
| 6 | 100 | *friedman\_mse* | None | 0.8396513408995016 |
| 7 | 100 | *absolute\_error* | None | 0.8337326515925146 |
| 8 | 100 | *absolute\_error* | Sqrt | 0.8462817998180086 |
| 9 | 100 | *absolute\_error* | Log2 | 0.8462817998180086 |
| 10 | 100 | ***Poisson*** | Sqrt | 0.8532447609105054 |
| 11 | 100 | ***Poisson*** | Log2 | 0.8532447609105054 |
| 12 | 100 | ***Poisson*** | None | 0.8312818392895119 |

The model selected is Random Forest with the below parameter as this is having the best R value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S# | n\_estimator | CRITERION | MAX FEATURES | R VALUE |
| 1 | 100 | *friedman\_mse* | Sqrt | 0.85937719395222 |