Problem Statement or Requirement: A client’s requirement is, he wants to predict the insurance charges based on the several parameters.

1. Identify your problem statement

To predict the insurance charges

1. Tell basic info about the dataset (Total number of rows, columns)

Dataset can be predicted by machine learning -supervised learning. It has

5 input column, 1 output column, 1338 rows, 5 column.

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

To convert the categorical column "Smoker" in dataset to a format suitable for machine learning using pandas get\_dummies() method with the drop\_first=True parameter

1. Develop a good model with r2\_score

To find the following the machine learning regression using in r2 value

1.Multiple Linear Regression : 0.7894

2.Support Vector Machine :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **HyperParameter** | **Linear (R2 Value)** | **RBF(Non Linear) (R2 Value)** | **Poly (R2 Value)** |
| 1 | c= 1 | - 0.1116 | - 0.0884 | - 0.06429 |
| 2 | c= 10 | -0.0016 | - 0.0819 | -0.0931 |
| 3 | c=1000 | -0.6340 | -0.1174 | 0.0555 |

3.Decision Tree

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Criterion | Splitter | R2 Value |
| 1 | squared\_error | random | 0.7046 |
| 2 | squared\_error | best | 0.6983 |
| 3 | friedman\_mse | random | 0.6840 |
| 4 | friedman\_mse | best | 0.6881 |
| 5 | absolute\_error | random | 0.7387 |
| 6 | absolute\_error | best | 0.6594 |
| 7 | poisson | random | 0.7142 |
| 8 | poisson | best | 0.6959 |

4.Random Forest

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **n\_estimators** | **Criterion** | **Max Features** | **Random\_state** | **R2 Value** |
| 1 | 50 | squared\_error | auto | 0 | 0.8498 |
| 2 | 50 | squared\_error | sqrt | 0 | 0.8695 |
| 3 | 50 | squared\_error | log2 | 0 | 0.8695 |
| 4 | 50 | squared\_error | auto | 1 | 0.8551 |
| 5 | 50 | squared\_error | sqrt | 1 | 0.8678 |
| 6 | 50 | squared\_error | log2 | 1 | 0.8678 |
| 7 | 50 | friedman\_mse | auto | 0 | 0.8500 |
| 8 | 50 | friedman\_mse | sqrt | 0 | 0.8702 |
| 9 | 50 | friedman\_mse | log2 | 0 | 0.8702 |
| 10 | 50 | friedman\_mse | auto | 1 | 0.8547 |
| 11 | 50 | friedman\_mse | sqrt | 1 | 0.8686 |
| 12 | 50 | friedman\_mse | log2 | 1 | 0.8686 |
| 13 | 50 | absolute\_error | auto | 0 | 0.8526 |
| 14 | 50 | absolute\_error | sqrt | 0 | 0.8708 |
| 15 | 50 | absolute\_error | log2 | 0 | 0.8708 |
| 16 | 50 | absolute\_error | auto | 1 | 0.8549 |
| 17 | 50 | absolute\_error | sqrt | 1 | 0.8696 |
| 18 | 50 | absolute\_error | log2 | 1 | 0.8696 |
| 19 | 50 | poisson | auto | 0 | 0.8491 |
| 20 | 50 | poisson | sqrt | 0 | 0.8632 |
| 21 | 50 | poisson | log2 | 0 | 0.8632 |
| 22 | 50 | poisson | auto | 1 | 0.8524 |
| 23 | 50 | poisson | sqrt | 1 | 0.8655 |
| 24 | 50 | poisson | log2 | 1 | 0.8655 |

**5).** Best model analyed for insurance problem is **Random forest with highest R score value**

**As** 0.8708. Its prediction is better than other models.