# ***Regression Assignment***

1.**Problem Statement**:

To predict the insurance charges based on the input parameters like age, sex, bmi, children, smoker.

2.**Basic Info:**

|  |  |
| --- | --- |
| No. of Rows | 1338 |
| No. of Columns | 6 |

3. **pre-processing method**:

Two columns **sex** and **Smoker** have categorical data which need to converted into numbers through label encoding because both have only two categories of data like male/female (sex) and yes/no(smoker).

4.**Model Creation using Machine Learning Algorithm:**

1.***Multiple Linear regression*:** The r2 score value predicted is : 0.7100923694

2.**Support Vector Machine(Regressor):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **sno** | **C value** | **Kernel=Linear** | **Kernel=rbf(default)** | **Kernel=sigmoid** | **Kernel=poly** |
| 1 | 100 | 0.543281 | -0.124803677 | -0.11814554 | -0.0997617 |
| 2 | 500 | 0.627046 | -0.1246416 | -0.4562944 | -0.082028798 |
| 3 | 1000 | 0.6340369 | -0.11749092 | -1.66590813 | -0.0555059 |
| 4 | 1500 | 0.6394211 | -0.1123899 | -3.31637405 | -0.0287324 |
| 5 | 5000 | 0.764893 | -0.073101073 | -31.56828207 | 0.146223786 |

**3.Decision Tree:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sno.** | **criterion** | **splitter** | **R2 score** |
| 1. | *squared\_error* | best | 0.73182146 |
| 2. | *squared\_error* | random | 0.7580079 |
| 3. | *friedman\_mse* | best | 0.71914332 |
| 4. | *friedman\_mse* | random | 0.70604365 |
| 5. | *absolute\_error* | best | 0.70055001 |
| 6. | *absolute\_error* | random | 0.75877442 |
| 7. | *poisson* | best | 0.766901678 |
| 8. | *poisson* | random | 0.75830571 |

4.**Random Forest:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sno.** | **N\_ESTIMATORS** | **CRITERION** | **MAX\_FEATURES** | **R2 SCORE** |
| 1. | 50 | squared\_error | sqrt | 0.84550549 |
| 2. | 100 | squared\_error | log2 | 0.8500236 |
| 3. | 50 | absolute\_error | sqrt | 0.8492508 |
| 4. | 100 | absolute\_error | log2 | 0.85299949 |
| 5. | 50 | friedman\_mse | sqrt | 0.8472547 |
| 6. | 100 | friedman\_mse | log2 | 0.84458564 |
| 7. | 50 | poisson | sqrt | 0.8435631 |
| 8. | 100 | poisson | log2 | 0.848083785 |

**Conclusion:**

On comparing all the above model the best model is Random Forest and its R2 value is **0.85299**