As a data scientist, the goal is to develop a predictive model that uses these parameters to accurately estimate the **insurance charges**, helping the client make informed decisions about pricing and risk assessment.

Data set contain **1338 rows × 6** columns data with 5 input and 1 output columns.

It have categorical data in Sex column so we need to convert it into **INT** by using **get\_dummies()** and it return table data as **ONE HOT ENCODING** method**.**

To find best **R\_square\_value** from the following models

1. **MULTIPLE LINEAR REGRESSION :** R^2 VALUE= **0.79**
2. **SUPPORT VECTOR MACHINE:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.NO | HYPER\_PARAMETER | LINER(R^2\_VALUE) | RBF(NON\_LINEAR) | POLY | SIGMOID |
| 1 | C10 | 0.50 | -0.04 | 0.045 | 0.04 |
| 2 | C100 | 0.64 | 0.35 | 0.65 | 0.53 |
| 3 | C200 | 0.69 | 0.50 | 0.78 | 0.53 |
| 4 | C500 | 0.63 | -0.102 | -0.057 | -0.6616 |
| 5 | C800 | 0.75 | 0.80 | 0.86 | 0.37 |
| 6 | C1000 | 0.66 | -0.096 | -0.025 | -2.25 |
| 7 | C2000 | 0.74 | 0.86 | 0.86 | -0.87 |
| 8 | C5000 | 0.75 | -0.045 | 0.21 | -42.4 |
| 9 | C10000 | 0.749 | 0.017 | 0.43 | -161.9 |

**DECISION TREE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SL.NO** | **CRITERION** | **Max\_feature** | **SPLITTERS** | **R2\_VALUE** |
| 1 | friedman\_mse | sqrt | best | 0.70 |
| 2 | friedman\_mse | sqrt | random | 0.73 |
| 3 | friedman\_mse | Log2 | best | 0.7482 |
| 4 | friedman\_mse | Log2 | random | 0.7230 |
| 5 | friedman\_mse | none | best | 0.7268 |
| 6 | friedman\_mse | none | random | 0.7857 |
| 7 | squared\_error | sqrt | best | 0.71 |
| 8 | squared\_error | sqrt | random | 0.67 |
| 9 | squared\_error | Log2 | best | 0.7290 |
| 10 | squared\_error | Log2 | random | 0.6876 |
| 11 | squared\_error | none | best | 0.7304 |
| 12 | squared\_error | none | random | 0.6978 |
| 13 | absolute\_error | sqrt | random | 0.81 |
| 14 | absolute\_error | sqrt | best | 0.65 |
| 15 | absolute\_error | Log2 | random | 0.7813 |
| 16 | absolute\_error | Log2 | best | 0.7238 |
| 17 | absolute\_error | none | random | 0.8344 |
| 18 | absolute\_error | none | best | 0.6583 |
| 19 | poisson | sqrt | best | 0.7897 |
| 20 | poisson | sqrt | random | 0.7266 |
| 21 | poisson | Log2 | best | 0.7979 |
| 22 | poisson | Log2 | random | 0.7630 |
| 23 | poisson | none | best | 0.7820 |
| 24 | poisson | none | random | 0.7693 |

**Random forest:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SL.NO** | **CRITERION** | **Max\_feature** | **N\_ESTIMATORS** | **R2\_VALUE** |
| 1 | absolute\_error | none | 50 | 0.8742 |
| 2 | absolute\_error | none | 100 | 0.8743 |
| 3 | absolute\_error | sqrt | 50 | 0.88 |
| 4 | absolute\_error | Log2 | 50 | 0.88 |
| 5 | absolute\_error | sqrt | 100 | 0.89 |
| 6 | absolute\_error | Log2 | 100 | 0.89 |
| 7 | squared\_error | none | 50 | 0.8773 |
| 8 | squared\_error | none | 100 | 0.8770 |
| 9 | squared\_error | sqrt | 50 | 0.89 |
| 10 | squared\_error | Log2 | 50 | 0.89 |
| 11 | squared\_error | sqrt | 100 | 0.894 |
| 12 | squared\_error | Log2 | 100 | 0.89 |
| 13 | friedman\_mse | none | 50 | 0.8753 |
| 14 | friedman\_mse | none | 100 | 0.8764 |
| 15 | friedman\_mse | sqrt | 50 | 0.89 |
| 16 | friedman\_mse | Log2 | 50 | 0.89 |
| 17 | friedman\_mse | sqrt | 100 | 0.89 |
| 18 | friedman\_mse | Log2 | 100 | 0.89 |
| 19 | poisson | sqrt | 50 | 0.8945 |
| 20 | poisson | sqrt | 100 | 0.8915 |
| 21 | poisson | Log2 | 50 | 0.8914 |
| 22 | poisson | Log2 | 100 | 0.8937 |
| 23 | poisson | none | 50 | 0.873 |
| 24 | poisson | none | 100 | 0.873 |

By the analysis of all model r2 value the final best model is **Random\_forest** because it return highest score:**0.89** in **friedman\_mse** with N\_ESTIMATOR as **100** and max\_feature in **sqrt.**