# 1.) Identify your problem statement

Supervised Learning

Machine Learning

Regression

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

RangeIndex: 1338 entries, 0 to 1337

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 age 1338 non-null int64

1 bmi 1338 non-null float64

2 children 1338 non-null int64

3 charges 1338 non-null float64

4 sex\_male 1338 non-null int64

5 smoker\_yes 1338 non-null int64

dtypes: float64(2), int64(4)

memory usage: 62.8 KB

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

# Yes i converted column sex, smoker string column to number

# 5.) All the research values (r2\_score of the models) should be documented.

# (You can make tabulation or screenshot of the results.)

Hyper Tuning parameter

MultiLinearRegression == 0.78

Support Vector Regression.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no | C | kernel | gamma | Output |
| 1 | 1000 | rbf | *scale* | 0.81 |
| 2 | 100 | rbf | *scale* | 0.32 |
| 3 | 1000 | rbf | *auto* | 0.81 |
| 4 | 1000 | linear | *scale* | 0.76 |
| 5 | 100 | linear | *scale* | 0.62 |
| 6 | 1000 | linear | *auto* | 0.76 |
| 7 | 1000 | poly | *scale* | 0.85 |
| 8 | 100 | poly | *scale* | 0.61 |
| 9 | 1000 | poly | *auto* | 0.85 |
| 10 | 1000 | sigmoid | *scale* | 0.28 |
| 11 | 100 | sigmoid | *scale* | 0.52 |
| 12 | 1000 | sigmoid | *auto* | 0.28 |
|  |  |  |  |  |

Decision tree regressor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no | criterion | splitter | max\_features | output |
| 1 | friedman\_mse | random | None | 0.68 |
| 2 | friedman\_mse | random | sqrt | 0.70 |
| 3 | friedman\_mse | best | log2 | 0.72 |
| 4 | friedman\_mse | best | sqrt | 0.76 |
| 5 | friedman\_mse | random | log2 | 0.67 |
| 6 | squared\_error | random | None | 0.71 |
| 7 | squared\_error | random | sqrt | 0.70 |
| 8 | squared\_error | random | log2 | 0.63 |
| 9 | squared\_error | best | sqrt | 0.69 |
| 10 | squared\_error | best | log2 | 0.76 |
| 11 | absolute\_error | random | None | 0.73 |
| 12 | absolute\_error | random | sqrt | 0.78 |
| 13 | absolute\_error | random | log2 | 0.70 |
| 14 | absolute\_error | best | sqrt | 0.66 |
| 15 | absolute\_error | best | log2 | 0.76 |
| 16 | poisson | random | None | 0.76 |
| 17 | poisson | random | sqrt | 0.73 |
| 18 | poisson | random | log2 | 0.73 |
| 19 | poisson | best | sqrt | 0.63 |
| 20 | poisson | best | log2 | 0.67 |

# RandomForestRegressor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no | n\_estimators | criterion | max\_features | output |
| 1 | 50 | friedman\_mse | 1.0 | 0.85 |
| 2 | 50 | friedman\_mse | sqrt | 0.86 |
| 3 | 50 | friedman\_mse | log2 | 0.86 |
| 4 | 50 | squared\_error | 1.0 | 0.85 |
| 5 | 50 | squared\_error | sqrt | 0.87 |
| 6 | 50 | squared\_error | 1og2 | 0.87 |
| 7 | 50 | absolute\_error | 1.0 | 0.85 |
| 8 | 50 | absolute\_error | sqrt | 0.87 |
| 9 | 50 | absolute\_error | log2 | 0.87 |
| 10 | 50 | poisson | 1.0 | 0.85 |
| 11 | 50 | poisson | sqrt | 0.86 |
| 12 | 50 | poisson | log2 | 0.87 |

# 6.) Mention your final model, justify why u have chosen the same.

i suggest Random Forest Regressor because i got more R2 value 0.87