1. Identify your problem statement

stage 1 : **Machine Learning**

stage 2 : **Supervised**

stage 3: **Regression**

1. Tell basic info about the dataset (Total number of rows, columns)
2. Dataset contain 1338 Rows and 6 column
3. 5 input column ( age, sex, bmi, children, smoker) and one output column ( Charges)
4. Sex and smoker are categorical (Ordinal) column so we have to convert as numerical data
5. Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)

Dataset for Before Preprocessing

A screenshot of a screen

Description automatically generated

Dataset for After Preprocessing

1. Converting string to number – Ordinal Mapping (Label Encoder)

A screenshot of a graph

Description automatically generated

1. Develop a good model with r2\_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.
2. **Multiple Liner Regression**

**R2 score is : 0.7894790349867009**

1. **SVM**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No | Hyper Paramater | Linear | Poly | Rbf | Sigmoid |
| 1 | 0.01 | 0.079762068 | -0.089291853 | -0.089695612 | -0.089712414 |
| 2 | 100 | 0.543282021 | -0.100542558 | 0.124704961 | -0.118167633 |
|  |
| 3 | 1000 | 0.634044611 | -0.057971666 | -0.117452608 | -1.688128493 |  |
| 4 | 10000 | 0.744474112 | 0.771666474 | -0.016403495 | -122.3158023 |  |
|  |

**R2 score is 0.771666474**

1. **Descension tree**

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | criterion | splitter | R.score |
| 1 | squared\_error | best | 0.723376349 |
| 2 | squared\_error | random | 0.796537965 |
| 3 | friedman\_mse | best | 0.722677501 |
| 4 | friedman\_mse | random | 0.812447479 |
| 5 | absolute\_error | best | 0.717056762 |
| 6 | absolute\_error | random | 0.798602915 |
| 7 | poisson | best | 0.747182908 |
| 8 | poisson | random | 0.813564941 |

**R2 score is 0.812447479**

1. **Random Forest**

**R2 score is 0.8505475436721328**

**Final Result**

**Random Forest is the best model which gives high R2 score  compare to other model**