1. Need to Predict insurance charges based on given independent parameters

Stage 1 -Domain selection – Given data is in table and Numeric – Machine learning

Stage 2 - Learning method - Input and output are clearly given – Supervised learning

Stage 3 – Model - Output is Numeric – Regression Model

1. Input Fields- age, sex, bmi, children, smoker and output- charges
2. Sex and smoker fields are nominal data , converted to numerical using dummy function in pandas

4/5.R2 results:

|  |  |  |  |
| --- | --- | --- | --- |
| Multi-Linear Regression -- r\_score - 0.7894790349867009 | | | |
|  |  |  |  |
| SVM Regression | | | |
| S.No. | Kernel | C | r\_score |
| 1 | Linear | 1000000 | 0.741404313 |
| 2 | Poly | 1000000 | 0.857938526 |
| 3 | Rbf | 1000000 | 0.869634389 |
| 4 | sigmoid | 1000000 | -368141.4908 |
| 5 | precomputed | 1000000 | NA-it is not a n x n matrix |
|  |  |  |  |
| Decision Tree | | | |
| S.No. | criterion | splitter | r\_score |
| 1 | default=squared\_error | default=best | 0.688139286 |
| 2 | squared\_error | random | 0.777478736 |
| 3 | friedman\_mse | best | 0.703906905 |
| 4 | friedman\_mse | random | 0.688689934 |
| 5 | absolute\_error | best | 0.692562994 |
| 6 | absolute\_error | random | 0.733283279 |
| 7 | Poisson | best | 0.685563229 |
| 8 | Poisson | random | 0.723418879 |
|  |  |  |  |
| Random Forest | | | |
| S.No. | n\_estimators | max\_features | r\_score |
| 1 | 100 | sqrt | 0.871799968 |
| 2 | 100 | log2 | 0.874166294 |
| 3 | 100 | default=1.0 | 0.854355463 |
| 4 | 70 | sqrt | 0.873557758 |
| 5 | 70 | log2 | 0.869811424 |
| 6 | 70 | default=1.0 | 0.855941326 |

6. Final Model is Random Forest with n\_ estimators = 100 and max\_features = log2. Since this showing good evaluation metrics than others.