A client’s requirement is, he want to predict the insurance charges based on several parameters, the client has provided the dataset of the same.

Stage 1- Machine Learning (age, sex, bmi, insurance charges)

Stage 2 - Requirements (Client wants to know insurance charges, the requirement is very clear)

stage3 - Regression (output is in numbers)

**Machine Learning, supervised regression.**

**Multiple Linear Regression** R2 score =0.78913454847886

**Support Vector Machine**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **penalty** | ***linear*** | ***Poly*** | ***rbf*** | ***sigmoid*** |
| 1 | C10 | 0.4624 | 0.0386 | -0.032 | 0.0394 |
| 2 | C100 | 0.629 | 0.6165 | 0.3197 | 0.5268 |
| 3 | C1000 | 0.7648 | 0.8547 | 0.8107 | 0.212 |
| 4 | C2000 | 0.7439 | 0.8583 | 0.8541 | -0.622 |
| 5 | C3000 | 0.7413 | 0.8581 | 0.8646 | -2.143 |

**The best r \_score value identified by using rbf=C3000=0.8646.**

**Decision Tree**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | **criterion** | **Splitter** | **Max\_features** | **r\_score** |
| 1 | 'squared\_error' | Random | sqrt | 0.69284937 |
| 2 | 'squared\_error' | Random | log2 | 0.70828955 |
| 3 | 'squared\_error' | best' | sqrt | 0.66198434 |
| 4 | squared\_error' | best' | log2 | 0.74676672 |
| 5 | ***friedman\_mse*** | Random | sqrt | 0.63930378 |
| 6 | ***friedman\_mse*** | Random | log2 | 0.71299523 |
| 7 | ***friedman\_mse*** | best' | sqrt | 0.7335166 |
| 8 | ***friedman\_mse*** | best' | log2 | 0.68881151 |
| 9 | ***absolute\_error*** | Random | sqrt | 0.67730044 |
| 10 | ***absolute\_error*** | Random | log2 | 0.61229729 |
| 11 | ***absolute\_error*** | best' | sqrt | 0.63831922 |
| 12 | ***absolute\_error*** | best' | log2 | 0.72476148 |
| 13 | ***poisson*** | Random | sqrt | 0.6564799 |
| 14 | ***poisson*** | Random | log2 | 0.6564799 |
| 15 | ***poisson*** | best' | sqrt | 0.75716815 |
| 16 | ***poisson*** | best' | log2 | 0.64424718 |

**The best r \_score value identified by using poisson-best-sqrt=0.75716815**

**Random forest**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S,No | n\_estimators | criterion | Max\_features | r\_score |
|
| 1 | 50 | squared\_error' | sqrt | 0.86038426 |
| 2 | 50 | 'squared\_error' | log2 | 0.86038426 |
| 3 | 100 | squared\_error' | sqrt | 0.86617181 |
| 4 | 100 | 'squared\_error' | log2 | 0.86617181 |
| 5 | 50 | friedman\_mse | sqrt | 0.86038426 |
| 6 | 50 | friedman\_mse | log2 | 0.86038426 |
| 7 | 100 | friedman\_mse | sqrt | 0.86617181 |
| 8 | 100 | friedman\_mse | log2 | 0.86617181 |
| 9 | 50 | absolute\_error | sqrt | 0.8621231 |
| 10 | 50 | absolute\_error | log2 | 0.8621231 |
| 11 | 100 | absolute\_error | sqrt | 0.86666187 |
| 12 | 100 | absolute\_error | log2 | 0.86666187 |
| 13 | 50 | poisson | sqrt | 0.8548262 |
| 14 | 50 | poisson | log2 | 0.8548262 |
| 15 | 100 | poisson | sqrt | 0.85914341 |
| 16 | 100 | poisson | log2 | 0.85914341 |

**The best r \_score value identified by using absolute error, sqrt, log2=0.86662**

Upon checking all the regressions, the final model that performed the best is the Random Forest. The R-score value obtained is 0.86662, indicating strong predictive performance. As a result, the model has been saved, and a deployment has been created for it.