**Insurance Charges Prediction**

1. End goal is Insurance Charge prediction.
2. 1338 Rows and 6 Columns.
3. Nominal Data- String convert to Numbers.
4. R2 researched value given below,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **R2 values:** | | | | | | |
|  |  |  |  |  |  |  |
| **1. Multiple Linear Regression - R2 value=0.789479** | | | | | | |
|  |  |  |  |  |  |  |
| **2. Support Vector Machine-Regression:** | | | | | | |
|  |  |  |  |  |  |  |
| **SL.NO.** | **HYPER PARAMETER** | **LINEAR** | **NON-LINEAR** | **POLY** | **SIGMOID** | **PRECOMPUTED** |
| 1 | C=0.1 | -0.122076 | -0.089567 | -0.086252 | -0.089743 | NONE |
| 2 | C=1000 | 0.634036 | -0.117490 | -0.055505 | -1.665908 | NONE |
| 3 | C=2000 | 0.689326 | -0.107787 | -0.002702 | -5.616431 | NONE |
| 4 | C=3000 | 0.759089 | -0.096212 | 0.048928 | -12.01904 | NONE |
| 5 | C=4000 | 0.765165 | -0.085312 | 0.098408 | -20.72061 | NONE |
| 6 | C=10000 | 0.744482 | -0.017278 | 0.352902 | -119.5185 | NONE |
|  |  |  |  |  |  |  |
| **SVM Regression- R2 value=0.765165 (Linear, Hyper parameter=4000)** | | | | |  |  |
|  |  |  |  |  |  |  |
| **3. Decision Tree:** | | | |  |  |  |
|  |  |  |  |  |  |  |
| **SL.NO** | **CRITERION** | **SPLITER** | **R2 VALUES** |  |  |  |
| 1 | ***squared\_error*** | ***best*** | 0.721080 |  |  |  |
| 2 | ***squared\_error*** | ***random*** | 0.691421 |  |  |  |
| 3 | ***friedman\_mse*** | ***best*** | 0.686401 |  |  |  |
| 4 | ***friedman\_mse*** | ***random*** | 0.760271 |  |  |  |
| 5 | ***absolute\_error*** | ***best*** | 0.656121 |  |  |  |
| 6 | ***absolute\_error*** | ***random*** | 0.679909 |  |  |  |
| 7 | ***poisson*** | ***best*** | 0.703009 |  |  |  |
| 8 | ***poisson*** | ***random*** | 0.694218 |  |  |  |
|  |  |  |  |  |  |  |
| **Decision Tree- R2 value=0.760271(Criterion=friedman\_mse, splitter-random)** | | | | |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **4. Random Forest:** | | | |  |  |  |
|  |  |  |  |  |  |  |
| **SL.NO** | **CRITERION** | **n\_estimators** | **R2 VALUES** |  |  |  |
| 1 | ***squared\_error*** | ***100*** | 0.8535520 |  |  |  |
| 2 | ***squared\_error*** | ***50*** | 0.848606 |  |  |  |
| 3 | ***friedman\_mse*** | ***100*** | 0.859751 |  |  |  |
| 4 | ***friedman\_mse*** | ***50*** | 0.849704 |  |  |  |
| 5 | ***absolute\_error*** | ***100*** | 0.852664 |  |  |  |
| 6 | ***absolute\_error*** | ***50*** | 0.853649 |  |  |  |
| 7 | ***poisson*** | ***100*** | 0.852775 |  |  |  |
| 8 | ***poisson*** | ***50*** | 0.849333 |  |  |  |
|  |  |  |  |  |  |  |
| **RandoForestRegressior- R2 value=0.859751(Criterion=friedman\_mse, n\_estimators-100)** | | | | | |  |

Final model is **RandoForestRegressior- R2 value=0.859751(Criterion=friedman\_mse, n\_estimators-100)**

**Final Output:**

**independent=dataset[['age', 'bmi', 'children','sex\_male', 'smoker\_yes']]**

**I/P Values: [50,30.970,2,1,1]**

**dependent=dataset[["charges"]]**

**Output Values:**  **[42969.5701375]**

**Note: Accuracy improved from 60% to 80%**