**REGRESSION ASSIGNMENT**

1. Identify your problem statement

**The charges for insurance have to be predicted.**

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

**Total no. of rows(inputs) =1312, no. of columns = 6.**

**(Input columns = 5 & output column =1)**

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

**Pre-processing:**

**Stage1: Machine learning**

**Stage2: Supervised learning (Requirement clear, I/p & O/p clear)**

**Stage3: Regression (as output requirement is numbers)**

**Input columns (Independent)🡪 age, sex, BMI, children, smoker**

**Output columns (Dependent) 🡪Charges**

4.) 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.

**Attached in the next sheet. Hopefully Forest tree regression is better with results 0.85 but not recommend as none came neither 90% nor 95%**

5.) All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

**Tabulation attached below.**

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

**Hopefully Forest tree regression is better with results 0.85 but not recommend as none came neither 90% nor 95%**

**ASSIGNMENT – HYPER TUNING OPERATORS**

**Support Vector Machine:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.No.** | **Kernel** | **C** | **degree** | **R^2 score** |
| 1 | Linear | 0.1/10/100 | 3 | -0.08/0.46/0.62 |
| ***2*** | Poly | 0.1/10/100 | 5 | -0.08/0.08/0.64 |
| 3 | Rbf | 0.1/10/100 | 2 | -0.08/-0.03/0.32 |
| 4 | ***Sigmoid*** | 100/150/200 | 1 | 0.52/0.53/0.54 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.No.** | **Regressor Type (criterion)** | **Splitter type** | **Depth\_Max** | **R^2 score** |
| 1 | Squared error | best | None/10 | 0.69/0.72 |
| ***2*** | ***Squared error*** | random | None/100 | 0.63/0.75 |
| 3 | ***absolute\_error*** | best | None/100 | 0.67/0.69 |
| 4 | ***absolute\_error*** | random | None/100 | 0.74/0.71 |
| 5 | ***Poisson*** | best | None/100 | 0.67/0.69 |
| 6 | ***Poisson*** | random | None/100 | 0.58/0.62 |
| 7 | ***friedman\_mse*** | best | None/50 | 0.69/0.71 |
| 8 | ***friedman\_mse*** | random | None/50 | 0.70/0.70 |

**Decision Tree Regressor**

**Forest Tree Regressor**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.no** | ***n\_estimators*** | ***criterion*** | ***max\_depth(None/int)*** | **R^2 score** |
| 1 | 100 | ***friedman\_mse*** | None/100 | 0.85/0.85 |
| 2 | 50 | ***friedman\_mse*** | None/10 | 0.85/0.86 |
| 3 | 10 | ***friedman\_mse*** | None/10 | 0.84/0.85 |
| 4 | 100 | ***poisson*** | None/100 | 0.83/0.83 |
| 5 | 50 | ***poisson*** | None/10 | 0.83/0.82 |
| 6 | 10 | ***poisson*** | None/10 | 0.81/0.82 |
| 7 | 100 | ***squared\_error*** | None/100 | 0.85/0.85 |
| 8 | 50 | ***squared\_error*** | None/10 | 0.85/0.85 |
| 9 | 10 | ***squared\_error*** | None/10 | 0.84/0.85 |