Regression Assignment

1. Identify Problem Statement – Since Requirement is clear, I/P and O/P are well defined falls under

Supervised Learning.

1. Basic info about dataset – dataset has 6 columns and 1338 rows
2. Preprocessing method – As there are 2 categorical columns, One hot encoding is used for those.

Scaling is also used.

**Multi Linear**

R2Value- 0.7894790349867009

**SVM**

|  |  |  |
| --- | --- | --- |
| Kernel | C | R2 Value |
| Rbf | 100 | 0.3200317832050831 |
| poly | 0.6179569624059795 |
| linear | 0.6288792857320359 |
| Sigmoid | 0.5276103546510407 |
| Rbf | 1000 | 0.8102064851758545 |
| poly | 0.8566487675946604 |
| linear | 0.764931173859732 |
| Sigmoid | 0.28747069486976695 |
| Rbf | 10000 | 0.8779952403530633 |
| poly | 0.8591715075953347 |
| linear | 0.744482485232016 |
| Sigmoid | -34.151535978496256 |

**Decision Tree**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criterion** | **Splitter** | **ccp\_alpha** | **R2 value** |
| *squared\_error* | Random(random\_state = 50)  best | 5.5 | 0.7241849114674901  0.6958271421479616 |
| *friedman\_mse* | Random(random\_state = 50)  best | 5.5 | 0.7183255548258962  0.6923104505410319 |
| *absolute\_error* | Random(random\_state = 50)  best | 5.5 | 0.7699587195315052  0.7055319904426092 |
| poisson | Random(random\_state = 50)  best | 5.5 | 0.8695895967821924  0.8056102030872151 |

**Random Forest Regressor**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criterion** | **ccp\_alpha** | **Random\_state** | **R2 value** |
| *squared\_error* | 15.0 | 40 | 0.8597710804106606 |
| *friedman\_mse* | 15.0 | 40 | 0.8591655219159275 |
| *absolute\_error* | 15.0 | 40 | **0.8860891505724752** |
| poisson | 15.0 | 40 | 0.8856023085905353 |

Random Forest gives better R2 Value compared to other algorithms.