**Machine Learning Regression**

1.Identify your Problem Statement

3 - Stages of Problem Identification

Stage 1-Machine Learning

Stage 2-Supervised learning

Stage 3-Regression

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

There are 1338 rows and 6 columns in the Dataset

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

I used One Hot Encoding to convert nominal data in sex and smoker Column

**Multiple Learning Regression**

Multiple Linear Regression r\_score value= 0.7894790349867009

**Support Vector Machine-Regression**

Support Vector Regression

R score value Without Standarisation= -0.11166128719608448

Standardised by passing arguments

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Parameter** | **Linear** | **rbf** | **Poly** | **Sigmoid** |
| 1 | **C=1** | **-0.010102665316081394** | **-0.08338238593619329** | **-0.07569965570860893** | **-0.07542924281107188** |
| 2 | **C=10** | **0.462468414233968** | **-0.03227329390671052** | **0.038716222760231345** | **0.03930714378274347** |
| 3 | **C=100** | **0.6288792857320359** | **0.3200317832050831** | **0.6179569624059795** | **0.5276103546510407** |
| 4 | **C=500** | **0.763105797597537** | **0.6642984645143137** | **0.826368354126896** | **0.44460610338694795** |
| 5 | **C=1000** | **0.7649311738608445** | **0.8102064851758545** | **0.8566487675946551** | **0.28747069486976695** |
| 6 | **C=2000** | **0.7440418308107487** | **0.854776642539298** | **0.860557927577388** | **-0.5939509731283508** |
| 7 | **C=3000** | **0.7414236599249815** | **0.8663393953081686** | **0.8598930084494408** | **-2.1244194786689863** |

**SVM Regression using Hyper Tuning Parameters**

**SVR(C=3000, kernel='rbf') Model has the Highest Accuracy**

**R\_score value is 0.8663393953081686**

**Decision Tree Regression**

R\_Score Value ,without passing any Parameters=0.68225721169395

Standarised R\_Score Value ,without passing any Parameters= 0.7149863334547812

Finding the Best model by Standarised and using Hyper Tuning Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | **Criterion** | **Splitter** | **max\_features** | **R\_score** |
| 1 | **squared\_error** | **Best** | **None** | 0.6882468013438084 |
| 2 | **squared\_error** | **random** | **None** | 0.6350861896159944 |
| 3 | **squared\_error** | **Best** | **Sqrt** | 0.6861211492032002 |
| 4 | **squared\_error** | **random** | **Sqrt** | 0.6691223172490448 |
| 5 | **squared\_error** | **Best** | **log2** | 0.6083716375423366 |
| 6 | **squared\_error** | **Random** | **log2** | 0.6809039559624114 |
| 7 | **friedman\_mse** | **best** | **None** | 0.7185125720055741 |
| 8 | **friedman\_mse** | **random** | **None** | 0.7200747053945815 |
| 9 | **friedman\_mse** | **best** | **Sqrt** | 0.6724064992834135 |
| 10 | **friedman\_mse** | **random** | **Sqrt** | 0.6827149209338792 |
| 11 | **friedman\_mse** | **best** | **log2** | 0.6502424951784631 |
| 12 | **friedman\_mse** | **random** | **log2** | 0.6045152032902725 |
| 13 | **absolute\_error** | **best** | **None** | 0.6844630566610516 |
| 14 | **absolute\_error** | **random** | **None** | 0.7553107619270651 |
| 15 | **absolute\_error** | **best** | **sqrt** | 0.7423897960308923 |
| 16 | **absolute\_error** | **random** | **sqrt** | 0.6997865322956003 |
| 17 | **absolute\_error** | **best** | **log2** | 0.5524392075238371 |
| 18 | **absolute\_error** | **random** | **log2** | 0.6224368161247387 |
| 19 | **Poisson** | **best** | **None** | 0.7339978683929528 |
| 20 | **Poisson** | **random** | **None** | 0.7152916153352498 |
| 21 | **Poisson** | **best** | **sqrt** | 0.7146998218266408 |
| 22 | **Poisson** | **random** | **sqrt** | 0.6434102735459901 |
| 23 | **Poisson** | **best** | **log2** | 0.7637655770462719 |
| 24 | **Poisson** | **Random** | **log2** | 0.6870540191516312 |

**The Best model in Decision Tree Regression is Criterion = Poisson, Splitter =best, max\_features=log2**

**R\_score value is**  0.7637655770462719

**Random Forest Regression**

**Default r\_score value=0.8498329315421834**

**Standarised Default r\_score value=** **0.8509112530145887**

**Finding the Best model by using Hyper Tuning Parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.no** | **Criterion** | **n\_estimators** |  | R\_score |
| 1 | **squared\_error** | 50 | **None** | 0.8509112530145887 |
| 2 | **squared\_error** | 100 | **None** | 0.8584187638082341 |
| 3 | **squared\_error** | 50 | **Sqrt** | 0.863547084044146 |
| 4 | **squared\_error** | 100 | **Sqrt** | 0.8700823821127073 |
| 5 | **squared\_error** | 50 | **log2** | 0.8705152681431384 |
| 6 | **squared\_error** | 100 | **log2** | 0.8725053540489766 |
| 7 | **friedman\_mse** | 50 | **None** | 0.8463543973160762 |
| 8 | **friedman\_mse** | 100 | **None** | 0.8544292837510767 |
| 9 | **friedman\_mse** | 50 | **Sqrt** | 0.8665728553506091 |
| 10 | **friedman\_mse** | 100 | **Sqrt** | 0.8701064795661044 |
| 11 | **friedman\_mse** | 50 | **log2** | 0.8715977439671387 |
| 12 | **friedman\_mse** | 100 | **log2** | 0.873026039925459 |
| 13 | **absolute\_error** | 50 | **None** | 0.8541782532783613 |
| 14 | **absolute\_error** | 100 | **None** | 0.8577713056040487 |
| 15 | **absolute\_error** | 50 | **sqrt** | 0.8677030951110538 |
| 16 | **absolute\_error** | 100 | **sqrt** | 0.87478367719952 |
| 17 | **absolute\_error** | 50 | **log2** | 0.8722915856314009 |
| 18 | **absolute\_error** | 100 | **log2** | 0.8756932802547971 |
| 19 | **Poisson** | 50 | **None** | 0.8578504305627456 |
| 20 | **Poisson** | 100 | **None** | 0.8571625312412804 |
| 21 | **Poisson** | 50 | **sqrt** | 0.871821282525396 |
| 22 | **Poisson** | 100 | **sqrt** | 0.8740796644059626 |
| 23 | **Poisson** | 50 | **log2** | 0.8700339017674392 |
| 24 | **Poisson** | 100 | **log2** | 0.8701135923821007 |

**The Best model in Random Forest Regression is Criterion = absolute\_error, n\_estimators =100, max\_features=log2**

**R\_score value is**  0.8756932802547971

* **Random Forest Regression is the best model to predict the Insurance Charges when compared to the other models.**