**Machine Learning Regression**

**Assignment-Regression Algorithm**

**Insurance charges prediction**

**Best R2-Score value RANDOM FOREST REGRESSION**

**R2-Score** **Vaule** =**(mae,sqrt)= (0.8639)**

**All the machine learning regression algoritham to comparatively random forest regression best prediction of Insurance charges prediction**

**1.Multiple linear Regression**

**R2-Score = 0.7894790349867009**

**2.SUPPORT VECTOR MACHINE REGRESSION**

**R2-Score Value** = **(POLY,C=0.1,AUTO) =(0.8629)**

**3. DECISION TREE REGRESSION**

**R2-Score Value** = **(Friedmans, sqrt, random) =(0.7771)**

**4. RANDOM FOREST REGRESSION**

**R2-Score** **Vaule** =**(mae,sqrt)= (0.8639)**

**Tabulasation and all steps of prediction stages**

1.Find out the 3 -Stage of Problem Identification

Stage1- Machine Learning

Stage2- supervised Learning

Stage3- Regression

2.Name the project

Insurance charges prediction

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

1. Total number of rows

1338 rows

2. Total number of columns – 6 Columns

1. age 2. Sex 3. Bmi 4. Children 5. Smoker 6. Charges

|  | **age** | **sex** | **bmi** | **children** | **smoker** | **charges** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 19 | female | 27.900 | 0 | yes | 16884.92400 |
| 1 | 18 | male | 33.770 | 1 | no | 1725.55230 |
| 2 | 28 | male | 33.000 | 3 | no | 4449.46200 |
| 3 | 33 | male | 22.705 | 0 | no | 21984.47061 |
| 4 | 32 | male | 28.880 | 0 | no | 3866.85520 |
| ... | ... | ... | ... | ... | ... | ... |
| 1333 | 50 | male | 30.970 | 3 | no | 10600.54830 |
| 1334 | 18 | female | 31.920 | 0 | no | 2205.98080 |
| 1335 | 18 | female | 36.850 | 0 | no | 1629.83350 |
| 1336 | 21 | female | 25.800 | 0 | no | 2007.94500 |
| 1337 | 61 | female | 29.070 | 0 | yes | 29141.36030 |

1338 rows × 6 columns

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |

**4.use to categorical to numerical coverted**

dataset=pd.get\_dummies(dataset)

dataset

| **age** | **bmi** | **children** | **charges** | **sex\_female** | **sex\_male** | **smoker\_no** | **smoker\_yes** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 19 | 27.900 | 0 | 16884.92400 | 1 | 0 | 0 | 1 |
| 1 | 18 | 33.770 | 1 | 1725.55230 | 0 | 1 | 1 | 0 |
| 2 | 28 | 33.000 | 3 | 4449.46200 | 0 | 1 | 1 | 0 |
| 3 | 33 | 22.705 | 0 | 21984.47061 | 0 | 1 | 1 | 0 |
| 4 | 32 | 28.880 | 0 | 3866.85520 | 0 | 1 | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1333 | 50 | 30.970 | 3 | 10600.54830 | 0 | 1 | 1 | 0 |
| 1334 | 18 | 31.920 | 0 | 2205.98080 | 1 | 0 | 1 | 0 |
| 1335 | 18 | 36.850 | 0 | 1629.83350 | 1 | 0 | 1 | 0 |
|  |  |  |  |  |  |  |  |  |
| 1336 | 21 | 25.800 | 0 | 2007.94500 | 1 | 0 | 1 | 0 |
| 1337 | 61 | 29.070 | 0 | 29141.36030 | 1 | 0 | 0 | 1 |

1338 rows × 8 columns

**5.Input/output split the dataset**

**Input split**

| **age** | **bmi** | **children** | **sex\_female** | **sex\_male** | **smoker\_no** | **smoker\_yes** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 19 | 27.900 | 0 | 1 | 0 | 0 | 1 |
| 1 | 18 | 33.770 | 1 | 0 | 1 | 1 | 0 |
| 2 | 28 | 33.000 | 3 | 0 | 1 | 1 | 0 |
| 3 | 33 | 22.705 | 0 | 0 | 1 | 1 | 0 |
| 4 | 32 | 28.880 | 0 | 0 | 1 | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 1333 | 50 | 30.970 | 3 | 0 | 1 | 1 | 0 |
| 1334 | 18 | 31.920 | 0 | 1 | 0 | 1 | 0 |
| 1335 | 18 | 36.850 | 0 | 1 | 0 | 1 | 0 |
| 1336 | 21 | 25.800 | 0 | 1 | 0 | 1 | 0 |
| 1337 | 61 | 29.070 | 0 | 1 | 0 | 0 | 1 |

1338 rows × 7 columns

**Output Split**

| **charges** |
| --- |
| 0 | 16884.92400 |
| 1 | 1725.55230 |
| 2 | 4449.46200 |
| 3 | 21984.47061 |
| 4 | 3866.85520 |
| ... | ... |
| 1333 | 10600.54830 |
| 1334 | 2205.98080 |
| 1335 | 1629.83350 |
| 1336 | 2007.94500 |
| 1337 | 29141.36030 |

1338 rows × 1 columns

**6.Train-set and test-set split**

from sklearn.model\_selection import train\_test\_split

x\_train,x\_test,y\_train,y\_test=train\_test\_split(independent,dependent,test\_size=0.30,random\_state=0)

**Machine Learning Regression**

**Assignment-Regression Algorithm**

**Insurance charges prediction**

**TABULATION**

**1.Multiple linear Regression R2-Score = 0.7894790349867009**

**2. SUPPORT VECTOR MACHINE REGRESSION**

The Decision Tree Regression best

**R2-Score Value** = **(POLY,C=0.1,AUTO) =(0.8629)**

Below the table using a all parameter Kernel, C and Gamma use R2-Score Value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **KERNEL** | **C** | **GAMMA** | **R2-SCORE** |
| 1 | linear | 0.01 | auto | -0.0797 |
| 2 | linear | 10 | Scale | -0.0016 |
| 3 | linear | 100 | Scale | 0.5432 |
| 4 | linear | 1000 | scale | 0.6340 |
| 5 | linear | 2000 | scale | 0.6893 |
| 6 | linear | 10 | auto | -0.0016 |
| 7 | linear | 100 | auto | 0.5432 |
| 8 | linear | 2000 | auto | 0.6893 |
| 9 | poly | 0.01 | scale | -0.0893 |
| 10 | poly | 10 | scale | -0.0931 |
| 11 | poly | 100 | scale | -0.0997 |
| 12 | poly | 0.1 | auto | 0.8629 |
| 13 | poly | 0.01 | auto | 0.8377 |
| 14 | rbf | 2000 | auto | -0.1077 |
| 15 | rbf | 2000 | scale | 0.00028 |
| 16 | sigmoid | 2000 | auto | -5.616 |
| 17 | sigmoid | 0.01 | auto | 0.0897 |
| 18 | sigmoid | 0.01 | scale | -5.6164 |
| 19 | Linear(standard) | 2000 | scale | 0.7440 |

The Decision Tree Regression best

**R2-Score Value** = **(POLY,C=0.1,AUTO) =(0.8629)**

**3. DECISION TREE REGRESSION**

The Decision Tree Regression best

**R2-Score Value** = **(Friedmans, sqrt, random) =(0.7771)**

Below the table using a all parameter Criterion, Max\_Features and Splitter use R2-Score Value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **CRITERION** | **MAX\_FEATURES** | **SPLITTER** | **R2-SCORE** |
| 1 | mse | auto | best | 0.7050 |
| 2 | mse | auto | random | 0.7120 |
| 3 | mse | sqrt | best | 0.6936 |
| 4 | mse | sqrt | random | 0.6833 |
| 5 | mse | log2 | best | 0.6917 |
| 6 | mse | log2 | random | 0.6651 |
| 7 | mae | auto | best | 0.6851 |
| 8 | mae | auto | random | 0.7643 |
| 9 | mae | sqrt | best | 0.7113 |
| 10 | mae | sqrt | random | 0.7607 |
| 11 | mae | log2 | best | 0.7127 |
| 12 | mae | log2 | random | 0.6650 |
| 13 | Friedman\_mse | auto | best | 0.6982 |
| 14 | Friedman\_mse | auto | random | 0.7151 |
| 15 | Friedman\_mse | sqrt | best | 0.7142 |
| 16 | Friedman\_mse | sqrt | random | 0.7771 |
| 17 | Friedman\_mse | Log2 | best | 0.6805 |
| 18 | Friedman\_mse | Log2 | random | 0.6287 |

The Decision Tree Regression best

**R2-Score Vaule** =**(Friedmans, sqrt, random)=(0.7771)**

**4. RANDOM FOREST REGRESSION**

1.The Decision Tree Regression best **R2-Score** Vaule =**(mae,sqrt)= (0.8639)**

2.And Also, The Decision Tree Regression best

**R2-Score** Vaule = **max\_depth=( 0.8668)**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO** | **CRITERION** | **MAX\_FEATURES** | **R2-SCORE** |
| 1 | mse | sqrt | 0.8405 |
| 2 | mse | log2 | 0.8452 |
| 3 | mae | sqrt | 0.8639 |
| 4 | mae | log2 | 0.8539 |
| 5 | friedman\_mse | sqrt | 0.8594 |
| 6 | friedman\_mse | log2 | 0.8499 |
|  |  |  |  |

**R2-SCORE**

1)n\_estimators=100 0.8561

2)max\_depth=1 0.6699

3)max\_depth=2 0.8668

4)min\_impurity\_decrease=0 0.8299

5)bootstrap=True 0.8428

6)bootstrap=False 0.7009