Prediction of insurance charages

1. Proplem statement

To predict insurance charages is estimated with the given dataset.

The inputs was cleared in number format and some inputs are in caregorical. The output is a only in numerical value that’s means Reegression.

**Stage1 :** Machine Learning - ( input-number )

**Stage2 :** Supervised Learning - ( input output is cleard )

**Stage3 :** Regression - ( output numerical value )

1. Basic info abourt dataset.

* Total number of Rows 1339.
* Total number of Coulums 6.
* Total number of Inputs 5. ( Age,Bmi,sex,Children,Smoker )
* Total number of Outputs 1. ( Charages )

1. Pre-processing

The categorical value inputs are convert to ordinal data.

Multiple linear

The multiple linear Regression R2 Value is = 0.78913

Support vector machine

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.NO | KERNEL | GAMMA | C | R2\_Score |
| 1 | Linear | Scale | 10 | -0.00170 |
|  | Linear | Scale | 100 | 0.54322 |
|  | Linear | Scale | 1000 | 0.63386 |
|  | Linear | Scale | 3000 | 0.75902 |
| 2 | Poly | Scale | 10 | -0.09309 |
|  | Poly | Scale | 100 | -0.09923 |
|  | Poly | Scale | 1000 | -0.05465 |
|  | Poly | Scale | 3000 | 0.04941 |
| 3 | Rbf | Scale | 10 | -0.08188 |
|  | Rbf | Scale | 100 | -.124500 |
|  | Rbf | Scale | 1000 | -0.11761 |
|  | Rbf | Scale | 3000 | -0.09623 |
| 4 | Sigmoid | Scale | 10 | -0.09093 |
|  | Sigmoid | Scale | 100 | -0.11850 |
|  | Sigmoid | Scale | 1000 | -1.71123 |
|  | Sigmoid | Scale | 3000 | -12.5445 |
| 5 | Linear | Auto | 10 | -0.00170 |
|  | Linear | Auto | 100 | 0.54322 |
|  | Linear | Auto | 1000 | 0.63386 |
|  | Linar | Auto | 3000 | 0.75902 |
| 6 | Rbf | Float | 10 | -0.08752 |
|  | Rbf | Float | 100 | -0.07503 |
|  | Rbf | Float | 1000 | -0.04196 |
|  | Rbf | Float | 3000 | 0.04296 |
| 7 | Sigmid | Float | 10 | -0.08972 |
|  | Sigmid | Float | 100 | -0.08972 |
|  | Sigmoid | Float | 1000 | -0.08972 |
|  | Sigmoid | Float | 3000 | -0.08972 |

The SVM Regression use R2 value ( kernel-linear,gamma-scsle and hyper parameter(C-3000))= 0.75902

RandomForest

The RF Regression use R2 value ( criterion-poission, n\_estimators)= 0.98209

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | CRITERION | N\_estimators | R2\_score |
| 1 | Squared\_error | 100 | 0.98180 |
| 2 | Absoult\_error | 100 | 0.98111 |
| 3 | Friedman\_msc | 100 | 0.98186 |
| 4 | poission | 100 | 0.98144 |
| 5 | Squared\_error | 200 | 0.98246 |
| 6 | Absolute\_error | 200 | 0.98157 |
| 7 | Friedman\_msc | 200 | 0.98209 |
| 8 | poission | 200 | 0.98209 |

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | CRITERION | SPLITTER | R2\_SCORE |
| 1 | Squared\_error | Best | 0.69155 |
| 2 | Friedman | Best | 0.71546 |
| 3 | Absoult\_error | Best | 0.65950 |
| 4 | Poission | Best | 0.72686 |
| 5 | Squared\_error | Random | 0.69092 |
| 6 | Friedman | Random | 0.71307 |
| 7 | Absoult\_error | Random | 0.72198 |
| 8 | Poission | Random | 0.76970 |

DecisionTree

The SVM Regression use R2 value ( criterion-poission,splitter-random)= 0.75902

THE FINAL MODEL OF EVALATION ALGRITHM IS RANDOM FOREST