**REGRESSION ALGORITHM**

DATE:10.8.22

STATEMENT: **CLIENT WANT TO PREDICT INSURANCE CHARGES**

DATASET: http://raw.githubusercontent.com/RamishaRaniK/dataset/main/insurance\_pre.csv

ABOUT MY DATASET : **MACHINE LEARNING-SUPERVISED LEARNING** (I/P AND O/P clear) Datasets.shape : **.shape** used to find **no of rows and columns** in our Dataset.

Pre-processing step : My data set have 2 columns of string data. Machine learning support only numerical data.so, we should **convert string into byte stream**. If the data values are non-comparable (yes/no )then go for **nominal-one heart encoding(using .getdummies**())

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SIMPLE LINEAR REGRESSION**(*one input &one output*)  Based on Linear Pattern(straight line) | **MULTIPLE LINEAR REGRESSION**(*multiple input & one output*)  Based on Linear dPattern (straight line) | **SUPPORT VECTOR MACHINE** (Goal: Create *a best line* or Hyperplane b/w the datapoints(vector).  Based on Linear and Non-Linear Pattern. | **DECISION TREE** (It divides the complete dataset into smaller subsets-Tree like model)  SUPERVISED LEARNING (input and output are clear) | **RANDOM FOREST**  (Taking random columns to build many Decisions Tree) |
| My Dataset has 4  inputs and one output, so, I cannot go for simple linear algorithm. | According to my dataset multiple inputs, I preferred MLR. | Here,MLR gave moderate result . so I preferred SVM. | Checked with Decision Tree Algorithm to get better r2\_score value. | Checked with Random forest Algorithm to get better r2\_score value. |
| NIL | **R2\_score value**:78%  My model predicted as 0.78 %, so I go for SVM algorithm. | **R2\_score value**:-97%  My model predicted -0.097% This is the very poor model algorithm.  Have to improve **model performance by**  **Changing kernel value and playable parameters** | **R2\_score value**:68%  My model predicted 0.688%. This is poor r2\_score value | **R2\_score value:**85%  My model predicted 0.85%. This is moderate r2\_score value |

**The final machine learning best method of Regression:**

1.**Random forest R2 \_ value=85%**

**TO IMPROVE THE MODEL R2\_SCORE VALUE ,YOU SHOULD CHANGE THE KERNEL AND PLAYABLE PARAMETER VALUES.**

**SVM:** **kernel*{‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’} or callable, default=’rbf’***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| palyable parameter | Linear  R2 value | Rbf  R2 value | Sigmoid  R2 value | Poly  R2 value |
| C=10 | 0.4 | 0.604 | 0.20 | 0.0143 |
| C=12000 | 0.76133 |  |  |  |
|  |  |  |  |  |

**FINALLY I SUGGEST RANDOM FOREST ALGORITHM BY COMPARED R2\_SCORE VALUE.**

