**Identify your problem statement:**

To predict the insurance charges based on the age,sex,gender,no of children,bmi & smoke habit

**Tell basic info about the dataset:**

Total no of rows-1339

No of coloumns:6

**Mention the pre-processing method if you’re doing any (like converting**

**string to number – nominal data):**

Convert the Sex,Smoker into Nominal data

**Multi Linear Algorithm:**

1338 rows × 8 columns

R2\_Score: 0.78

**SVM:**

|  |  |  |  |
| --- | --- | --- | --- |
| **kernel** | **Gamma** | **C** | **r2\_score** |
| linear | scale |  | -0.01 |
| rbf | scale |  | -0.083 |
| poly | scale |  | -0.075 |
| ***sigmoid*** | scale |  | -0.075 |
| ***precomputed*** | scale |  | unsupported |
| linear | auto |  | -0.01 |
| rbf | auto |  | -0.083 |
| poly | auto |  | -0.075 |
| ***sigmoid*** | auto |  | -0.075 |
| ***precomputed*** | auto |  | unsupported |
| linear | scale | 10 | 0.46 |
| rbf | scale | 10 | -0.032 |
| poly | scale | 10 | 0.038 |
| ***sigmoid*** | scale | 10 | 0.039 |
| ***precomputed*** | scale | 10 | unsupported |
| linear | auto | 10 | 0.46 |
| rbf | auto | 10 | -0.032 |
| poly | auto | 10 | 0.038 |
| ***sigmoid*** | auto | 10 | 0.039 |
| ***precomputed*** | auto | 10 | unsupported |
| linear | scale | 100 | 0.62 |
| rbf | scale | 100 | 0.32 |
| poly | scale | 100 | 0.61 |
| ***sigmoid*** | scale | 100 | 0.52 |
| ***precomputed*** | scale | 100 | unsupported |
| linear | auto | 100 | 0.62 |
| rbf | auto | 100 | 0.32 |
| poly | auto | 100 | 0.61 |
| ***sigmoid*** | auto | 100 | 0.52 |
| ***precomputed*** | auto | 100 | unsupported |
| linear | scale | 1000 | 0.76 |
| rbf | scale | 1000 | 0.81 |
| poly | scale | 1000 | 0.85 |
| ***sigmoid*** | scale | 1000 | 0.28 |
| ***precomputed*** | scale | 1000 | unsupported |
| linear | auto | 1000 | 0.76 |
| rbf | auto | 1000 | 0.81 |
| poly | auto | 1000 | 0.85 |
| ***sigmoid*** | auto | 1000 | 0.28 |
| ***precomputed*** | auto | 1000 | unsupported |
| linear | scale | 2000 | 0.74 |
| rbf | scale | 2000 | 0.85 |
| poly | scale | 2000 | 0.86 |
| ***sigmoid*** | scale | 2000 | -0.59 |
| ***precomputed*** | scale | 2000 | unsupported |
| linear | auto | 2000 | 0.74 |
| rbf | auto | 2000 | 0.85 |
| poly | auto | 2000 | 0.86 |
| ***sigmoid*** | auto | 2000 | -0.59 |
| ***precomputed*** | auto | 2000 | unsupported |

Highest R2\_score value :0.86

Decision tree:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SNO | criterion | splitter | max\_features | r value |
| 1 | ***friedman\_mse*** | random | None | 0.69 |
| 2 | ***friedman\_mse*** | random | auto | 0.70 |
| 3 | ***friedman\_mse*** | random | ***sqrt*** | 0.51 |
| 4 | ***friedman\_mse*** | random | ***log2*** | 0.62 |
| 5 | ***friedman\_mse*** | best | None | 0.68 |
| 6 | ***friedman\_mse*** | best | auto | 0.68 |
| 7 | ***friedman\_mse*** | best | ***sqrt*** | 0.68 |
| 8 | ***friedman\_mse*** | best | ***log2*** | 0.67 |
| 9 | squared\_error | best | None | 0.71 |
| 10 | squared\_error | best | auto | 0.67 |
| 11 | squared\_error | best | ***sqrt*** | 0.65 |
| 12 | squared\_error | best | ***log2*** | 0.75 |
| 13 | squared\_error | random | None | 0.65 |
| 14 | squared\_error | random | auto | 0.70 |
| 15 | squared\_error | random | ***sqrt*** | 0.66 |
| 16 | squared\_error | random | ***log2*** | 0.69 |
| 17 | absolute\_error | best | None | 0.68 |
| 18 | absolute\_error | best | auto | 0.66 |
| 19 | absolute\_error | best | ***sqrt*** | 0.75 |
| 20 | absolute\_error | best | ***log2*** | 0.71 |
| 21 | absolute\_error | random | None | 0.73 |
| 22 | absolute\_error | random | auto | 0.67 |
| 23 | absolute\_error | random | ***sqrt*** | 0.61 |
| 24 | absolute\_error | random | ***log2*** | 0.67 |
| 25 | ***poisson*** | best | None | 0.71 |
| 26 | ***poisson*** | best | auto | 0.74 |
| 27 | ***poisson*** | best | ***sqrt*** | 0.68 |
| 28 | ***poisson*** | best | ***log2*** | 0.67 |
| 29 | ***poisson*** | random | None | 0.68 |
| 30 | ***poisson*** | random | auto | 0.65 |
| 31 | ***poisson*** | random | ***sqrt*** | 0.57 |
| 32 | ***poisson*** | random | ***log2*** | 0.59 |

Highest r2\_score value: 0.75

|  |  |  |  |
| --- | --- | --- | --- |
| absolute\_error | best | ***sqrt*** | 0.75 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2 | ***friedman\_mse*** | random | auto | 0.92 |

RF:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SNO | n\_estimators | criterion | **max\_features** | r2\_score |
| 1 | 50 | ***squared\_error*** | None | 0.84 |
| 2 | 50 | ***squared\_error*** | ***sqrt*** | 0.86 |
| 3 | 50 | ***squared\_error*** | log2 | 0.86 |
| 4 | 50 | ***absolute\_error*** | None | 0.85 |
| 5 | 50 | ***absolute\_error*** | ***sqrt*** | 0.87 |
| 6 | 50 | ***absolute\_error*** | log2 | 0.87 |
| 7 | 50 | ***friedman\_mse*** | None | 0.84 |
| 8 | 50 | ***friedman\_mse*** | ***sqrt*** | 0.87 |
| 9 | 50 | ***friedman\_mse*** | log2 | 0.87 |
| 10 | 50 | ***poisson*** | None | 0.82 |
| 11 | 50 | ***poisson*** | ***sqrt*** | 0.82 |
| 12 | 50 | ***poisson*** | log2 | 0.82 |

RF best r2\_score is : 0.87

**SVM and RF are the best algorithm to identify the Charge values.**

|  |  |  |  |
| --- | --- | --- | --- |
| poly | scale | 2000 | 0.86 |

|  |  |  |  |
| --- | --- | --- | --- |
| poly | auto | 2000 | 0.86 |