REGRESSION ASSIGNMENT

1. Problem Identification :

Stage 1: DOMAIN SELECTION ---Machine Learning - As dataset in numbers and excel format

Stage 2: LEARNING SELECTION --- Supervised Learning -As prediction from client (insurance charges) is clear and input and output is well defined.

Stage 3: REGRESSION OR CLASSIFICATION---Supervised Learning -Regression – As output is in form of numerical value.

1. Dataset Info:

It consist of 6 column with 5 input column (Age,bmi,sex,children,smoker) and one output column (charges)with multiple rows.

1. Preprocessing Method:

* In multiple linear regression to convert the string into numerical data used nominal data as it is without any comparsion and coded

Dataset=pd.get\_dummies(Dataset,dtype=int,drop\_first=True)

* In SVMR to increase the value r2 score used standardization preprocessing method and coded

from sklearn.preprocessing import StandardScaler

sc=StandardScaler()

x\_train=sc.fit\_transform(x\_train)

x\_test=sc.transform(x\_test)

1. Machine Learning Algorithms:
2. MULTIPLE LINEAR REGRESSION:

The R2\_score of multiple linear regression is 0.78.

2.SUPPORT VECTOR MACHINE:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hyper parameters© | R2\_score  Linear | Non-linear  Rbf | Poly | Sigmoid |
| C=0.01 | -0.08 | -0.08 | 0.08 | -0.08 |
| C=1.0 | -0.01 | -0.08 | 0.07 | -0.07 |
| C=10 | 0.46 | -0.03 | 0.03 | 0.03 |
| C=100 | 0.63 | 0.32 | 0.62 | 0.52 |
| C=1000 | 0.76 | 0.81 | 0.85 | 0.28 |

3.DECISION TREE:

|  |  |  |
| --- | --- | --- |
| Criterion | Splitter | R2\_score |
| Squared\_error | Best | 0.70 |
| Squared\_error | Random | 0.73 |
| Friedman\_mse | Best | 0.71 |
| Friedman\_mse | Random | 0.67 |
| Absolute\_error | Best | 0.67 |
| Absolute\_error | Random | 0.77 |
| poisson | Best | 0.71 |
| poisson | Random | 0.67 |

4.RANDOM FOREST:

|  |  |  |
| --- | --- | --- |
| Criterion | N\_estimators | R2\_score |
| Squared\_error | 100 | 0.85 |
| Squared\_error | 10 | 0.84 |
| Friedman\_mse | 100 | 0.85 |
| Friedman\_mse | 10 | 0.85 |
| Absolute\_error | 100 | 0.85 |
| Absolute\_error | 10 | 0.84 |
| poisson | 100 | 0.71 |
| poisson | 10 | 0.85 |

1. Final Model:

Finalising Random Forest Machine learning algorithm as my final model which will meet end user as it has high r2\_score value compared to others in all the parameters.(0.85)- r2\_score