# Assignment-Regression Algorithm

## 1.Problem Statement-

Client wants to predict insurance charges based on several parameters given in the data set.

Domain Selection – Machine learning.

Learning Selection-Supervised Learning-Regression.

## 2.Basic info about dataset-

Parameters – Age, Sex, BMI, Children, Smoker and charges.

Independent (Input)- Age, Sex, BMI, Children and Smoker.

Dependent (Output)- Charges

Rows-1339

Columns-6

## 3.Pre-Processing Method-

Parameter Age and Smoker has staring values we are converting nominal data to number data
Algorithm used-One hot Encoding

## 4.Model used and R2 score-

Machine learning algorithm used a for-insurance charges prediction and finalized model.

Algorithms used- Multilinear Regression, Random Forest, Decision Tree and Support Vendor Machine.

Finalized Model- Support Vendor Machine

R2 Score- 0.872

## 5.Documentation on R2scores.

Multilinear Regression- R2 Score- 0.789

```
                [11952.89779244]])

[72]: from sklearn.metrics import r2_score
      r_score=r2_score(y_test,y_pred)

[73]: r_score

[73]: 0.7894790349867009

[ ]:
```

Random Forest - R2 score-0.854

| S.NO | N_ Estimators | Random State | Criterion | R Value |
|---|---|---|---|---|
| 1 | 1 | 0 | Squared Error | 0.744 |
| 2 | 1 | 0 | Absolute Error | 0.706 |
| 3 | 1 | 0 | Friedman MSE | 0.75 |
| 4 | 1 | 0 | Poisson | 0.76 |
| 5 | 10 | 0 | Squared Error | 0.833 |
| 6 | 10 | 0 | Absolute Error | 0.835 |
| 7 | 10 | 0 | Friedman MSE | 0.833 |
| 8 | 10 | 0 | Poisson | 0.831 |
| 9 | 100 | 0 | Squared Error | 0.85 |
| 10 | 100 | 0 | Absolute Error | 0.852 |
| 11 | 100 | 0 | Friedman MSE | 0.854 |
| 12 | 100 | 0 | Poisson | 0.852 |

Support Vector Machine-R2 score-0.877

| S.No | Hyper Parameter | Liner R Value | RBF (Non Linear) R Value | Ploy R Value | SIGMOID R Value |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| 1 | C=10 | 0.462 | -0.03 | 0.03 | 0.039 |
| 2 | C=100 | 0.628 | 0.32 | 0.617 | 0.527 |
| 3 | C=500 | 0.763 | 0.664 | 0.826 | 0.444 |
| 4 | C=1000 | 0.764 | 0.81 | 0.856 | 0.287 |
| 5 | C=5000 | 0.741 | 0.874 | 0.859 | -7.53 |
| 6 | C=10000 | 0.741 | 0.877 | 0.859 | -34.15 |

Decision Tree-0.735

| S.No | Criterion | Splitter | R Value |
|---|---|---|---|
| 1 | Squarred error | Best | 0.725 |
| 2 | friedman mse | best | 0.735 |
| 3 | absolute error | best | 0.663 |
| 4 | Poisson | best | 0.701 |
| 5 | Squarred error | random | 0.652 |
| 6 | friedman mse | random | 0.723 |
| 7 | absolute error | random | 0.7 |
| 8 | Poisson | random | 0.754 |

## 5.Final Model -

Support Vector Machine-R2 score-0.877

Among trying all the algorithms Support vector Machine predicted the R value –0.877 which good for the dataset given.