

DOCUMENT ON insurance charges prediction using regression algorithms .

(A)PROBLEM STATEMENT:

STAGE 1- MACHINE LEARNING

STAGE 2- SUPERVISED LEARNING

STAGE 3- REGRESSION

(B) Total number of rows: 1338

Total number of columns: 6

(C) Changed string to number -smoker ,sex

(D)R2_score best model - Random Forest (0.906200)

To find the machine learning method R2_ value using R2_value:

a) Multiple linear regression - R2- value 0.78947.

B) Support vector machine :

	R_score	C=0.01	C=1	C=100	C=10
Linear	-0.09770	-0.09770	-0.09770	-0.09770	-0.09770
rbf	-0.09770	-0.09770	-0.09770	-0.09770	-0.09770
Poly	-0.09770	-0.09770	-0.09770	-0.09770	-0.09770
Sigmoid	-0.09770	-0.09770	-0.09770	-0.09770	-0.09770

C) Decision Tree

Criterion	Spliter	R2_value
Squared error	Random	0.080024
Friedman _mse	Random	0.080024
Possion	Random	0.080024



Edit with WPS Office

Absolute error	Random	0.080024
Squared error	Best	0.080024
Friedman _mse	Best	0.080024
Poisson	Best	0.080024
Absolute error	Best	0.080024

C) Random forest:

N_estimators	R2_value
10	0.906200
20	0.906119
30	0.906111
40	0.906100
50	0.906200
60	0.906200
70	0.906119
80	0.906118
90	0.906200
100	0.906200

Final the best method of machine learning Random Forest Regression algorithm.

Random Forest R2_value = 0.906200.



Edit with WPS Office