

REGRESSION ASSIGNMENT

[Click here to see the Dataset.](#)

PROBLEM IDENTIFICATION

STAGE 1:

DOMAIN SELECTION : Machine Learning

STAGE 2:

LEARNING SELECTION : Supervised Learning

STAGE 3:

CATEGORICAL SELECTION : Regression

BASIC INFORMATION ABOUT DATASET

PROJECT NAME : Insurance Charge Prediction

TOTAL NUMBER OF COLUMNS : 6

TOTAL NUMBER OF ROWS : 1338

NOMINAL COLUMN : Sex, Smoker

MULTIPLE LINEAR REGRESSION R^2 VALUE : 0.7891

SUPPORT VECTOR MACHINE

HYPER PARAMETER	LINEAR($R^2 V$)	RBF($R^2 V$)	POLY($R^2 V$)	SIGMOID($R^2 V$)
C100	0.6289	0.3196	0.6164	0.5268
C1000	0.7648	0.8107	0.8546	0.2120
C3000	0.7413	0.8646	0.8580	-2.1431
C4000	0.7413	0.8707	0.8587	-5.4667
C5000	0.7413	0.8737	0.8587	-8.1606

SVM Regression use R^2 value(RBF) and Hyper Parameter(C5000)= 0.8737

DECISION TREE

CRITERION	SPLITTER	MAX-FEATURES	R ² VALUE
Friedman_mse	random	log2	0.5136
Friedman_mse	best	log2	0.7000
Friedman_mse	random	sqrt	0.6588
Friedman_mse	best	sqrt	0.6585
Squared_error	random	log2	0.6731
Squared_error	best	log2	0.6746
Squared_error	random	sqrt	0.6644
Squared_error	best	sqrt	0.6872
Absolute_error	random	log2	0.7188
Absolute_error	best	log2	0.7286
Absolute_error	random	sqrt	0.6720
Absolute_error	best	sqrt	0.6974

DECISION TREE Regression use R² value(Absolute error,best,log2) = 0.7286

RANDOM FOREST

CRITERION	N_ESTIMATORS	MAX-FEATURES	R ² VALUE
Friedman_mse	10	sqrt	0.8531
Friedman_mse	100	sqrt	0.8639
Friedman_mse	10	Log2	0.8308
Friedman_mse	100	Log2	0.8645
Squared_error	10	sqrt	0.8415
Squared_error	100	sqrt	0.8654
Squared_error	10	Log2	0.8310
Squared_error	100	Log2	0.8637
poisson	10	sqrt	0.8391
poisson	100	sqrt	0.8626
poisson	10	Log2	0.8483
poisson	100	Log2	0.8665

RANDOM FOREST Regression use R² value(POISSON,N_ESTIMATORS=100,MAX_FEATURES=LOG2) = 0.8665

BEST MODEL : SVM

REASON: I GOT HIGHEST R²VALUE IN SVM FOR (insurance) DATASET