

 Karachi Institute of Economics and Technology <i>Collaboration with Pakistan Air Force</i>	College of Computing and Information Sciences		
	Assignment-2		
Type	Theory	Course Title	Intro to Data Science
Program	Intro. To Data Science	Campus / Shift	main
Date		Total Points	
Duration		Faculty Name	Dr. Affan Alim
Assigned Date	021-12-2021	Due date	

[Problem-1 ; Marks=]

Use the provided dataset titled "carprice.csv". Calculate the efficiency of models linear regression, decision tree, and KNN using RMSE, MSE, MAE, and R^2 .

	RMSE	MSE	MAE	R^2
Linear Regression				

0. Write the python code for the regression based models using Linear Regression
1. Discuss your results regarding the variance of error, and model acceptance.
2. Your model is overfitted, under-fitted, or in acceptance condition?

Note: Code will also be submitted in PDF format

[Problem-1 ; Marks=]

Consider the following dataset and fill the NaN using following step.

Step-1: Find correlation of (X3, X1), (X3,X2), and (X3,X4)

Step-2: Selection the high correlated attribute with X3.

Step-3: suppose X1 is the highly correlated with X3, then x3 will be target class and X1 is single attribute

Step-4: the NaN indices data will take as test data and remaining as train

Step-5: build a model using linear regression of training data

Step-6: predict the value of missing data using known attributes

X1	X2	X3	X4
6.5	8.6	NaN	4
1.2	5.0	5.2	9
4.0	2.0	8.7	5
3.2	3.6	7.3	1
5.2	7.1	NaN	3
7.3	2.0	6.4	8
5.5	4.1	NaN	5