SRN | | | | | | | | | | | |

PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS905**

August 2022: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER I

UE20CS905 - MACHINE LEARNING - I

Time: 3 Hrs	Answer All Questions	Max Marks: 80

Instructions

1. Answer all the questions.
2. Section A should be handwritten in the answer script provided.
3. Section B and C are coding questions which have to be answered in the system and uploaded in Olympus Login.
4. Smartly use GridSearchCV as it might impact the system performance.

Section A (20 marks)			
1	a)	Explain assumptions of linear regression.	4
	b)	Explain the relation between Bias and Variance in a linear regression model.	4
	c)	Discuss the need for data transformations in a linear regression model. Also, write about various transformation techniques.	4
	d)	How can regularization help in tackling overfitting?	4
	e)	A multiple regression equation was fit for n = 21 observations using 5 independent variables X1, X2, ,….., X5 gave SS(Total) = 1500 and SS(Residual) = 375. Calculate the value of the coefficient of determination and What do you conclude from the result?	4
Section B (30 Marks)			
2		Problem Statement: A Bike Rental firm is trying to understand predict the rental count on daily based on the environmental and seasonal settings. They want to utilize rational approach facilitated by machine learning in predicting the daily count. They collect a data set with different features along with the bike counts. The challenge is to learn a relationship between the important features and the bike count and use it to predict the future daily count. Develop a machine learning model to predict the bike count using the features provided in the dataset. Prior to build the ML model EDA need to carried out to understand and clean the data.	
	(i)	Read the dataset and perform the following. • Display the data types of all the features. (1 Mark) • Replace the numbers in month column to their respective names of month. (1 Mark) • Replace the date column with only the date and remove month and associated year. (1 Mark) • Typecast the relevant numerical attributes to category. (1 Mark)	4
	(ii)	Perform the following analysis to understand the data. • Convert the codes in season column to: 1:spring, 2:summer, 3:fall, 4:winter. (1 Mark)	4

SRN | | | | | | | | | | | | |

		• Convert the codes in weather column to: 1: Clear, 2: Mist ,3: Light Snow,4: Heavy Rain. (1 Mark) • Convert the codes in weekday column to:0:sunday, 1: monday, 2: tuesday, 3: wednesday, 4: thursday, 5: friday, 6:saturday. (2 Mark)	
	(iii)	Analyze the relationship between: --Month Vs demand of bike, --Weekday Vs demand of bike , --Holiday Vs demand of bike. Write some inference based on diagram.	4
	(iv)	Perform the following. --Use boxplot to visualize the outliers of numeric columns and Perform outlier elimination using IQR method. (3 Marks) --Display the percentage of missing values in each column and Implement a strategy to deal with the missing values. (3 Marks)	6
	(v)	Perform the following. --Plot a correlation plot and highlight the correlations through colour map. (3 Marks) --Display and remove the features showing high multi-collinearity by using VIF. (3 Marks)	6
	(vi)	Check the demand variable for normality and apply appropriate transformation.	6
Section C (30 marks)			
3	(i)	Perform the following: Drop all irrelevant features though the EDA performed in above steps. (2 marks) Perform relevant scaling on the numerical columns. (2 marks) Use encoding technique to encode the categorical variables. (2 marks)	6
	(ii)	Use OLS stat models package and build the Linear Regression model and also check for normality of residuals.	6
	(iii)	Write code to display significant features and build the model only with significant features	6
	(iv)	Use sklearn model for Linear Regression, compute evaluation parameters and comment about model overfitting nature.	6
	(v)	Validate models performance using Elastic net cross validation and print the RMSE. Also elaborate inferences.	6