SRN												

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

Question Paper
Executive Master of Business Administration

Linear Predictive Model

Time: 3 Hrs	Answer All Questions	Max Marks: 100

This is only for reference, the actual ESA paper will consist of 5 questions of 20 marks each.

INSTRUCTIONS

- All questions are compulsory.
- Part A should be handwritten in the answer script provided
- Part B and C are coding questions which have to be answered in the system.

Part A

1	a)	Consider the following model Y = 3.5+4*X1+3*X2 – X3 Predict Y for X1 =2, X2 =2 and X3 =1. Now, assuming all three variables are significant, which variable among these would provide the highest increase in Y and highest decrease in Y? Lastly, predict values when all three variables are zero.	5
	b)	Explain the type of supervised Machine Learning models with example.	5
	c)	The following is an output of the confusion matrix Predicted Class / Actual Class Positive / Positive: 150 Positive / Negative: 20 Negative / Positive: 10 Negative / Negative: 100 Based on the output, how many cases are correctly classified and how many are wrongly classified? Also, determine accuracy, precision and specificity.	5

SRN											

	d)	How can the problem of overfitting be reduced in Linear regression? Explain the different variants of Linear regression to solve the over fitting problem.	5
	e)	Write about the different variance measures involved in the operation of Linear Regression. List at least 5 assumptions of Linear Regression.	10
		Part B - 50 marks	
2	a)	Read the file and identify the categorical and numerical variable. Identify the missing values and see for any patterns using heatmap. Implement a strategy to deal with the missing value. Encode the categorical variables.	10
	b)	Create the train and test data set with suitable y and x variables. Develop a linear regression model based on the variables. Based on the output, identify which variables are significant and which are insignificant (use alpha = 0.05). Check for correlation among the numeric variables, based on the correlation accordingly decide which variables to be included, use correlation of 0.6 as cutoff. Based on the variables accordingly remove insignificant variables and reconduct analysis. Conclude accordingly using the output and provide which variables would affect the regression model.	30
	c)	Apply Lasso regression and determine the significant variables.	10
		Part C – 20 marks	
3	a	Consider the operating financial ratio data of 33 firms which went bankrupt after 2 years and the other 33 remained insolvent during the same period Datafile : bankruptcy.csv Perform below steps and build a logistic regression model -Read the data set and check for missing values -Split the data set into train and test -Analyze the data set based on the output, Use alpha = .05, to identify the significant variables, what do you conclude about the model -Calculate the odd for each of the variable and interpret the meaning	20