LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **DATA SCIENCE**

FIRST SEMESTER – **NOVEMBER 2022**

PDS1MC01 – FOUNDATIONS OF DATA SCIENCE

Date: 23-11-2022 Dept. No. Max. : 100 Marks

Time: 01:00 PM - 04:00 PM

SECTION - A			
Answer ALL the Questions			
1	**Answer the following**		**(5 x 1 = 5)**
a)	Write any two application of Data Science in healthcare domain.	K1	CO1
b)	List any two strategies used to handle missing data.	K1	CO1
c)	Define-Skewness	K1	CO1
d)	What is Insample evaluation.	K1	CO1
e)	Differentiate between MSE and RMSE.	K1	CO1
2	**Answer the following**		**(5 x 1 = 5)**
a)	What is the role of Business Analyst?	K2	CO1
b)	Write the formula of max-min to standardize the data.	K2	CO1
c)	How will you interpret the regression equation $y=200.3+4.15X_1+0.06X_2$	K2	CO1
d)	Differentiate between training set and test set.	K2	CO1
e)	What is hyper parametric tuning?	K2	CO1
SECTION - B			
	Answer any THREE of the following in 500 words		**(3 x 10 = 30)**
3	Write in detail the application of Data Science in any five domain.	K3	CO2
4	Explain Data pre-processing steps involved in Data Science.	K3	CO2
5	Elaborate how heatmap is applied in Data Science project	K3	CO2
6	Illustrate and explain any three regression models used in Data analysis .	K3	CO2
7	Elaborate Lasso regression model.	K3	CO2
	SECTION - C		
Answer any TWO of the following in 500 words			**(2 x 12.5 = 25)**
8	Discuss in detail the various Data security issues and measures.	K4	CO3
9	Describe in detail about Exploratory Data Analysis?	K4	CO3
10	Write the mathematical model, assumptions and uses of Simple linear regression model.	K4	CO3
11	Explain any two methods to solve overfitting issues.	K4	CO3

	SECTION - D		
	Answer any ONE of the following		**(1 x 15 = 15)**
12	Explain polynomial regression model. Also compare it with linear regression models.	K5	CO4
13	Write the procedure of solving two way ANOVA and discuss how it is used as a feature selection technique.	K5	CO4
	SECTION - E		
	Answer any ONE of the following		**(1 x 20 = 20)**
14	You are given a data set on cancer detection. You've built a classification model and achieved an accuracy of 96%. Why shouldn't you be happy with your model performance? Discuss as what you can do about it?	K6	CO5
15	You are given a data set. The data set contains many variables, some of which are highly correlated and you know about it. Propose the ways to handle such high dimensional data.	K6	CO5

@@@@@@