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

?>>>>?

PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS908**

Sept 2021: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER II

UE20CS908 - MACHINE LEARNING - II

Time: 3 Hrs	Answer All Questions	Max Marks: 100

Instructions

1. Answer all the questions.
2. Section A should be handwritten in the answer script provided and signed at the end of the same.
3. Section B and C are coding questions which have to be answered in the system and uploaded in Olympus Login.
4. Smartly use Grid SearchCV as it might impact the system performance.
5. Write appropriate inferences.

Section A (30 marks)			
1	a)	Explain K Nearest Neighbor Algorithm with an example.	5
	b)	i)Explain F1 Score metric with formula. ii)Calculate Precision metric based on below confusion matrix Predicted 0 Predicted 1 Actual 0 56850 4 Actual 1 30 78	5
	c)	Describe Ada boost and Gradient Boosting with differences and advantages	5
	d)	Explain Stacking Classifier and its advantages.	5
	e)	Describe working of Decision tree with an example.	5
	f)	Compare Decision tree and Random Forest	5
Section B (30 Marks)			
2		This data set contains details of a bank's customers and the target variable is a binary variable reflecting the fact whether the customer left the bank (closed his account) or he continues to be a customer. Develop a machine learning model to predict whether customer left the bank 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 print the following (4 marks) *Shape of the data * Number of numerical and categorical variable * Descriptive stats of numerical data	5

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

		* Descriptive stats of categorical data	
	(ii)	Perform appropriate encoding on 'Geography' and 'Gender' column.	5
	(iii)	Examine outliers by plotting and also by using z score. Examine is the Target variable evenly balanced.	5
	(iv)	Check for defects in the data like missing values, removing unnecessary features/columns. Perform necessary actions to 'fix' these defects.	6
	(v)	Examine the correlation and summarize the relationship between variables. Use appropriate plots to justify the same.	6
	(vi)	Split dataset into train and test (70:30)	3
Section C (40 marks)			
3	(i)	Fit a base model and explain the reason of selecting that model. Please write your key observations. Calculate Cohen Kappa value with the best model achieved.	15
	(ii)	How do you improve the accuracy of the model? Write clearly the changes that you will make before re-fitting the model. Fit the final model.	15
	(iii)	Summarize as follows 1.with respect to features 2.Evaluation metrics 3.Overall Results and Observations	10