JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA

PGDM / PGDM (M) / PGDM (SM)

FIFTH TRIMESTER (Batch 2020-22)

END TERM EXAMINATIONS, JANUARY, 2022, SET - 2

Course Name	MACHINE LEARNING (MLA)	Course Code	BA-501
Max. Time	**2 hours**	Max. Marks	**40 MM**

INSTRUCTIONS:

a. Students are required to work on their own personal Laptop. This is an open book exam
b. Attempt all the questions on a single Jupyter Notebook
c. The data for the exam can be downloaded from Moodle.
d. Write down your Roll no., course name and code on top of Jupyter Notebook
e. Save your Jupyter notebook with .ipynb extension and as pdf file
f. Upload both the files on Moodle.
g. Label the files as ML_roll no.

Q 1: Read the following case

Patient Satisfaction and Switching Behavior

Objective: To analyze the patient switching behavior using machine learning techniques and to figure out the possible reasons of switching the clinic by the patient.

ABC clinic has experienced a significant level of disenrollment; the director of the clinic wants to find out the factors which are associated with a patient's decision to switch clinics. A random sample of 495 patients from the clinic's population in the previous year was surveyed and data were collected using their responses to the questions. The data is given in an Excel file named **ABCKlinic.** The description is given below:

Variable	**Definition**
Patid	Unique patient identifier
Switchx	1 if person switched (dropped you as their designated primary care clinic), 0 if not
Age	Beneficiary's age (years)
Male	1 if male, 0 if female
Educ	1 if high school diploma or less, =2 some college, =3 Bachelor's degree 4: Graduate degree
Hlthstat	1 if excellent, =2 very good, =3 good, fair, or poor
Moveres	1 if patient moved home residence in past year, 0 if not
Hvdis	1 if person has a chronic disease, 0 if not
Havedoc	1 if there is one physician or care provider within this clinic that you think of as your personal doctor or nurse, 0 if not
Clinicrate	Scale of 0 (worst) to 10 (best), rating of the clinic
Appointment	In the past year, when you made an appointment for an office visit, how often did you get an appointment as soon as you needed? =1 never, =2 sometimes, =3 usually, =4 always.
Waitgt15	In the past year, how often did you have to wait more than 15 minutes to see a provider? =1 never, =2 if sometimes, =3 usually, =4 always; skipped for those who didn't have a doctor visit
Stafresp	In the last 12 months, how often did office staff at a doctor's office treat you with courtesy and respect? =1 never, =2 sometimes, =3 usually, =4 always; skipped for those who didn't have a doctor visit
Gethelp	In the last 12 months, when you phoned the provider's office during regular hours, how often did you get the help you needed? =1 never, =2 sometimes, =3 usually; 4: always

Analyze the dataset in python and answer the following questions:

1. Find out number of cases available in dataset for switchx = 1 and switchx = 0. Plot the findings using the bar plot. Is the data set balanced? (5 marks)
2. Balance the dataset by up sampling of switch = 1 cases. After up sampling the frequency count of switch = 1 cases should be 200. (3 marks)
3. Shuffle the balanced data after up sampling. (2 marks)
4. Split the balanced data into train and test set. (2 marks)
   (*Note: Random State value while splitting the data should be the last two digits of your roll no.)
5. Apply decision tree, random forest and any boosting classification algorithm. (10 marks)
6. Use the grid search mechanism to find most optimal values of the parameters for the two algorithms. (5 marks)
7. Find the best model on model accuracy score. (5 marks)
8. Find the feature importance from one of the models used above. And report the topmost important features. (5 marks)
9. What are your key learnings from the case? (3 marks)

Note: You are required to illustrate every output of the above questions with the interpretation mentioned as comments on the Jupyter notebook.