

JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA

PGDM / PGDM (M) / PGDM (SM)

FIFTH TRIMESTER (Batch 2021-23)

END TERM EXAMINATIONS, FEBRUARY 2023

Course Name	Machine Learning (Set 1)	Course Code	20287
Max. Time	2 Hours	Max. Marks	40 MM

INSTRUCTIONS:

- The mode of the paper is pen and paper
- The paper has two sections. Attempt both the sections.
- For section B, analyze the case in Jupyter Notebook on Python and answer the given questions on the answer sheet provided to you.

SECTION – A

- "The long term strength of Amazon is its ability to recognize customers one-on-one and effectively cross-sell across categories based on targeted recommendations." Demonstrate with the help of an example from e-commerce how cross-sell is a data analytics problem. **(5 Marks)**
- Match the machine learning methods with correct applications. **(5 Marks)**

Methods	Applications
A. Classification	a. A module that provides method to build confusion matrix
B. Clustering	b. Defining groups with similar behavior
C. Regression	c. Method to help with up-sampling
D. Sklearn.metrics	d. Predicting Salary of Employee
E. from sklearn.utils import resample	e. Predicting an email as Spam or Ham

- You were collecting data of students studying in primary college for your data analytics project. You later realized that your final data has some missing values and you cannot build data analytics model without fixing it. For the following data with seven rows, develop a method for treating the missing values (marked as XXX). Discuss your method and estimate the missing values **(5 marks)**

Student Name	Statistics Marks	Marketing Marks	Finance Marks
Alex	50	65	70
Maria	80	70	90
John	90	90	50
Jessica	85	XXX	92
Steve	45	42	49
Isabel	XXX	57	78
Sheldon	100	60	100

- Clustering is very popular decision-making technique of machine learning amongst the marketers. Discuss with the help of business use case how this method is useful in: **(5 marks)**
- Customer segmentation
- Product segmentation

SECTION – B

Read the following case

Employee Attrition

HR analytics is the process of gathering and analysing Human Resource (HR) data in order to improve an organization's workforce performance. In HR, attrition, is a term used to indicate voluntary departure of employees from a company. You are working as HR analysts for a Business Process Outsourcing (BPO) company. Your task is to classify which employee is likely to leave the company in near future, based on data collected about the employees. The "MLdata.csv" data set represents employee data of a BPO. The data description is given below:

Features	Description
Age_years	Age of an Employee as on Jan 01, 2023
Gender	Categorical (Female & Male)
Marital_status	Categorical (Single, Divorced & Married)
Work_loc	Categorical (Amritsar, Chandigarh, Gurgaon & Noida)
Buss_unit	Categorical (Fintech, Logistics & Telecom)
Tot_work_exp	Total Work Experience
Exp_curr_emp	Experience with current Employer
Per_Salary_Hike	Percentage Salary Hike in the last year
Work_environ_satis	Work Environment Satisfaction rating given by an employee. Where 1 indicates Extremely Dissatisfied and 5 indicates Extremely Satisfied
Perform_rating	Performance Rating given to an Employee. Where 1 is Extremely bad and 5 is extremely good
Dist_home	Distance from home
Training	No. of training/ courses completed in last one year
Attrition	Categorical (Yes & No)

Analyze the case using python and create a managerial report. You are required to prepare five slides (using pen and paper). Your report should present the following: (5 x 4 =20 marks)

1. Your observations from descriptive analytics
2. Model building process of the above business problem
3. Determine the number of True Positive, True Negative, False Positive and False Negative cases in both the techniques applied.
4. Which model is the best model for prediction? Why?
5. What is the need for hyper-parameter tuning?
6. The managerial implications of the case.