****

**JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA**

**PGDM / PGDM (M) / PGDM (SM)**

**FIFTH TRIMESTER (Batch 2022-24)**

**END TERM EXAMINATIONS, JANUARY 2024**

**SET – 1 (Group 2)**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Name | Machine Learning | Course Code | **20827** |
| Max. Time | **2 hours** | Max. Marks | **40 MM** |

**INSTRUCTIONS:**

1. There are two sections in the question paper. Both are compulsory to attempt.
2. It’s a pen and paper exam. Answer all questions on the answer sheet provided to you.
3. For Section B refer to the Annexure given with the question paper.

**Section A**

1. Read the following case let and answer the questions given at the end of the case:

**Case: The Dream Employee**

A group of MBA students from management institute in Delhi attended a conclave on “Demystifying Artificial Intelligence”. The event featured a panel of experts specializing in artificial intelligence, machine learning, data science, and analytics who shared insights into their experiences and ongoing projects.

One of the sessions focused on exploring the applications of Machine Learning in the e-commerce sector. The panel experts discussed various aspects, including the significance of analyzing digital information for effective decision-making. The experts from, marketing, finance, HR and operations managers addressed the relevance of these technologies in their respective domains.

They reiterated in many ways that a **dream employee** is one who has the domain expertise, which comes with time and experience, can look at varied sources of data to suggest means for filtering out relevant information, is able to suggest the analytical techniques to process the data and finally read and communicate the outcomes for its appropriate use in decision-making. The experts acknowledged that finding a resource possessing all these qualities is challenging. However, they stressed that in the future as we move towards a paperless environment, such muti-talented individuals will be highly valued. Therefore, efforts should be made to cultivate and enhance these capabilities.

**Answer the following questions based on the above case let:**

1. Recommend the kind of capabilities and the skills to be possessed by an individual wanting to join the Supply Chain Analytics department of an e-commerce company. **(5 Marks)**
2. Discuss two opportunities of e-commerce business which can be resolved by applying machine learning algorithm. **(5 Marks)**
3. Techsolution, a call center in Noida, sent selected employees for a training program to enhance their efficiency and productivity. After the training the productivity of employees was measured by estimating the number of leads converted in the last month. The data is provided in the table below. Upon analysis, the team of analysts identified an outlier in the dataset. Consequently, it is imperative to address this issue before constructing a machine learning model. Your task is to develop a method for handling the outlier value. Elaborate on your chosen method for treating outlier values using the given data, which consists of seven rows. (**5** **marks**)

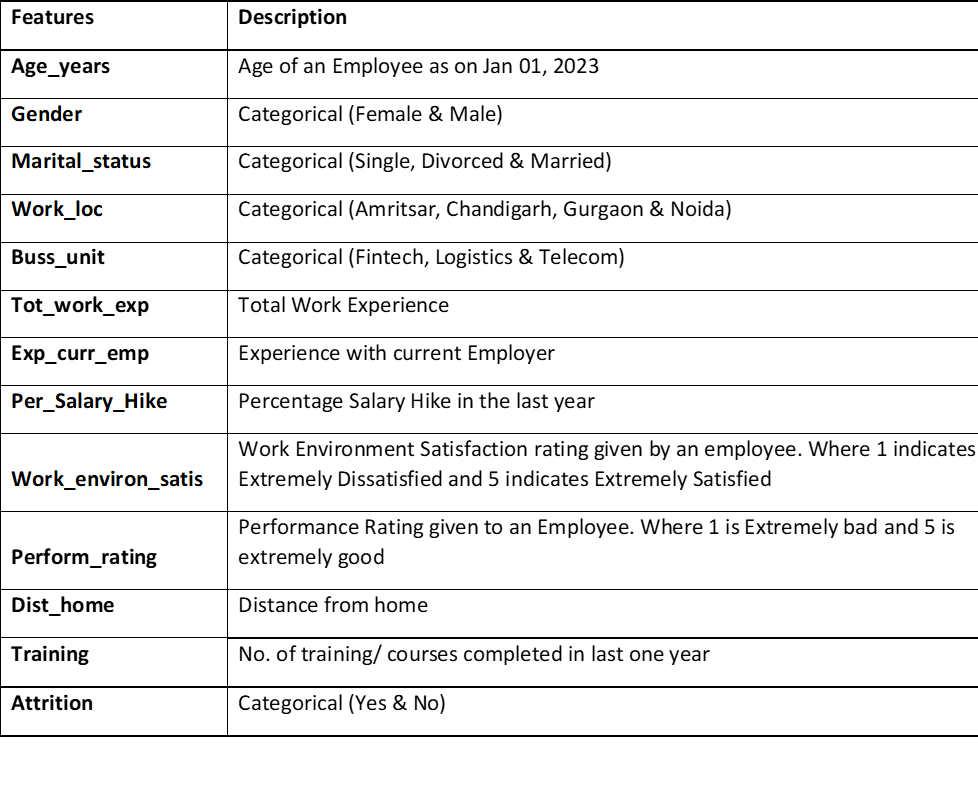
|  |  |
| --- | --- |
| **Employee** | **Productivity** |
| Sonia | 25 |
| Kanika | 17 |
| Rahul | 19 |
| Jessica | 90 - outlier |
| Mohit | 27 |
| Rohit | 17 |
| Sachin | 22 |

**Section B**

1. Read the following case:

**Employee attrition Case**

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 machine learning analyst for a Business Process Outsourcing (BPO) company. Your task is to classify which employee is likely to leave the company in the near future, based on data collected about the employees. The data description is given below:



Analyze the case using python codes and output enclosed with the question paper and answer the following questions:

Q1. Discuss why this is a classification problem. Compare it with other types of machine learning problems.

**(5 marks)**

Q2. Estimate the number of numeric features given for the above problem? **(1 mark)**

Q3. Which region has the maximum attrition rate? Justify by giving the valid reason. **(2 marks)**

Q4. Does distance from home impact attrition? Estimate the value of median distance from home.  **(2 marks)**

Q5. Refer to the confusion matrix of K-NN model and determine the following values:

1. True Positive (TP) - \_\_\_\_\_\_\_\_\_\_
2. False Positive (FP) - \_\_\_\_\_\_\_\_\_\_
3. False Negative (FN) - \_\_\_\_\_\_\_\_\_\_
4. True Negative (TN) - \_\_\_\_\_\_\_\_\_\_
5. Accuracy - \_\_\_\_\_\_\_\_\_\_

Further, discuss how the values of precision, recall and f1-score are calculated and their relevance in context to the business problem. **(5 marks)**

**Q6.** Highlight the best algorithm for this problem? **(2 marks)**

**Q7.** Determine the optimal parameters of the best algorithm. **(2 marks)**

**Q8.** What does cv = 10 indicate in code block [85]. Discuss in detail its importance in machine learning? **(3 marks)**

**Q9**. Discuss graphically how the K-NN technique works for classifying the dependent variable. You may choose any business case of your choice. **(3 marks)**