****

**JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA**

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

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

**END TERM EXAMINATIONS, JANUARY 2024**

**SET – 2**

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

**INSTRUCTIONS:**

1. Attempt all the questions.
2. It’s a closed book examination.
3. Suppose you have applied for an internal job promotion in your current organization for the position of Machine Learning Analyst. The interview panel has asked you to explain the following with the help of business scenarios: **(10 Marks = 2.5\*2)**
4. Overfitting versus underfitting in machine learning
5. Classification versus clustering algorithms
6. Balanced versus unbalanced problem.
7. Decision-tree versus random forest algorithm.

Write down your response for the questions asked.

**Q2.** You are a data scientist working with mid-sized company offering online courses located in Pune, India. The company has recognized the need to tackle lead conversion issue, i.e. enrolling students for their courses. You are at the forefront of their efforts, with the background of machine learning. You are tasked with harnessing the power of data to predict and address lead conversion. **(20 Marks = 5\*4)**

1. Using ML technique, describe the methodology you would employ to analyze the said problem within the Organization.
2. What data sources and features would you utilize, and what algorithms and models would you consider for predicting turnover?
3. Following the analysis in the above questions, what data driven strategies and recommendations would you propose to reduce employee turnover and enhance retention within the organization?
4. How would these strategies benefit from the application of machine learning?

**Q3.** LargeBazaar is seeing a significant number of customer churn since past six months. The store owner is worried and asked the research team to develop a predictive model to predict the churn behavior of the customers. The team developed and trained a classification model to know the significant predictors of churn behavior. The results were then validated using the test data. The confusion matrix of a test data is given below. Note that **‘No’** indicates **negative** case and **‘Yes’** indicates **positive** case.

A graph of a number of labels

Description automatically generated with medium confidence

1. Explain and Determine the values of **(10 marks)**
2. True Positive (TP) - \_\_\_\_\_\_\_\_\_\_
3. False Positive (FP) - \_\_\_\_\_\_\_\_\_\_
4. False Negative (FN) - \_\_\_\_\_\_\_\_\_\_
5. True Negative (TN) - \_\_\_\_\_\_\_\_\_\_
6. Accuracy - \_\_\_\_\_\_\_\_\_\_
7. Recall (P) – \_\_\_\_\_\_\_\_\_\_
8. Precision (P) – \_\_\_\_\_\_\_\_\_\_
9. F1-score (P) - \_\_\_\_\_\_\_\_\_\_