|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |

# PRESIDENCY UNIVERSITY BENGALURU

# SET B

**Semester :** Semester V

**Course Code :** BBB3004 **Course Name :**Machine Learning

# SCHOOL OF MANAGEMENT

**MAKE-UP EXAMINATION JULY 2024**

**Date :** 03 JULY 2024

**Time :** 9:30AM - 12:30 PM

## Max Marks : 100

**Program :** BBA - Business Analytics **Weightage :** 50%

## Instructions:

1. *Read all questions carefully and answer accordingly.*
2. *Question paper consists of 3 parts.*
3. *Scientific and non-programmable calculator are permitted.*
4. *Do not write any information on the question paper other than Roll Number.*

**PART A**

**ANSWER ALL THE QUESTIONS 5 X 2M = 10M**

* 1. Write any two major applications using Machine Learning tools.
	2. Write the code for creating a 2x3 matrix in Python

(CO1) [Knowledge] (CO2) [Knowledge]

* 1. Define the primary purpose of Principal Component Analysis (PCA) in simplifying data, and how does it achieve dimensionality reduction?

(CO3) [Knowledge]

* 1. Describe the main function of a histogram, and how does it assist in understanding data patterns?

(CO4) [Knowledge]

* 1. What is the primary purpose of Logistic Regression, and how does it differ from linear regression?

(CO5) [Knowledge]

**PART B**

**ANSWER ALL THE QUESTIONS 5 X 10M = 50M**

* 1. Subject: Predicting Customer Churn in a Telecommunication Company.

A leading telecommunication company is experiencing high customer churn rates, which is negatively impacting its revenue and market share. They aim to reduce churn by identifying customers at risk of leaving and implementing targeted retention strategies. To address this, the company decides to create a machine learning model for predicting customer churn.

Object: Explain the general process of creating a machine learning model, starting from data collection to model deployment, in a few sentences.

(CO1) [Comprehension]

* 1. The local electricity provider charges consumers based on their electricity usage. The billing rates are as follows:

For consumers using 100 units or less, the rate is Rs. 1.50 per unit.

For consumers using between 101 and 200 units (inclusive), the rate is Rs. 2.50 per unit. For consumers using more than 200 units, the rate is Rs. 5.00 per unit.

Calculate the electricity bill for the following consumers, based on their electricity usage:

|  |  |
| --- | --- |
| **Consumer Name** | **Electricity Usage (in units)** |
| Kiran Kumar | 85 |
| Lalit | 150 |
| Mohan | 250 |
| Raju | 850 |

Calculate the total bill for each consumer based on the given conditions and provide the output in separate Total Bill Column.

(CO2) [Comprehension]

* 1. Explain the concept of normalization in the context of databases and data management.

(CO3) [Comprehension]

* 1. Describe how Matplotlib contribute to data visualization in Python, and what are its main features that make it widely used for creating different types of charts and plots?

(CO4) [Comprehension]

* 1. Describe Decision Tree algorithm, and what are the key steps involved in the decision-making process?

(CO5) [Comprehension]

**PART C**

**ANSWER ALL THE QUESTIONS 2 X 20M = 40M**

* 1. Elucidate the following syntaxes under dyplr() function, providing a comprehensive description of each along with practical applications.

filter() mutate() group\_by()

summarize() arrange()

(CO3) [Application]

* 1. You are provided with a dataset containing information about individuals' salaries and various features, including education level, years of experience, and the bank where they hold an account. Your task is to use a Random Forest regression model to predict salary levels based on these features.

Dataset:

The dataset includes the following features:

Education Level Years of Experience Bank Salary

|  |  |  |  |
| --- | --- | --- | --- |
| High School | 3 | A | 50,000 |
| Bachelor's | 5 | B | 70,000 |
| Master's | 8 | C | 90,000 |
| Ph.D. | 10 | A | 120,000 |
|  |  |  |  |
| Education Level |  |  | Count |
| High School |  |  | 100 |
| Bachelor's |  |  | 250 |
| Master's |  |  | 150 |
| Ph.D. |  |  | 50 |
|  |  |  |  |
| Bank |  |  | Count |
| A |  |  | 180 |
| B |  |  | 120 |
| C |  |  | 250 |
| Questions: |  |  |  |

* + 1. Briefly describe the dataset, including the types of features and the distribution of salary levels.
		2. Explore the distribution of education levels and the number of individuals associated with each bank.
		3. Evaluate the model's performance using appropriate metrics (e.g., Mean Absolute Error, R- squared).

(CO5) [Application]