**MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING**



(A Unit of Rajalaxmi Education Trust®, Mangalore)

Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE, New Delhi

Accredited by NAAC with A+ Grade & ISO 9001:2015 Certified Institution

**Model Question Paper**

**Fourth Semester MBA Degree Examination**

Machine Learning with Business Applications

**Time: 3 Hours Max. Marks: 100**

***Note: 1. Answer any FOUR full questions from Q1 to Q7.***

***2. Question No. 8 is compulsory.***

***3. M: Marks, L: RBT (Revised Bloom’s Taxonomy) level, C: Course outcomes****.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | **M** | **L** | **C** |
| Q1 | a. | Illustrate how machine learning differs from traditional programming. | 03 | L2 | CO1 |
| b. | Compare supervised, unsupervised, and reinforcement learning with suitable business examples for each. | 07 | L2 | CO1 |
| c. | Explain how machine learning is transforming sales and marketing functions. Give scenarios in at least two industry verticals. | 10 | L3 | CO1 |
| Q2 | a. | Illustrate the difference between a probabilistic model (e.g., Naive Bayes) and a geometric model. | 03 | L2 | CO1 |
| b. | Summarize the end-to-end workflow of a machine learning project and highlight where strategic business decisions are crucial. | 07 | L2 | CO1 |
| c. | Outline how can machine learning enhance customer experience and drive revenue in e-commerce? How machine learning application can be used in fraud-detection or dynamic pricing in e-commerce? | 10 | L3 | CO4 |
| Q3 | a. | Outline the difference between a confidence interval and a prediction interval in linear regression. | 03 | L2 | CO2 |
| b. | Explain how Deviance, Wald test and Hosmer Lemeshow test help in evaluating how best the logistics model fits the requirement. | 07 | L2 | CO2 |
| c. | Summarize the process of feature selection in multiple linear regression. Why it is important? What methods can be used? Explain with an example. | 10 | L2 | CO2 |
| Q4 | a. | Summarize the role of feature importance scores in Random Forest and how they can be used for decision-making. | 03 | L2 | CO2 |
| b. | A company uses Gradient Boosting to predict B2B sales conversion probability. Describe how boosting differs from bagging in improving model performance. | 07 | L3 | CO2 |
| c. | Give a business scenario where it would require building a classification model using both Logistic Regression and Random Forest. Outline how you would compare and choose the better-performing model. | 10 | L3 | CO2 |
| Q5 | a. | Illustrate how an ARIMAX model differ from ARIMA, and in what business situations would you choose ARIMAX over ARIMA? | 03 | L2 | CO3 |
| b. | Explain how a moving average method smooths time series data. Give an example in retail demand forecasting. | 07 | L2 | CO3 |
| c. | A supermarket chain observes weekly demand fluctuations for ice cream with clear peaks during summer. Describe how you would model this series, specifying which components (trend, seasonality) you would account for. | 10 | L3 | CO4 |
| Q6 | a. | Outline the role of the distance metric (e.g., Euclidean, Manhattan) in clustering algorithms. | 03 | L2 | CO3 |
| b. | Explain the trend, cyclical, and seasonal components of a time series. How can identifying this help in business forecasting? | 07 | L2 | CO2 |
| c. | Explain how logistic regression can help predict employee attrition and guide HR policies. | 10 | L3 | CO4 |
| Q7 | a. | Compare K-means clustering and hierarchical clustering in terms of approach, scalability, and output. | 03 | L2 | CO3 |
| b. | A retail chain wants to group its stores based on sales performance, product mix, and customer demographics. Describe how you would approach this problem using K-means. | 07 | L3 | CO3 |
| c. | An e-commerce platform wants to group products based on purchase patterns and customer reviews. Suggest a clustering approach and explain your reasoning. How would you validate whether the clusters found are meaningful? | 10 | L3 | CO3 |
| Q8 | a.  b.  c.  d. | CASE STUDY (Compulsory)  A retail chain has collected the following data on monthly advertising spend (in ₹ lakhs) and monthly sales revenue (in ₹ lakhs) for format:   | Month | Ad Spend (X) | Sales Revenue (Y) | | --- | --- | --- | | 1 | 2.0 | 24 | | 2 | 2.5 | 28 | | 3 | 3.0 | 35 | | … | … | … |   Write a Python script, to fit a simple linear regression model with Ad Spend as the independent variable and Sales Revenue as the dependent variable.  Write an example regression equation for this given scenario. Using this equation, calculate the sales if ad spend is Rs. 8 lakhs. Give an example coefficient of determination (R²) for the model. What conclusion business can draw based on the score you have taken?  A t-test for the slope was conducted. Write Null Hypothesis and Alternate Hypothesis for this given scenario. The p-value was found to be 0.0002. What would you conclude based on the same? Also form a sample confidence interval and prediction interval.  Further the retail chain also would like to check the impact of product price and inflation on sales. Suggest a multiple linear regression model to handle this scenario. | 5  5  5  5 | L4 | CO4 |

-\*\*\*\*\*-