

**INSTITUTE OF AERONAUTICAL ENGINEERING**

(Autonomous)

Dundigal-500043, Hyderabad

B.Tech VII SEMESTER END EXAMINATIONS (REGULAR) - DECEMBER 2023

Regulation: UG-20

PREDICTIVE DATA ANALYTICS

**Time: 3 Hours**    **COMPUTER SCIENCE AND ENGINEERING (AI & ML)**    **Max Marks: 70****Answer ALL questions in Module I and II****Answer ONE out of two questions in Modules III, IV and V****All Questions Carry Equal Marks****All parts of the question must be answered in one place only****MODULE – I**

1. (a) Interpret the results by conceiving a predictive analytics model for preventing employee attrition. [BL: Understand| CO: 1|Marks: 7]
- (b) Apply the steps to assess the feasibility of predictive analytic solution and show the results for “Motor insurance fraud project”. [BL: Apply| CO: 1|Marks: 7]

**MODULE – II**

2. (a) List the features of advanced data exploration techniques. Explain the concept of binning and sampling in data preparation. [BL: Understand| CO: 2|Marks: 7]
- (b) Discuss about the type of data used for predictive analytics with example. Compare observation period and outcome period in prediction models. [BL: Understand| CO: 2|Marks: 7]

**MODULE – III**

3. (a) Outline the principle of similarity based learning and construct a predictive model. Indicate the features of ensemble based learning. [BL: Understand| CO: 3|Marks: 7]
- (b) Apply the principle of decision tree induction and construct the steps for iterative Dichotomizer-3 algorithm. [BL: Apply| CO: 3|Marks: 7]
4. (a) Explain the concept of entropy in terms of uncertainty or disorder in a probability distribution. [BL: Understand| CO: 4|Marks: 7]
- (b) Assume that you have an image of a creature that looks similar to cat and dog, but you want to know either it is a cat or dog. So for this identification, interpret the best possible algorithm that works on a similarity measure and construct the steps. [BL: Apply| CO: 4|Marks: 7]

**MODULE – IV**

5. (a) Explain how Naïve Bayes classifier differs from the traditional Naïve Bayes model? Elaborate the reasons for using Naïve Bayes classifier for multiclass problems. [BL: Understand| CO: 5|Marks: 7]
- (b) Build a Naïve Bayes classifier model by using a synthetic dataset that encompasses weather conditions for playing a game of cricket. [BL: Apply| CO: 5|Marks: 7]

6. (a) What is factorization in the context of probability distributions? Explain the concept of binning applied in probability based learning. [BL: Understand| CO: 5|Marks: 7]
- (b) Discuss the role of smoothing in probability based learning models. Implement probability density function using open source programming language. [BL: Understand| CO: 5|Marks: 7]

### MODULE – V

7. (a) List out the assumptions made in linear regression. Explain in detail about multiple linear regression. [BL: Understand| CO: 6|Marks: 7]
- (b) Consider the following example: An organization wants to determine an employee's salary increase based on their performance. Also the organization wants to know whether an employee would get a promotion or not based on their performance. Apply logistic regression and implement the solution steps. [BL: Apply| CO: 6|Marks: 7]
8. (a) Describe the process of hypothesis testing in simple linear regression, including the role of p-values and confidence intervals. [BL: Understand| CO: 6|Marks: 7]
- (b) Compare and contrast the methods of error-based and non-error based learning. Analyze the implication of “learning rate” in Gradient descent algorithm. [BL: Understand| CO: 6|Marks: 7]

