**UPES**

**End Semester Examination, May 2024**

**Course:** Overview of Data Mining **Semester:** II

**Program:** MBA BA KPMG **Time: 03 hrs.**

**Course Code:** DSBA7011P **Max. Marks: 100**

**Instructions:**

**SECTION A**

**10Qx2M=20 Marks**

| S. No. | | Marks | CO |
|---|---|---|---|
| Q 1 | Statement of question | | |
| a. | What are the functions of Data Mining?<br>i. Association and correctional analysis classification<br>ii. Prediction and characterization<br>iii. Cluster analysis and Evolution analysis.<br>iv. All of the above | **2** | **CO1** |
| b. | What is the full form of OLTP?<br>i. Online transaction processing<br>ii. Offline transaction processing<br>iii. Online traffic processing<br>iv. None of the above | **2** | **CO1** |
| c. | List out the applications of data mining. | **2** | **CO1** |
| d. | How data mining differs from the databases. | **2** | **CO1** |
| e. | Compare OLAP and OLTP. | **2** | **CO1** |
| f. | Name the steps involved in data mining? | **2** | **CO1** |
| g. | State the applications of clustering. | **2** | **CO1** |
| h. | List the weaknesses of k-means. | **2** | **CO1** |
| i. | Neural networks are modeled on the _____.<br>i. neuron.<br>ii. network.<br>iii. human brain.<br>iv. machines. | **2** | **CO1** |
| j. | Which of the following is the not a type of clustering?<br>i. K-means.<br>ii. Hierarchical.<br>iii. Partitional.<br>iv. Splitting. | **2** | **CO1** |

**SECTION B**

**4Qx5M= 20 Marks**

| | | | |
|---|---|---|---|
| Q 2 | Statement of question | | |
| a. | List and explain the various activation functions used in modeling of artificial neuron. | **5** | **CO2** |
| b. | Explain different data pre- processing techniques. | **5** | **CO2** |
| c. | What is the role of nearest neighbor classifier? Explain it briefly? | **5** | **CO2** |

| | | | |
|---|---|---|---|
| d. | Explain the Knowledge Discovery in Databases (KDD) process with examples. | **5** | **CO2** |
| **SECTION-C<br>3Qx10M=30 Marks** | | | |
| Q 3 | Statement of question | | |
| a. | Explain in detail about the categories of major clustering methods. Explain the use of the elbow method for determining the optimal number of clusters in K-means clustering. | **10** | **CO2** |
| b. | Discuss the multifaceted goals and objectives of data mining, considering its potential to uncover hidden patterns, predict future trends, segment customer demographics, and drive informed decision-making across diverse industries. | **10** | **CO2** |
| c. | Explain the different models for multidimensional analysis.<br>OR<br>Explain the architecture of pre trained CNN Models. | **10** | **CO2** |
| **SECTION-D<br>2Qx15M= 30 Marks** | | | |
| Q 4 | Statement of question | | |
| a. | Given a dataset of monthly sales data for a retail store, how would you apply multiple linear regression to forecast future sales based on factors such as advertising spend, promotions, and seasonality? | **15** | **CO3** |
| b. | You have a dataset with categorical variables encoded as text labels. How would you perform one-hot encoding or label encoding to convert these categorical variables into numerical format for machine learning algorithms?<br>OR<br>Illustrate feature scaling and its significance in preparing heterogeneous datasets for analysis. You have a dataset with numerical features that have different scales. How would you perform feature scaling to ensure that all features have similar ranges or distributions? | **15** | **CO3** |