

|  |  |  |
| --- | --- | --- |
| **Semester: Dec-2024**  **Maximum Marks: 50 Examination: End Exam Date: 15/12/24 Duration: 2.5 Hours** | | |
| **Programme code: 18**  **Programme: MBA Working Executive Batch 1** | **Class: SY** | **Semester/Trimester: III**  **Batch- 2023-24** |
| **College:**  **K. J. Somaiya Institute of Management** | **Name of the department/Section/Center: Business Analytics** | |
| **Course Code:** | **Name of the Course: Application Based Analytics** | |
| **Instructions:**   1. **Attempt any 5 questions. All questions carry equal marks.** 2. **Make suitable assumptions if required and state them.** 3. **Use Excel if required and keep saving the file every ten minutes or so.** 4. **Use of a calculator is permitted.** | | |

|  |  |  |
| --- | --- | --- |
| **Question No.** |  | **Max.**  **Marks** |
| Q1 | Based on the following sociogram:  E  A  F  C  B  D     1. Calculate the Degree and Closeness Centrality for all the nodes and identify the most influential node(s). 2. Calculate Betweenness Centrality for nodes A and D. | 10 |
| Q2 | |  |  | | --- | --- | | **Transaction ID** | **Items** | | T1 | Hot Dogs, Buns, Ketchup | | T2 | Hot Dogs, Buns | | T3 | Hot Dogs, Coke, Chips | | T4 | Chips, Coke | | T5 | Chips, Ketchup | | T6 | Hot Dogs, Coke, Chips |   Use an appropriate algorithm to find frequent items bought together by using minimum support of 2 and 3. Show step-by-step iteration. | 10 |
| Q3 | 1. Which of the following (a or b) will you choose as a feature to detect the churn and why?     a    b   1. Differentiate between: 2. Eigenvector Centrality and PageRank Centrality. 3. Weighted and Unweighted Networks. | 5 + 5 |
| Q4 | For the following directed sociogram:  B  A  D  C  Calculate the overall Degree Centrality, In-Degree, and Out-Degree, Closeness Centrality and Betweenness Centrality for all the nodes. Identify the most influential node(s) based on each centrality. | 10 |
| Q5 | Explain the following Python codes and identify the business problem(s) where the codes are used:   |  |  |  |  | | --- | --- | --- | --- | |  | Code | Explanation | Business Problem | | a | G\_asymmetric = nx.DiGraph()  G\_asymmetric.add\_edge('A', 'B')  G\_asymmetric.add\_edge('A', 'D')  G\_asymmetric.add\_edge('C', 'A')  G\_asymmetric.add\_edge('D', 'E') |  |  | | b | Q1 = df.Number\_of\_cards.quantile(0.25)  Q3 = df.Number\_of\_cards.quantile(0.75)  df\_outlier = df[(df.Number\_of\_cards <Q1 - 1.5 \* (Q3-Q1))| (df. Number\_of\_cards > Q3 + 1.5 \* (Q3-Q1))]  df\_outlier |  |  | | c | from sklearn import preprocessing  le = preprocessing.LabelEncoder()  for i in categorical:  df[i]=le.fit\_transform(df[i]) |  |  | | d | frequent\_items = apriori(my\_basket\_sets.astype('bool'), min\_support = 0.01, use\_colnames = True)  frequent\_items |  |  | | e | lr = LogisticRegression(max\_iter = 1000)  lr.fit(x, y)  y\_pred = lr.predict(x)  lr |  |  | | 10 |
| Q6 | Consider a Big Bazar scenario where the item set is I = {Milk, Egg, Bread, Butter, Ketchup, Cookies}. The database comprises twelve transactions is given below:   |  |  | | --- | --- | | **Transaction ID** | **Items** | | t1 | Milk, Egg, Bread, Butter | | t2 | Milk, Egg, Butter, Ketchup | | t3 | Bread, Butter, Ketchup | | t4 | Milk, Bread, Butter | | t5 | Bread, Butter, Cookies | | t6 | Milk, Bread, Butter, Cookies | | t7 | Milk, Cookies | | t8 | Milk, Bread, Butter | | t9 | Egg, Bread, Butter, Cookies | | t10 | Milk, Bread, Butter | | t11 | Milk, Bread, Butter | | t12 | Milk, Bread, Ketchup, Cookies |   Considering the following association rules:   |  |  | | --- | --- | | **Association Rules** | | | **Antecedent** | **Consequent** | | {Milk, Bread} | {Butter} | | {Cookies} | {Milk, Bread} | | {Egg} | {Bread} |   Compute support, confidence, and lift for the association rule Antecedent → Consequent and identify the frequent itemset based on the minimum lift threshold. | 10 |