

|  |
| --- |
| **Semester: Jan-Mar 25****Maximum Marks: 50 Examination: ETE Exam Date: 3rd April 2025 Duration: 3 Hrs** |
| **Programme code: 01****Programme: MBA BADS Minor** | **Class:** SY | **Semester/Trimester: VI** |
| **College:**  **K. J. Somaiya Institute of Management** | **Name of the department/Section/Center: Business Analytics** |
| **Course Code: 217P01M625**  | **Name of the Course: Application Based Analytics**  |
| **Instructions:** 1. **Attempt 5 questions in all.**
2. **All questions carry equal marks.**
3. **Make suitable assumptions if required and state them.**
4. **Use of a calculator is permitted.**
 |

|  |  |  |
| --- | --- | --- |
| **Question No.** |  |  **Max.****Marks** |
| 1 | **SocioNet**, a growing social media platform, is analyzing how viral content spreads. The following **network represents the spread of information and communication** among five key influencers: **Alice, Brian, Chris, Diana, and Emma.** SocioNet wants to determine who plays the most critical role in content virality and influence. 1. Which influencer has the widest reach? Which centrality measure have you used and why?
2. Who spreads information the fastest? Which centrality measure have you used and why?
3. To understand which influencers control the flow of information in the network, which centrality measure should be used? Justify your choice. Compute the centrality score for Chris and Brian.
4. Which centrality measure would you use to find an influencer connected to more influential people in the network?
 | 10 |
| 2 | FreshMart, a leading grocery chain, wants to optimize its product bundling strategy by understanding customer purchasing behavior. The store collects transaction data, where each record represents items bought together. The following sample dataset contains five transactions:

|  |  |
| --- | --- |
| **Transaction ID** | **Items Purchased** |
| T1 | Bread, Milk |
| T2 | Bread, Diaper, Beer, Eggs, Milk |
| T3 | Diaper, Beer, Coke |
| T4 | Bread, Milk, Diaper, Beer |
| T5 | Bread, Milk, Diaper, Coke |

1. Use an appropriate algorithm to find frequent items bought together by using minimum support of 2 and 3. Show step-by-step iteration.
2. Which algorithm have you used and why?
 | 10 |
| 3 | 1. Interpret the following output.
2.
3.
4. Write short notes on any two of the following:
5. Degree centrality in directed and undirected networks
6. Different metrics used in Apriori algorithm
7. Univariate and multivariate anomaly detection
 | 5 + 5 |
| 4 | Explain the following Python codes and identify the business problem(s) where the codes are used:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Code | Explanation | Business Problem |
| a | fraud = df[df['Class'] == 1] valid = df[df['Class'] == 0] print('Fraud Cases =', len(fraud))print('Valid Transactions =', len(valid)) |  |  |
| b | basket = df.groupby(['Order', 'Description'])['Description'].count().reset\_index(name ='Count') basket |  |  |
| c | from sklearn.feature\_selection import RFECVrfecv = RFECV (estimator = LogisticRegression (max\_iter = 1000), cv=10)rfecv.fit(x, y) |  |  |
| d | pos = nx.spring\_layout(G\_net)#8 communities detectedfor i in range(0,8): community\_i = sorted(c[i]) nx.draw\_networkx\_nodes(G\_net, pos, nodelist  = community\_i) |  |  |
| e | text = \_arg1scores = [] sid = SentimentIntensityAnalyzer()for word in text: ss = sid.polarity\_scores(word) scores.append(ss['compound'])return scores |  |  |

 | 10 |
| 5 | TechFlow Inc., a leading software company, is analyzing how information flows within its leadership structure to enhance communication efficiency. The company consists of four key roles:* A - Software Engineer
* B - Senior Manager
* C - Team Lead
* D - Project Manager

The directed network diagram represents the communication channels between these roles.Calculate all centrality measures for each node and determine the most influential node based on these metrics. | 10 |
| 6 | SuperMart, a large retail chain, wants to analyze customer purchasing patterns to optimize store layout and promotions. The store maintains a transaction database where 1 represents the presence of an item, and 0 represents its absence.The item set under consideration:I = {Tea, Coffee, Sugar, Biscuits, Cereal, Jam}The store recorded twelve transactions as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Transaction ID** | **Tea** | **Coffee** | **Sugar** | **Biscuits** | **Cereal** | **Jam** |
| T1 | 1 | 1 | 1 | 1 | 0 | 0 |
| T2 | 1 | 1 | 0 | 1 | 1 | 0 |
| T3 | 0 | 0 | 1 | 1 | 1 | 0 |
| T4 | 1 | 0 | 1 | 1 | 0 | 0 |
| T5 | 0 | 0 | 1 | 1 | 0 | 1 |
| T6 | 1 | 0 | 1 | 1 | 0 | 1 |
| T7 | 1 | 0 | 0 | 0 | 0 | 1 |
| T8 | 1 | 0 | 1 | 1 | 0 | 0 |
| T9 | 0 | 1 | 1 | 1 | 0 | 1 |
| T10 | 1 | 0 | 1 | 1 | 0 | 0 |
| T11 | 1 | 0 | 1 | 1 | 0 | 0 |
| T12 | 1 | 0 | 1 | 0 | 1 | 1 |

Considering the following association rules:

|  |
| --- |
| **Association Rules** |
| **Antecedent** | **Consequent** |
| {Tea, Sugar} | {Biscuits} |
| {Jam} | {Tea, Sugar} |
| {Coffee} | {Sugar} |

1. Compute support, confidence, and lift for the association rule Antecedent → Consequent.
2. Identify the frequent item set from the set of association rules and justify your selection.
 | 10 |