JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATIONS- 2026

B.Tech-IV Semester (CSE/IT/BI)

COURSE CODE (CREDITS): 25B1WCI431 (2) MAX MARKS: 35

COURSE NAME: Data Analytics using R and Python

COURSE INSTRUCTORS: **ATA**, AYS, RVS MAX. TIME: 2 Hours

***Note:*** *(a) All questions are compulsory. (b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problem (c) Use of calculator is allowed*



<table>
<tr><th>Q. No</th><th>Questions</th><th>CO</th><th>Marks</th></tr>
<tr><td>Q1</td><td>The following Python code is intended to build a frequency dictionary from a nested list of student scores, but it produces wrong output. Identify all bugs, explain why each one is logically wrong, and write the corrected version:<br>scores = [[85, 90, 85], [90, 70], [85, 70, 70]]<br>freq = {}<br>for group in scores:<br>for s in group:<br>if s not in freq:<br>freq[s] == 0<br>freq[s] =+ 1<br>print(freq)</td><td>2</td><td>5</td></tr>
<tr><td>Q2</td><td>A nutrition researcher studies the effect of three different diets — Low-Carb, Mediterranean, and High-Protein — on weight loss (in kg) over 8 weeks. Six participants were assigned to each diet group. After the study, each participant was also given a diet adherence score (out of 10) by a dietician, and their overall health improvement rank (1 = best) was recorded. The data for weight loss across the three groups Table 1 and the adherence/rank data for the Mediterranean group Table 2 are given below.<br><br>
<table>
<tr><th colspan="3">Table 1</th></tr>
<tr><th>Low-Carb</th><th>Medit</th><th>High-Prot</th></tr>
<tr><td>4.2</td><td>3.8</td><td>5.1</td></tr>
<tr><td>5</td><td>4.5</td><td>5.8</td></tr>
<tr><td>3.8</td><td>3.2</td><td>4.6</td></tr>
<tr><td>6.1</td><td>5</td><td>6.3</td></tr>
<tr><td>4.7</td><td>4.1</td><td>5.5</td></tr>
<tr><td>5.3</td><td>3.6</td><td>6</td></tr>
</table>
<table>
<tr><th colspan="2">Table 2</th></tr>
<tr><th>Adherence (X)</th><th>Health Rank (Y)</th></tr>
<tr><td>6</td><td>5</td></tr>
<tr><td>8</td><td>3</td></tr>
<tr><td>4</td><td>6</td></tr>
<tr><td>9</td><td>1</td></tr>
<tr><td>5</td><td>4</td></tr>
<tr><td>7</td><td>2</td></tr>
</table></td><td>3</td><td>5</td></tr>
</table>

| | | | |
|---|---|---|---|
| | **a)**Using Table 1, perform a one-way ANOVA at $\alpha = 0.05$. Use Fcritical (2, 15) = 3.682 and state your conclusion.<br>**b)**Using Table 2, compute Spearman's Correlation, Test significance at $\alpha = 0.05$ using df = 4, tcritical = 2.776. Interpret the result. | | |
| Q3 | An IT hardware supplier tracks the following server component bundles in a series of custom build orders:<br><br>**TID** \| **Components Purchased**<br>T100 \| GPU, RAM, SSD<br>T200 \| GPU, Motherboard, CPU, Cooler<br>T300 \| RAM, Motherboard, CPU, PSU<br>T400 \| GPU, RAM, Motherboard, CPU<br>T500 \| GPU, RAM, Motherboard, PSU<br><br>Using the Apriori Algorithm, identify all frequent itemsets with a minimum support threshold of 60%. Show the candidate and frequent sets for each pass. | 4 | 4 |
| Q4 | Consider the following dataset of software installations:<br><br>**User** \| **Installed Tools**<br>U1 \| VS Code, Git, Docker<br>U2 \| VS Code, Git<br>U3 \| VS Code, Docker<br>U4 \| Git, Docker<br>U5 \| VS Code, Git<br><br>Calculate the Support of {VS Code, Git} and the confidence of the rule {VS Code → Git}, and explain using the Apriori property why all subsets of a frequent itemset must also be frequent. | 4 | 4 |
| Q5 | Using the following transaction database, construct the FP-Tree and find all frequent 2-itemsets using the FP-Growth approach. Assume minimum support count = 3.<br><br>**TID** \| **Items Purchased**<br>T1 \| {A, B, D}<br>T2 \| {B, C, E}<br>T3 \| {A, B, C, E}<br>T4 \| {B, E} | 4 | 4 |

| | | | | |
|---|---|---|---|---|
| | T5 | {A, B, C, E} | | |
| Q6 | Write down 2 packages and 4 functions of R language for machine learning with functionality. | | 2 | 3 |
| Q7 | As a data scientist, what skills are required to complete any project of data analytics. | | 1 | 2 |
| Q8 | How can Kendall Tau be applied in machine learning model evaluation for ranking problems? | | 3 | 2 |
| Q9 | A dataset contains the following records for loan approval: | | 5 | 6 |

| **Income** | **Credit Score** | **Loan Approved** |
|---|---|---|
| High | Good | Yes |
| High | poor | Yes |
| Medium | Good | Yes |
| Low | Good | No |
| Low | Poor | No |

a)Determine the best attribute split using Gini Index
b)Construct the first level of CART tree

ALL THE BEST