Subject Title: Machine Learning
Subject Code: 20CST-316
Semester: V

Time: 3 Hours **Maximum Marks: 60**

Instructions: Attempt all questions

<table>
<tr><th>Q. No</th><th>Statement</th><th>CO mapping</th></tr>
<tr><td colspan="3">Section A 5 x 2 = 10 marks</td></tr>
<tr><td>1</td><td>Differentiate a missing value with an outlier</td><td>CO2</td></tr>
<tr><td>2</td><td>Define Ensemble Learning</td><td>CO2</td></tr>
<tr><td>3</td><td>List out various advantages of using Decision Trees</td><td>CO2</td></tr>
<tr><td>4</td><td>Differentiate Unsupervised Learning and Reinforcement Learning</td><td>CO3</td></tr>
<tr><td>5</td><td>Define any problems using Naive Bayes for Classification</td><td>CO3</td></tr>
<tr><td colspan="3">Section B 4 x 5 = 20 marks</td></tr>
<tr><td>6</td><td>Compare SVM and Logistic Regression in handling outliers</td><td>CO2</td></tr>
<tr><td>7</td><td>Is Feature Scaling required for the KNN Algorithm? Explain with proper justification.</td><td>CO3</td></tr>
<tr><td>8</td><td>Explain how the Random Forests give output for Classification and Regression problems</td><td>CO2</td></tr>
<tr><td>9</td><td>Differentiate Manhattan Distance and Euclidean Distance in Clustering with an example</td><td>CO3</td></tr>
<tr><td colspan="3">Section C 3 x 10 = 30 marks</td></tr>
<tr><td>10</td><td>Given the data in Table, reduce the dimension from 2 to 1 using the Principal Component Analysis (PCA) algorithm.
<table>
<tr><th>Feature</th><th>Example 1</th><th>Example 2</th><th>Example 3</th><th>Example 4</th></tr>
<tr><td>X_1</td><td>4</td><td>8</td><td>13</td><td>7</td></tr>
<tr><td>X_2</td><td>11</td><td>4</td><td>5</td><td>14</td></tr>
</table></td><td>CO3</td></tr>
<tr><td>11</td><td>Find the frequent itemsets and generate the association rules using the Apriori algorithm using given dataset which has various transactions.
<table>
<tr><th>TID</th><th>ITEMSETS</th></tr>
<tr><td>T1</td><td>A, B</td></tr>
<tr><td>T2</td><td>B, D</td></tr>
<tr><td>T3</td><td>B, C</td></tr>
<tr><td>T4</td><td>A, B, D</td></tr>
<tr><td>T5</td><td>A, C</td></tr>
<tr><td>T6</td><td>B, C</td></tr>
<tr><td>T7</td><td>A, C</td></tr>
<tr><td>T8</td><td>A, B, C, E</td></tr>
<tr><td>T9</td><td>A, B, C</td></tr>
</table>
Given: Minimum Support= 2, Minimum Confidence= 50%</td><td>CO4</td></tr>
<tr><td>12</td><td>Justify with elaboration the following statement:
The k-means algorithm is based on the strong initial condition to decide the Number of clusters through the assignment of ‘k’ initial centroids or means.</td><td>CO4</td></tr>
</table>