**TKM COLLEGE OF ENGINEERING, KOLLAM-5**

**Department of Computer Applications**

**III Semester MCA**

**Internal Assessment Retest (Offline) Feb 2022**

**Course with Code:** 20MCA 201 DATA SCIENCE &MACHINE LEARNING

**Time:2Hrs** **Maximum Marks: 50**

<table>
<tr><th>Qn. No.</th><th>PART – A<br>Answer all questions</th><th>Marks</th><th>BL</th><th>CO</th></tr>
<tr><td>1</td><td>Write a note on histogram visualization technique.</td><td>3</td><td>L1</td><td>1</td></tr>
<tr><td>2</td><td>Explain the measures of spread and their significance in analyzing the data set.</td><td>3</td><td>L1</td><td>1</td></tr>
<tr><td>3</td><td>Why Laplace estimator is needed in classification using Bayes theorem. Justify your answer.</td><td>3</td><td>L2</td><td>2</td></tr>
<tr><td>4</td><td>State the limitation of KNN.</td><td>3</td><td>L2</td><td>2</td></tr>
<tr><td>5</td><td>How to choose best split in decision tree.</td><td>3</td><td>L2</td><td>3</td></tr>
<tr><td>6</td><td>Write down IR algorithm.</td><td>3</td><td>L1</td><td>3</td></tr>
<tr><td>7</td><td>Differentiate between hyperplane and maximum margin hyperplane.</td><td>3</td><td>L2</td><td>4</td></tr>
<tr><td>8</td><td>Explain about artificial neuron.</td><td>3</td><td>L1</td><td>4</td></tr>
<tr><td>9</td><td>Distinguish between precision and recall.</td><td>3</td><td>L2</td><td>5</td></tr>
<tr><td>10</td><td>Explain the various methods to perform cross validation.</td><td>3</td><td>L1</td><td>5</td></tr>
<tr><td></td><td>PART – B</td><td></td><td></td><td></td></tr>
<tr><td></td><td>MODULE-I</td><td></td><td></td><td></td></tr>
<tr><td>11 a</td><td>For the following set of training samples, find which attribute can be chosen as the root for decision tree classification.
<table>
<tr><th>Instance</th><th>Classification</th><th>a1</th><th>a2</th></tr>
<tr><td>1</td><td>+</td><td>T</td><td>T</td></tr>
<tr><td>2</td><td>+</td><td>T</td><td>T</td></tr>
<tr><td>3</td><td>–</td><td>T</td><td>F</td></tr>
<tr><td>4</td><td>+</td><td>F</td><td>F</td></tr>
<tr><td>5</td><td>–</td><td>F</td><td>T</td></tr>
<tr><td>6</td><td>–</td><td>F</td><td>T</td></tr>
</table>
OR</td><td>5</td><td>L3</td><td>3</td></tr>
<tr><td>b</td><td>Given the set of values X = (3, 9, 11, 5, 2) and Y = (1, 8, 11, 4, 3). Find the regression line that best fit the given sample data and</td><td>5</td><td>L3</td><td>3</td></tr>
</table>

| | | | | |
|---|---|---|---|---|
| | predict the value of, if X=7. | | | |
| | **MODULE IV** | | | |
| 12 a | Describe the significance of soft margin hyperplane and explain how they are computed. | 5 | L1 | 4 |
| | **OR** | | | |
| b | What is a Perceptron? Explain the working of a perceptron with a neat diagram. | 5 | L1 | 4 |
| | **MODULE-V** | | | |
| 13a | Suppose 1000 patients get tested for flu; out of them, 900 are actually healthy and 100 are actually sick. For the sick people, a test was positive for 62 and negative for 38. For the healthy people, the same test was positive for 18 and negative for 882. Construct a confusion matrix for the data and compute the precision and recall for the data. | 5 | L3 | 5 |
| b | **OR**<br>Use K Means clustering to cluster the following data into two groups. Assume cluster centroid are m1=2 and m2=4. The distance function used is Euclidean distance. { 2, 4, 10, 12, 3, 20, 30, 11, 25 } | 5 | L3 | 5 |
| | **MODULE-IV&V** | | | |
| 14 a | Explain how Support Vector Machine can be used for classification of linearly separable data. | 5 | L2 | 4 |
| | **OR** | | | |
| b | Describe the random forest algorithm to improve classifier accuracy. | 5 | L2 | 5 |
| | ******* | | | |

**TKM COLLEGE OF ENGINEERING, KOLLAM-5**

**Department of Computer Applications**

**I Semester MCA**

**Internal Assessment (Offline) March 2021**

**Course with Code:** 20MCA107 ADVANCED SOFTWARE ENGINEERING

**Scheme of Valuation/Answer Key**

**Time:2Hrs** **Maximum Marks: 50**

| Qn.No | PART –A<br>Answer all questions |
|---|---|
| 1 | Definition of custom assertion -1.5 mark,example-1.5 mark |
| 2 | Unit test explanation-1 mark, use of unit tests-2 mark |
| 3 | At least three difference between factory method and abstract factory method-1 mark each |
| 4 | Concept of Anti pattern – 2 mark, example -1 mark |
| 5 | Dataflow testing and its usage carries 1.5 marks each |
| 6 | Explanation of refactor method carries 3 mark |
| 7 | At least three characteristics of agility in agile frame work -1 mark each |
| 8 | Comparison between pair wise and state transition testing -2 mark, with example -1 mark |
| 9 | Usage of version control –at least three points – 1 mark each |
| 10 | Differentiating continuous delivery and continuous deployment with neat sketch –(2 mark,sketch-1 mark) |
| | **PART -B** |
| 11a | Definition of structural design pattern- 1mark<br>Types of structural design pattern-1 mark<br>Explanation of any two with its structure -3 mark |
| 11b | Explanation of assertion with example-2.5 mark<br>Explanation of expected error test with example-2.5 mark |
| 12 a | About SCRUM-2 mark<br>Phases with explanation -3 mark |
| 12 b | Any three testing methodologies -3.5 mark<br>If explaining with example – 1.5 mark each |
| 13 a | Use of xunit architecture- 1 mark<br>Phases with explanation and sketch – 4 mark |
| 13 b | Explanation of automated regression testing -2 mark<br>Features carries 3 mark |
| 14 a | Role of continuous integration in SCM- 2 mark<br>Strategy for its implementation-3 mark |
| 14 b | Explanation of deployment pipeline -1 mark<br>Detailed sketch carries 1 mark ,stages with its explanation carries 3 mark |