**K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22**
**(Autonomous College Affiliated to University of Mumbai)**

Subject Code: CEC701 Subject Name: Machine Learning Date:01/12/2022

Nov – Dec 2022
Program: B.Tech (Computer Engineering)
Examination: LY Semester: **VII**
Course Code: CEC701 and Course Name: Machine Learning
Duration: 2.5 Hours Max. Marks: 60

Instructions:
(1)All questions are compulsory.
(2)Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

| | | Max. Marks | CO | BT level |
|---|---|---|---|---|
| Q 1 | Solve any six questions out of eight: | 12 | | |
| i) | Define Machine Learning according to Tom Mitchell. Describe Tasks, Performance and Experience with respect to grapes crop disease detection. | 2M | CO4 | An |
| ii) | Explain Sigmoid activation function and its usage in any real world application. | 2M | CO2 | U |
| iii) | Define GINI index. What is its importance in creating a decision tree? | 2M | CO2 | U |
| iv) | Explain any two distance metrics for clustering. | 2M | CO3 | U |
| v) | Explain decision tree with its related terminologies. | 2M | CO2 | U |
| vi) | Differentiate between classification and clustering. | 2M | CO3 | An |
| vii) | Explain the terms 1.Hyperplane 2.Support vectors | 2M | CO2 | U |
| viii) | Justify necessity for dimensionality reduction in context to machine learning. | 2M | CO6 | U |
| Q.2 | Solve any four questions out of six. | 16 | | |
| i) | Explain Overfitting and Underfitting suitable diagram. | 4M | CO1 | U |
| ii) | Differentiate between linear regression and Logistic regression | 4M | CO2 | An |
| iii) | Explain Radial Basis Function to solve XOR problem. | 4M | CO5 | Ap |
| iv) | Write a short note on Expectation Maximization Algorithm. | 4M | CO3 | U |

<table>
<tr><td>v)</td><td>Consider confusion metrics for Cancer prediction problem.<br><table><tr><td></td><td>Actual Cancer=Yes</td><td>Actual Cancer=No</td></tr><tr><td>Predicted Cancer=Yes</td><td>True Positive= 57</td><td>False Positive= 14</td></tr><tr><td>Predicted Cancer=No</td><td>False Negative=23</td><td>True Negative=171</td></tr></table>Explain and Calculate Accuracy, Recall, Precision and F1 score</td><td>4M</td><td>CO5</td><td>Ap</td></tr>
<tr><td>vi</td><td>Differentiate between Principal Component Analysis and Linear Discriminate Analysis.</td><td>4M</td><td>CO6</td><td>An</td></tr>
<tr><td>Q.3</td><td>Solve any two questions out of three.</td><td>16</td><td></td><td></td></tr>
<tr><td>i)</td><td>Write a short note on<br>1.Steps to design machine learning application<br>2.Issues in Machine Learning</td><td>8M</td><td>CO1</td><td>U</td></tr>
<tr><td>ii)</td><td>Explain Rule based classification and related Sequence cover algorithm in detail.</td><td>8M</td><td>CO2</td><td>U</td></tr>
<tr><td>iii)</td><td>What is clustering? Explain how cluster is formed using Density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm.</td><td>8M</td><td>CO3</td><td>U</td></tr>
<tr><td>Q.4</td><td>Solve any two questions out of three.</td><td>16</td><td></td><td></td></tr>
<tr><td>i)</td><td>Write a short note on Singular Value Decomposition.</td><td>8M</td><td>CO6</td><td>U</td></tr>
<tr><td>ii)</td><td>Explain Linear regression and Multivariate Linear Regression. Apply linear regression to following data and predict Y for X = 10<br><table><tr><td>Sr no</td><td>TT1(X)</td><td>TT2(Y)</td></tr><tr><td>1</td><td>25</td><td>16</td></tr><tr><td>2</td><td>24</td><td>23</td></tr><tr><td>3</td><td>15</td><td>16</td></tr><tr><td>4</td><td>14</td><td>20</td></tr><tr><td>5</td><td>20</td><td>12</td></tr><tr><td>6</td><td>18</td><td>13</td></tr><tr><td>7</td><td>22</td><td>23</td></tr><tr><td>8</td><td>14</td><td>16</td></tr></table></td><td>8M</td><td>CO5</td><td>Ap</td></tr>
<tr><td>iii)</td><td>Identify suitable technique to boost the performance of machine learning model and explain related methods.</td><td>8M</td><td>CO4</td><td>An</td></tr>
</table>

**********