



**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

**SYLLABUS**

**UNIT I INTRODUCTION TO MACHINE LEARNING & PREPARING TO MODEL**

Introduction: What is Human Learning? Types of Human Learning, what is Machine Learning? Types of Machine Learning, Problems Not to Be Solved Using Machine Learning, Applications of Machine Learning, State-of-The-Art Languages/Tools in Machine Learning, Issues in Machine Learning Preparing to Model: Introduction, Machine Learning Activities, Basic Types of Data in Machine Learning, Exploring Structure of Data, Data Quality and Remediation, Data Pre-Processing.

**UNIT II MODELING AND EVALUATION & BASICS OF FEATURE ENGINEERING**

Introduction, selecting a Model, training a Model (for Supervised Learning), Model Representation and Interpretability, Evaluating Performance of a Model, Improving Performance of a Model Basics of Feature Engineering: Introduction, Feature Transformation, Feature Subset Selection

**UNIT III BAYESIAN CONCEPT LEARNING & SUPERVISED LEARNING:  
CLASSIFICATION**

Introduction, Why Bayesian Methods are Important? Bayes' Theorem, Bayes' Theorem and Concept Learning, Bayesian Belief Network .Supervised Learning: Classification: Introduction, Example of Supervised Learning, Classification Model, Classification Learning Steps, Common Classification Algorithms-k-Nearest Neighbour (kNN), Decision tree, Random forest model, Support vector machines

**UNIT IV SUPERVISED LEARNING: REGRESSION**

Introduction, Example of Regression, Common Regression Algorithms-Simple linear regression, Multiple linear regression, Assumptions in Regression Analysis, Main Problems in Regression Analysis, Improving Accuracy of the Linear Regression Model, Polynomial Regression Model, Logistic Regression, Maximum Likelihood Estimation.

**UNIT V UNSUPERVISED LEARNING**

Introduction, Unsupervised vs Supervised Learning, Application of Unsupervised Learning, Clustering – Clustering as a machine learning task, Different types of clustering techniques, Partitioning methods, K-Medoids: a representative object-based technique, Hierarchical clustering, Density-based methods-DBSCAN Finding Pattern using Association Rule- Definition of common terms, Association rule, Theapriori algorithm for association rule learning, Build the a priori principle rules.



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

Max Marks: **10**

S.No.	CO	Questions	BT
<b>Unit I: (INTRODUCTION TO MACHINE LEARNING &amp; PREPARING TO MODEL)</b>			
1	1	Explain human learning in detail. Discuss the different types of human learning and compare them with machine learning.	L2 + L4
2	1	With suitable illustration describe the end-to-end process of machine learning lifecycle?	L2
3	1	Discuss the applications of Machine Learning across various domains such as healthcare, finance, manufacturing, and education.	L2
4	1	Discuss the major issues and challenges in Machine Learning, including data-related, model-related, and ethical issues.	L2
5	1	Compare and contrast Supervised, Unsupervised, and Reinforcement Learning with suitable examples.	L4
6	1	Explain the problems that should not be solved using Machine Learning. Justify your answer with appropriate reasons and examples.	L5
7	1	Describe the basic types of data in Machine Learning. Explain structured, semi-structured, and unstructured data with examples.	L2
8	1	Describe the state-of-the-art languages and tools used in Machine Learning. Explain their roles in model development and deployment.	L2
9	1	Discuss data quality issues and data pre-processing techniques used in Machine Learning.	L2
10	1	Explain the concept of preparing data for machine learning models. Describe the key machine learning activities involved in the modeling process.	L2



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

S.No.	CO	Questions	BT
<b>Unit II: (MODELING AND EVALUATION &amp;BASICS OF FEATURE ENGINEERING)</b>			
1	2	Explain the process of selecting an appropriate machine learning model for a given problem, considering data characteristics and evaluation criteria.	L4
2	2	What is model representation in machine learning? Explain different representations and their impact on learning and interpretability.	L2
3	2	Compare and contrast following feature subset selection methods; filter, wrapper, embedded. Also, discuss their advantages and limitations in various scenarios.	L4
4	2	Compare and contrast feature creation, feature transformation, feature extraction, feature selection and feature scaling with suitable examples.	L4
5	2	Explain how data is split into training, validation, and test sets, and describe the different methods of validation techniques.	L2
6	2	Compare and contrast k-fold cross-validation, stratified k-fold cross-validation, leave-one-out cross-validation and holdout validation. Discuss their advantages and disadvantages.	L4
7	2	Explain various methods used to evaluate the performance of machine learning models for classification and regression tasks.	L2
8	2	A medical test is performed on 200 patients: 20 patients have the disease, 180 are healthy. The test correctly identifies 15 diseased patients (TP) and 160 healthy patients (TN). a) Construct the confusion matrix. b) Calculate accuracy, precision, recall, and F1-score. c) Comment on why accuracy might be misleading in this case.	L3



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

9	2	<p>A binary classifier produces the following results on a test set of 100 instances:</p> <ul style="list-style-type: none"><li>• True Positives (TP) = 40</li><li>• True Negatives (TN) = 45</li><li>• False Positives (FP) = 5</li><li>• False Negatives (FN) = 10</li></ul> <p>Draw the confusion matrix and calculate the accuracy, precision, recall, and F1-score.</p>	L3
10	2	<p>In a Covid test of 1000 patients, there were 45 positive tests, of which 30 patients had covid and 15 were falsely tested positive. Of the 955 negative tests there were 5 that were incorrect, these patients had covid but were tested negatively. Draw the confusion matrix and calculate the accuracy, precision, recall, sensitivity and F1 score from the matrix.</p>	L3
S.No.	CO	Questions	BT
<b>Unit III: (BAYESIAN CONCEPT LEARNING &amp; SUPERVISED LEARNING: CLASSIFICATION)</b>			
1	3	<p>Explain the importance of Bayesian Methods in machine learning. State and explain Bayes' theorem with a suitable mathematical expression also state real-life applications of Bayes' theorem.</p>	L2
2	3	<p>A man is known to speak the lies 1 out of 4 times. He throws a die and reports that it is a six. Find the probability that it is actually a six.</p>	L3



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

3	3	<p>A training dataset for classifying emails:</p> <table border="1"><thead><tr><th>Class</th><th>Spam</th><th>Not Spam</th></tr></thead><tbody><tr><td>Prior</td><td>0.4</td><td>0.6</td></tr></tbody></table> <p>Word probabilities: <math>P(\text{Offer} \text{Spam}) = 0.7</math>; <math>P(\text{Offer} \text{NotSpam}) = 0.1</math>; <math>P(\text{Free} \text{Spam}) = 0.8</math></p> <p><math>P(\text{Free} \text{NotSpam}) = 0.05</math>; For a new email containing words “Offer” and “Free”, classify it using Naïve Bayes.</p>	Class	Spam	Not Spam	Prior	0.4	0.6	L3																																				
Class	Spam	Not Spam																																											
Prior	0.4	0.6																																											
4	3	<p>Use Naïve Bayes to classify a new email as Spam or Non-Spam based on the frequencies of unique words: offer, win, prize, buy, cheap, meeting, schedule, project, and update.</p> <table border="1"><thead><tr><th>Email</th><th>Words Present</th><th>Class</th></tr></thead><tbody><tr><td>E<sub>1</sub></td><td>"offer win prize"</td><td>Spam</td></tr><tr><td>E<sub>2</sub></td><td>"buy cheap offer"</td><td>Spam</td></tr><tr><td>E<sub>3</sub></td><td>"meeting schedule project"</td><td>Non-Spam</td></tr><tr><td>E<sub>4</sub></td><td>"project meeting update"</td><td>Non-Spam</td></tr></tbody></table>	Email	Words Present	Class	E <sub>1</sub>	"offer win prize"	Spam	E <sub>2</sub>	"buy cheap offer"	Spam	E <sub>3</sub>	"meeting schedule project"	Non-Spam	E <sub>4</sub>	"project meeting update"	Non-Spam	L3																											
Email	Words Present	Class																																											
E <sub>1</sub>	"offer win prize"	Spam																																											
E <sub>2</sub>	"buy cheap offer"	Spam																																											
E <sub>3</sub>	"meeting schedule project"	Non-Spam																																											
E <sub>4</sub>	"project meeting update"	Non-Spam																																											
5	3	<p>Consider a data set with k-Nearest neighbor, calculate Euclidean distance with <math>k = 3</math> from the new point <math>X=[5.9, 3.0, 5.1, 1.8]</math>.</p> <table border="1"><thead><tr><th>No</th><th>Sepal Length (cm)</th><th>Sepal Width (cm)</th><th>Petal Length (cm)</th><th>Petal Width (cm)</th><th>Class (Species)</th></tr></thead><tbody><tr><td>1</td><td>5.1</td><td>3.5</td><td>1.4</td><td>0.2</td><td>Setosa</td></tr><tr><td>2</td><td>4.9</td><td>3.0</td><td>1.4</td><td>0.2</td><td>Setosa</td></tr><tr><td>3</td><td>5.8</td><td>2.7</td><td>4.1</td><td>1.0</td><td>Versicolor</td></tr><tr><td>4</td><td>6.0</td><td>2.9</td><td>4.5</td><td>1.5</td><td>Versicolor</td></tr><tr><td>5</td><td>6.3</td><td>3.3</td><td>6.0</td><td>2.5</td><td>Virginica</td></tr><tr><td>6</td><td>5.8</td><td>2.7</td><td>5.1</td><td>1.9</td><td>Virginica</td></tr></tbody></table>	No	Sepal Length (cm)	Sepal Width (cm)	Petal Length (cm)	Petal Width (cm)	Class (Species)	1	5.1	3.5	1.4	0.2	Setosa	2	4.9	3.0	1.4	0.2	Setosa	3	5.8	2.7	4.1	1.0	Versicolor	4	6.0	2.9	4.5	1.5	Versicolor	5	6.3	3.3	6.0	2.5	Virginica	6	5.8	2.7	5.1	1.9	Virginica	L3
No	Sepal Length (cm)	Sepal Width (cm)	Petal Length (cm)	Petal Width (cm)	Class (Species)																																								
1	5.1	3.5	1.4	0.2	Setosa																																								
2	4.9	3.0	1.4	0.2	Setosa																																								
3	5.8	2.7	4.1	1.0	Versicolor																																								
4	6.0	2.9	4.5	1.5	Versicolor																																								
5	6.3	3.3	6.0	2.5	Virginica																																								
6	5.8	2.7	5.1	1.9	Virginica																																								



**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

6	3	<p>Consider a data set with k-Nearest neighbor, Compute the Euclidean distance between the new day and all previous observations. Identify the 3 nearest neighbors (<math>k = 3</math>). Based on majority class, decide whether to Play Tennis (Yes/No). New Day: Temperature = <math>26^{\circ}\text{C}</math>, Humidity = 70%.</p> <table border="1"><thead><tr><th>Observation</th><th>Temperature (<math>^{\circ}\text{C}</math>)</th><th>Humidity (%)</th><th>Play Tennis</th></tr></thead><tbody><tr><td>1</td><td>30</td><td>70</td><td>Yes</td></tr><tr><td>2</td><td>25</td><td>65</td><td>No</td></tr><tr><td>3</td><td>28</td><td>80</td><td>Yes</td></tr><tr><td>4</td><td>22</td><td>90</td><td>No</td></tr><tr><td>5</td><td>27</td><td>75</td><td>Yes</td></tr></tbody></table>	Observation	Temperature ( $^{\circ}\text{C}$ )	Humidity (%)	Play Tennis	1	30	70	Yes	2	25	65	No	3	28	80	Yes	4	22	90	No	5	27	75	Yes	L3
Observation	Temperature ( $^{\circ}\text{C}$ )	Humidity (%)	Play Tennis																								
1	30	70	Yes																								
2	25	65	No																								
3	28	80	Yes																								
4	22	90	No																								
5	27	75	Yes																								
7	3	<p>Illustrate how Support Vector Machine (SVM) separates data points using the concept of hyperplanes and margins. Discuss with examples and graphical illustrations.</p>	L2																								
8	3	<p>Explain how SVM handles non-linearly separable data using feature transformation.</p> <p>Illustrate the transformation from low-dimensional space to high-dimensional space with an example.</p>	L2																								
9	3	<p>Consider the dataset below that predicts whether a student will pass an exam based on study hours and attendance. Calculate the Entropy of the dataset. Compute the Information Gain for each attribute. Construct the Decision Tree and justify the choice of the root node.</p> <table border="1"><thead><tr><th>Student</th><th>Study Hours</th><th>Attendance</th><th>Result</th></tr></thead><tbody><tr><td>S1</td><td>High</td><td>Regular</td><td>Pass</td></tr><tr><td>S2</td><td>Low</td><td>Irregular</td><td>Fail</td></tr><tr><td>S3</td><td>Medium</td><td>Regular</td><td>Pass</td></tr><tr><td>S4</td><td>Low</td><td>Regular</td><td>Fail</td></tr><tr><td>S5</td><td>High</td><td>Irregular</td><td>Pass</td></tr></tbody></table>	Student	Study Hours	Attendance	Result	S1	High	Regular	Pass	S2	Low	Irregular	Fail	S3	Medium	Regular	Pass	S4	Low	Regular	Fail	S5	High	Irregular	Pass	L2
Student	Study Hours	Attendance	Result																								
S1	High	Regular	Pass																								
S2	Low	Irregular	Fail																								
S3	Medium	Regular	Pass																								
S4	Low	Regular	Fail																								
S5	High	Irregular	Pass																								
10	3	<p>What are decision trees, and how are they related to Random Forests? Illustrate with an example how majority voting works in a Random Forest classifier.</p>	L3																								



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

S.N o.	C O	Questions	B T												
<b>Unit IV: (SUPERVISED LEARNING: REGRESSION)</b>															
1	4	Describe the concept of supervised learning and explain the difference between regression and classification tasks with examples.	<b>L2</b>												
2	4	Explain the concept of linear regression. How does the model estimate the relationship between dependent and independent variables?	<b>L2</b>												
3	4	Describe following evaluation metrics used in regression; MAE, MSE, RMSE, RMSLE, R <sup>2</sup> , Adjusted R <sup>2</sup> . Explain their significance and how do these metrics help in model assessment?	<b>L2</b>												
4	4	<p>A dataset contains the following points:</p> <table border="1" style="margin-left: auto; margin-right: auto;"><thead><tr><th>x</th><th>y</th></tr></thead><tbody><tr><td>1</td><td>2</td></tr><tr><td>2</td><td>3</td></tr><tr><td>3</td><td>5</td></tr><tr><td>4</td><td>4</td></tr><tr><td>5</td><td>6</td></tr></tbody></table> <p>Fit a simple linear regression model <math>y = \beta_0 + \beta_1 x</math> using the least squares method. Find <math>\beta_0</math> and <math>\beta_1</math>.</p>	x	y	1	2	2	3	3	5	4	4	5	6	<b>L3</b>
x	y														
1	2														
2	3														
3	5														
4	4														
5	6														



**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

5	4	<p>Consider the dataset and determine the following regression metrics; MAE, MSE, RMSE, RMSLE, <math>R^2</math>, Adjusted <math>R^2</math>.</p> <table border="1"><thead><tr><th><i>Height</i></th><th><i>Weight</i></th></tr></thead><tbody><tr><td>151</td><td>63</td></tr><tr><td>174</td><td>81</td></tr><tr><td>138</td><td>56</td></tr><tr><td>186</td><td>91</td></tr><tr><td>128</td><td>47</td></tr><tr><td>136</td><td>57</td></tr><tr><td>179</td><td>76</td></tr><tr><td>163</td><td>72</td></tr><tr><td>152</td><td>62</td></tr><tr><td>131</td><td>48</td></tr></tbody></table>	<i>Height</i>	<i>Weight</i>	151	63	174	81	138	56	186	91	128	47	136	57	179	76	163	72	152	62	131	48	<b>L3</b>
<i>Height</i>	<i>Weight</i>																								
151	63																								
174	81																								
138	56																								
186	91																								
128	47																								
136	57																								
179	76																								
163	72																								
152	62																								
131	48																								
6	4	<p>Compute the correlation coefficient for the given dataset; and find the relation between height and weight of ten students. Also predict the weight of a person with height of 169.</p> <table border="1"><thead><tr><th><i>Height</i></th><th><i>Weight</i></th></tr></thead><tbody><tr><td>151</td><td>63</td></tr><tr><td>174</td><td>81</td></tr><tr><td>138</td><td>56</td></tr><tr><td>186</td><td>91</td></tr><tr><td>128</td><td>47</td></tr><tr><td>136</td><td>57</td></tr><tr><td>179</td><td>76</td></tr><tr><td>163</td><td>72</td></tr><tr><td>152</td><td>62</td></tr><tr><td>131</td><td>48</td></tr></tbody></table>	<i>Height</i>	<i>Weight</i>	151	63	174	81	138	56	186	91	128	47	136	57	179	76	163	72	152	62	131	48	<b>L3</b>
<i>Height</i>	<i>Weight</i>																								
151	63																								
174	81																								
138	56																								
186	91																								
128	47																								
136	57																								
179	76																								
163	72																								
152	62																								
131	48																								



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

QUESTION BANK

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

7	4	<p>The marks obtained by a student are dependent on his/her study time. Given study time in minutes and marks out of 2000. Find the relationship between study time and marks using the concept of linear regression. Also predict the marks for a student if he/her studied for 790 minutes.</p> <table border="1"><thead><tr><th><i>Study Time (min)</i></th><th><i>Marks Obtained</i></th></tr></thead><tbody><tr><td>350</td><td>520</td></tr><tr><td>1070</td><td>1600</td></tr><tr><td>630</td><td>1000</td></tr><tr><td>890</td><td>850</td></tr><tr><td>940</td><td>1350</td></tr><tr><td>500</td><td>490</td></tr></tbody></table>	<i>Study Time (min)</i>	<i>Marks Obtained</i>	350	520	1070	1600	630	1000	890	850	940	1350	500	490	<b>L3</b>																
<i>Study Time (min)</i>	<i>Marks Obtained</i>																																
350	520																																
1070	1600																																
630	1000																																
890	850																																
940	1350																																
500	490																																
8	4	<p>Consider the dataset and fit a multiple linear regression model to predict Price based on Size, Bedrooms, and Age. Also, predict the price of a house with 1500 sq.ft., 3 bedrooms, and 10 years old.</p> <table border="1"><thead><tr><th>houseID</th><th>Size (sq.ft.)</th><th>Bedrooms</th><th>Age (years)</th><th>Price (\$1000s)</th></tr></thead><tbody><tr><td></td><td>1400</td><td>3</td><td>10</td><td>245</td></tr><tr><td></td><td>1600</td><td>3</td><td>5</td><td>312</td></tr><tr><td></td><td>1700</td><td>4</td><td>15</td><td>279</td></tr><tr><td></td><td>1875</td><td>4</td><td>7</td><td>308</td></tr><tr><td></td><td>1100</td><td>2</td><td>20</td><td>199</td></tr></tbody></table>	houseID	Size (sq.ft.)	Bedrooms	Age (years)	Price (\$1000s)		1400	3	10	245		1600	3	5	312		1700	4	15	279		1875	4	7	308		1100	2	20	199	<b>L3</b>
houseID	Size (sq.ft.)	Bedrooms	Age (years)	Price (\$1000s)																													
	1400	3	10	245																													
	1600	3	5	312																													
	1700	4	15	279																													
	1875	4	7	308																													
	1100	2	20	199																													



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

QUESTION BANK

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

9	4	Consider the dataset and fit a multiple regression model to predict Exam Score using Study Hours, Attendance, and Assignments Completed. Calculate the coefficients using the least squares method. Predict the score for a student with 6 study hours, 80% attendance, and 4 completed assignments.					L3
		Student ID	Study Hours	Attendance (%)	Assignments Completed	Exam Score	
		1	5	80	4	70	
		2	8	90	5	88	
		3	6	75	3	65	
		4	7	85	5	80	
		5	4	60	2	55	
10	4	Consider the dataset fit a multiple regression model to predict MPG using Engine Size, Weight, and Horsepower. Interpret the coefficients: which feature has the strongest negative impact on MPG. Predict the MPG for a car with Engine Size = 2.8 L, Weight = 1650 kg, Horsepower = 145.					L4
		CarID	Engine Size (L)	Weight (kg)	Horsepower	MPG (Fuel Efficiency)	
		1	2.0	1500	120	35	
		2	3.0	1700	150	28	
		3	2.5	1600	140	30	
		4	3.5	1800	170	25	
		5	1.8	1400	110	38	



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

11	4	Explain the concept of the logistic regression model. How does the logit function help in binary classification? Consider a data set and build a logistic regression model based on the samples, also predict admission probability for: Case 1: ( $x_1=70$ , $x_2=80$ ) and Case 2: ( $x_1=55$ , $x_2=60$ )			
			Exam1 ( $x_1$ )	Exam2 ( $x_2$ )	Admitted ( $y$ )
			50	85	0
			60	95	0
			80	70	1
			90	88	1
				<b>L3</b>	

S.No.	CO	Questions	BT
<b>Unit V: (UNSUPERVISED LEARNING)</b>			
1	5	Explain the concept of clustering in unsupervised learning. Differentiate between hard clustering and soft clustering with suitable examples.	L2
2	5	Illustrate the K-Means clustering process with an example dataset. Describe the key steps involved and apply the algorithm to show how centroids are updated during iterations.	L3
3	5	Compare and contrast the K-Means and K-Medoids clustering algorithms. Analyze their performance with respect to outliers, data types, and computational complexity.	L4
4	5	Describe the working principle of DBSCAN. Explain the role of parameters Epsilon ( $\epsilon$ ) and MinPts, and discuss how DBSCAN identifies core, border, and noise points.	L2



**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

5	5	Demonstrate how hierarchical clustering is performed using Agglomerative (Bottom Up) and Divisive (Top-Down) approaches. Include suitable dendrograms to illustrate the process.	L3																
6	5	Explain the concept of a dendrogram in hierarchical clustering. Discuss how different linkage criteria affect the structure of resulting clusters.	L2																
7	5	Given the following 1D dataset: Data Points: 2, 4, 10, 12, 20, 22	L3																
8	5	<p>Given the following distance matrix in Hierarchical Clustering;</p> <table border="1"><thead><tr><th>Particulars</th><th>A</th><th>B</th><th>C</th></tr></thead><tbody><tr><th>A</th><td>0</td><td>2</td><td>6</td></tr><tr><th>B</th><td>2</td><td>0</td><td>5</td></tr><tr><th>C</th><td>6</td><td>5</td><td>0</td></tr></tbody></table> <p>a) Perform Agglomerative Hierarchical Clustering using Single Linkage. b) Show step-by-step cluster merging. c) Draw the corresponding dendrogram. d) Analyze how Complete Linkage would change the result.</p>	Particulars	A	B	C	A	0	2	6	B	2	0	5	C	6	5	0	L3
Particulars	A	B	C																
A	0	2	6																
B	2	0	5																
C	6	5	0																
9	5	<p>Given the following 2D points: (1,2), (2,2), (2,3), (8,7), (8,8), (25,80); Let: Eps = 2 &amp; MinPts = 2</p> <p>a) Identify core points, border points, and noise points. b) Form clusters using DBSCAN. c) Explain why DBSCAN is preferred over K-means in discovering arbitrary-shaped clusters.</p>	L2																



SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES  
(AUTONOMOUS)

**QUESTION BANK**

Year / Semester: **III B.Tech VI Semester**

Regulation: **R23**

Subject and Code: **23OML361A**

10	5	Define and explain the following terms with examples: i) Itemset ii) Support iii) Confidence iv) Lift v) Frequent Itemset vi) Strong Association Rule; Explain how association rules are useful in market basket analysis.	L2
11	5	Explain the Apriori algorithm in detail. Discuss the Apriori principle and how it reduces search space. Evaluate its advantages and limitations when applied to large datasets	L2

Note: L1-Remembering, L2-Understanding, L3-Applying, L4-Analyzing, L5-Evaluating, and L6-Creating

**The Six Levels of Bloom's Taxonomy:**

1. **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory (e.g., list, define, name, locate).
2. **Understanding:** Constructing meaning, explaining ideas, or concepts (e.g., summarize, interpret, classify, compare).
3. **Applying:** Using information in new situations or implementing procedures to solve problems (e.g., solve, use, demonstrate, implement).
4. **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure (e.g., contrast, categorize, distinguish, diagram).
5. **Evaluating:** Making judgments based on criteria and standards through checking and critiquing (e.g., judge, critique, justify, defend, argue).
6. **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure (e.g., design, construct, develop, formulate).