

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

April 2026 ~~Nov - Dec 2025~~

(M.Tech) Program: Artificial Intelligence Scheme : II & HB

Supplementary ~~Regular~~ Examination: LY Semester: VII

Course Code: PCEC102 and Course Name: Machine Learning and Pattern Recognition

Date of Exam: ~~19/01/2026~~ 13-04-2026 Duration: 02.5 Hours

Max. Marks: 60

Instructions:

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

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Explain the steps involved in PCA with mathematical formulation. State two advantages and one limitation of PCA.		CO2	U
b)	What is a Support Vector Machine? Explain the concept of maximum margin hyperplane using a diagram and mathematical equations.		CO4	U
c)	Compare the decision tree algorithms ID3 and C4.5 based on following aspects: <ul style="list-style-type: none"> • Attribute selection criteria used by each algorithm • Handling of continuous and discrete attributes • Treatment of missing values • Pruning mechanism and overfitting control • Overall advantages of C4.5 over ID3 		CO3	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	What do you mean by overfitting and underfitting? Explain bias-variance trade-off.		CO3	U
b)	Consider the following dataset with 6; 2-dimensional points: $P_1=(1,2)$, $P_2=(2,1)$, $P_3=(4,5)$, $P_4=(5,4)$, $P_5=(8,9)$, $P_6=(9,8)$ Apply K-Means clustering with $K=2$, and the initial centroids: $C_1=(1,2)$, $C_2=(8,9)$ a) Assign each point to the nearest centroid. b) Compute the new centroids after the first iteration.		CO4	Ap
c)	Explain how Random Forests improve prediction accuracy compared to a single decision tree.	CO5	U	
Q.3	Solve any two questions out of three. (10 marks each)	20		

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

~~Nov-Dec 2025~~
April 2026

(M.Tech) Program: Artificial Intelligence Scheme : II & IIB

~~Supplementary Regular Examination: LY~~ Semester: VII

Course Code: PCEC102 and Course Name: Machine Learning and Pattern Recognition

Date of Exam: ~~13-04-2026~~ **13-04-2026** Duration: 02.5 Hours

Max. Marks: 60

a)	<p>i) A small bank wants to predict whether a customer will subscribe to a term deposit. The dataset is:</p> <table border="1"> <thead> <tr> <th>Customer</th> <th>Age</th> <th>Income</th> <th>Subscribe</th> </tr> </thead> <tbody> <tr><td>1</td><td>Young</td><td>High</td><td>No</td></tr> <tr><td>2</td><td>Young</td><td>High</td><td>No</td></tr> <tr><td>3</td><td>Middle</td><td>High</td><td>Yes</td></tr> <tr><td>4</td><td>Senior</td><td>Medium</td><td>Yes</td></tr> <tr><td>5</td><td>Senior</td><td>Low</td><td>Yes</td></tr> <tr><td>6</td><td>Senior</td><td>Low</td><td>No</td></tr> <tr><td>7</td><td>Middle</td><td>Low</td><td>Yes</td></tr> <tr><td>8</td><td>Young</td><td>Medium</td><td>No</td></tr> <tr><td>9</td><td>Young</td><td>Low</td><td>Yes</td></tr> <tr><td>10</td><td>Senior</td><td>Medium</td><td>Yes</td></tr> </tbody> </table> <p>Use the Naïve Bayes classifier to predict whether a new customer with Age = Young and Income = Medium will subscribe or not.(5mk)</p> <p>ii)) Consider the following training data points belonging to two classes A and B:</p> <p>Point X Y Class</p> <p>P1 1 1 A P2 2 2 A P3 3 3 A P4 6 6 B P5 7 7 B P6 8 8 B</p> <p>Using the K-Nearest Neighbors (KNN) algorithm with K = 3 and Euclidean distance, classify a new data point:Q(4,4).</p>	Customer	Age	Income	Subscribe	1	Young	High	No	2	Young	High	No	3	Middle	High	Yes	4	Senior	Medium	Yes	5	Senior	Low	Yes	6	Senior	Low	No	7	Middle	Low	Yes	8	Young	Medium	No	9	Young	Low	Yes	10	Senior	Medium	Yes	CO4	AP L
		Customer	Age	Income	Subscribe																																										
1	Young	High	No																																												
2	Young	High	No																																												
3	Middle	High	Yes																																												
4	Senior	Medium	Yes																																												
5	Senior	Low	Yes																																												
6	Senior	Low	No																																												
7	Middle	Low	Yes																																												
8	Young	Medium	No																																												
9	Young	Low	Yes																																												
10	Senior	Medium	Yes																																												
b)	<p>Given the following dataset of two classes, reduce the dimension of the dataset from 2D to 1D using Linear Discriminant Analysis (LDA).</p> <p>Class 1: $X_1 = \{(3,2), (1,5), (2,3), (4,4), (3,5)\}$</p> <p>Class 2: $X_2 = \{(7,9), (6,7), (8,6), (9,8), (7,7)\}$</p>	CO2	Ap																																												

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

~~Nov-Dec 2025~~
April 2026

(M.Tech) Program: Artificial Intelligence Scheme : II & III
~~Supplementary~~ Examination: LY Semester: VII

Course Code: PCEC102 and Course Name: Machine Learning and Pattern Recognition
 Date of Exam: ~~10/01/2026~~ **13-04-2026** Duration: 02.5 Hours Max. Marks: 60

c)	<p>A retail company wants to analyze its customer dataset to improve sales and marketing strategies. The dataset contains customer demographics, transaction history, product preferences, and purchase amounts.</p> <p>a) Identify which machine learning algorithms (classification, clustering, prediction, association) would be appropriate for the following scenarios. Justify your choice. (2mk)</p> <table border="1" data-bbox="256 688 1117 961"> <thead> <tr> <th>Scenario</th> <th>Task</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Predict whether a customer will buy a new product (Yes/No)</td> </tr> <tr> <td>2</td> <td>Group customers into segments based on purchasing behavior</td> </tr> <tr> <td>3</td> <td>Forecast monthly sales for the next 6 months</td> </tr> <tr> <td>4</td> <td>Find frequent itemsets of products bought together</td> </tr> </tbody> </table> <p>b) Choose one algorithm from each category above and describe two key parameters that can be tuned for better performance. (4mk)</p> <p>c) Explain the importance of parameter tuning in machine learning models and how it affects overfitting and underfitting. (4mk)</p>	Scenario	Task	1	Predict whether a customer will buy a new product (Yes/No)	2	Group customers into segments based on purchasing behavior	3	Forecast monthly sales for the next 6 months	4	Find frequent itemsets of products bought together	CO6	Ap																					
Scenario	Task																																	
1	Predict whether a customer will buy a new product (Yes/No)																																	
2	Group customers into segments based on purchasing behavior																																	
3	Forecast monthly sales for the next 6 months																																	
4	Find frequent itemsets of products bought together																																	
Q.4	Solve any two questions out of three. (10 marks each)	20																																
a)	<p>The following dataset is used to construct a Classification and Regression Tree (CART) to classify whether a customer will Purchase a Product (Yes / No) based on Age Group and Income Level.</p> <table border="1" data-bbox="256 1381 1117 1654"> <thead> <tr> <th>Attribute</th> <th>Category</th> <th>Yes</th> <th>No</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td rowspan="3">Age</td> <td>Young</td> <td>8</td> <td>2</td> <td>10</td> </tr> <tr> <td>Middle</td> <td>5</td> <td>5</td> <td>10</td> </tr> <tr> <td>Old</td> <td>2</td> <td>8</td> <td>10</td> </tr> <tr> <td rowspan="3">Income</td> <td>High</td> <td>9</td> <td>1</td> <td>10</td> </tr> <tr> <td>Medium</td> <td>4</td> <td>6</td> <td>10</td> </tr> <tr> <td>Low</td> <td>1</td> <td>9</td> <td>10</td> </tr> </tbody> </table> <p>a) Compute the Gini index for each category of Age and Income. b) Determine the best binary split for Age using CART. c) Determine the best binary split for Income using CART. d) Identify the root attribute selected by CART with justification.</p>	Attribute	Category	Yes	No	Total	Age	Young	8	2	10	Middle	5	5	10	Old	2	8	10	Income	High	9	1	10	Medium	4	6	10	Low	1	9	10	CO3	Ap
Attribute	Category	Yes	No	Total																														
Age	Young	8	2	10																														
	Middle	5	5	10																														
	Old	2	8	10																														
Income	High	9	1	10																														
	Medium	4	6	10																														
	Low	1	9	10																														

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

~~Nov 2025~~ **April 2026**
(M.Tech) Program: Artificial Intelligence Scheme : II & IIB
~~Regular Examination: LY~~ **Supplementary Examination: LY** Semester: VII
Course Code: PCEC102 and Course Name: Machine Learning and Pattern Recognition
Date of Exam: ~~10/01/2026~~ **13-04-2026** Duration: 02.5 Hours Max. Marks: 60

b)	<p>A university plans to develop a machine learning model to predict students' final exam scores based on attendance, internal assessment marks, study hours, and background information.</p> <p>Answer the following by applying appropriate machine learning concepts:</p> <p>a) Identify and apply basic machine learning definitions by specifying the learning type, features, target variable, and dataset used.</p> <p>b) Apply the concept of hypothesis space and inductive bias to justify the choice of linear regression for this problem.</p> <p>c) The collected dataset contains missing attendance values and inconsistent mark scales. Apply suitable data cleaning and data transformation techniques to prepare the data.</p> <p>d) Apply k-fold cross-validation to explain how the model performance should be evaluated and why it is preferred over a single train-test split.</p> <p>e) After training the model, apply evaluation and model visualization techniques to interpret model performance and regression coefficients.</p>	CO1	Ap
c)	<p>Explain the need and usefulness of ensemble classifiers in machine learning.(2mk) Discuss how ensemble methods improve classification performance compared to single classifiers.(2 mk) Describe the working principles, advantages, and typical use cases of the following ensemble techniques: Bagging, Boosting, and voting (8 mk)</p>	CO5	U
