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

Nov – Dec 2025
(B. Tech ) Program: Computer Engineering Scheme III
Regular Examination: TY Semester: V
Course Code: CEC502 and Course Name: Machine Learning
Date of Exam: 26/11/2025 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) | Describe different types of learning approaches used in Machine Learning. | | 1 | U |
| b) | Calculate the regression evaluation parameters i)MAE ii)MSE iii) RMSE and iv) $R^2$ for the given dataset points | | 2 | AP |
| c) | A machine learning model was evaluated on the following dataset of 10 instances | | 3 | AP |
| Q 2 | Solve any **two** questions out of three: (05 marks each) | 10 | | |
| a) | Describe Ensemble Learning and its types | | 4 | U |
| b) | Apply the DBSCAN algorithm on the following dataset using $\varepsilon = 1.5$ and **minPts = 3**. Identify core, border, noise points and form the final clusters. | | 5 | AP |

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Actual (Y) | 100 | 150 | 200 | 250 | 300 |
| Predicted ($\hat{Y}$) | 130 | 170 | 220 | 260 | 325 |

| ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Actual (Y)** | **Cat** | **Cat** | **Cat** | **Cat** | **Cat** | **Dog** | **Dog** | **Dog** | **Dog** | **Rabbit** | **Rabbit** | **Rabbit** |
| **Predicted ($\hat{Y}$)** | **Cat** | **Dog** | **Cat** | **Rabbit** | **Cat** | **Dog** | **Dog** | **Cat** | **Rabbit** | **Rabbit** | **Dog** | **Rabbit** |

I. Construct the confusion matrix for this dataset.
II. Calculate the following performance metrics of each class
i. Accuracy ii.Precision iii. Recall iv.F1Score
III. Interpret the results, what do these values indicate about the model's performance



**Seat No.:**

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

Nov – Dec 2025
(B. Tech ) Program: Computer Engineering Scheme III
Regular Examination: TY Semester: V
Course Code: CEC502 and Course Name: Machine Learning
Date of Exam: 26/11/2025 Duration: 02.5 Hours Max. Marks: 60

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ID | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
| **X** | **1** | **1** | **2** | **2** | **8** | **8** | **9** | **25** | **8** |
| **Y** | **1** | **2** | **2** | **1** | **7** | **8** | **7** | **80** | **6** |

c) Apply the K-Medoid clustering algorithm on the following 2-dimensional dataset. Use $C_1 = (3,4)$ and $C_2 = (7,4)$ as initial medoids. — 5 — AP

| ID | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | x10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **X1** | **2** | **3** | **3** | **4** | **6** | **6** | **7** | **7** | **8** | **7** |
| **X2** | **6** | **4** | **8** | **7** | **2** | **4** | **3** | **4** | **5** | **6** |

Q.3 Solve any **two** questions out of three. (10 marks each) — 20

a) i) Explain types of Feature Engineering with an example (5M)
ii) Compare Classification and Regression (5M) — 1 — U

b) i) Given area in square feet and rent in dollars. Find the relationship between rent and area using linear regression and also predict the property area 790 sq. feet (5M) — 2 — AP

| Sr.No. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Area** | **340** | **1080** | **640** | **880** | **990** | **510** |
| **Rent** | **500** | **1700** | **1100** | **800** | **1400** | **500** |

ii) Explain the Gradient Descent algorithm in detail. Discuss how the learning rate (α) affects the convergence of Gradient Descent. Give a suitable diagram to support your explanation. (5M)

c) Consider the following dataset — 3 — AP

| Age | Income | Student | Credit Rating | Buys Computer |
|---|---|---|---|---|
| <=30 | High | No | Fair | No |
| <=30 | High | No | Excellent | No |
| 31–40 | High | No | Fair | Yes |
| >40 | Medium | No | Fair | Yes |
| >40 | Low | Yes | Fair | Yes |
| >40 | Low | Yes | Excellent | No |
| 31–40 | Low | Yes | Excellent | Yes |
| <=30 | Medium | No | Fair | No |
| <=30 | Low | Yes | Fair | Yes |
| >40 | Medium | Yes | Fair | Yes |



**Seat No.:**

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

Nov – Dec 2025
(B. Tech ) Program: Computer Engineering Scheme III
Regular Examination: TY Semester: V
Course Code: CEC502 and Course Name: Machine Learning
Date of Exam: 26/11/2025 Duration: 02.5 Hours Max. Marks: 60

| | | | | |
|---|---|---|---|---|
| | i.Compute the entropy of the entire dataset.<br>ii. Calculate the Information Gain (IG) for each attribute (Age, Income, Student, Credit Rating).<br>iii. Identify the attribute that should be chosen as the root node<br>iv Construct the first split of the decision tree. | | | |
| Q.4 | Solve any **two** questions out of three. (10 marks each) | 20 | | |
| a) | Explain the Support Vector Machine (SVM) algorithm with a neat diagram and differentiate between Linear SVM and Non-Linear SVM with suitable examples. | | 4 | U |
| b) | Consider the following transactions:<br><br>T1: milk, bread, eggs<br>T2: milk, juice<br>T3: juice, butter<br>T4: milk, bread, eggs<br>T5: coffee, eggs<br>T6: coffee<br>T7: coffee, juice<br>T8: milk, bread, cookies, eggs<br>T9: cookies, butter<br>T10: milk, bread<br><br>Identify frequent item sets with minimum support = 30%, confidence = 75% and generate strong association rules using the Apriori algorithm. | | 5 | AP |
| c) | Apply LDA (Linear Discriminant Analysis) steps to convert the following 2D data into 1D data<br>Class W1: X1 = (x1,x2) ={ (4,2),(2,4),(2,3),(3,6),(4,4)}<br>Class W2: X2 = (x1,x2) ={ ( 9,10),(6,8),(9,5),(8,7),(10,8)} | | 6 | AP |

**********



**Seat No.:**