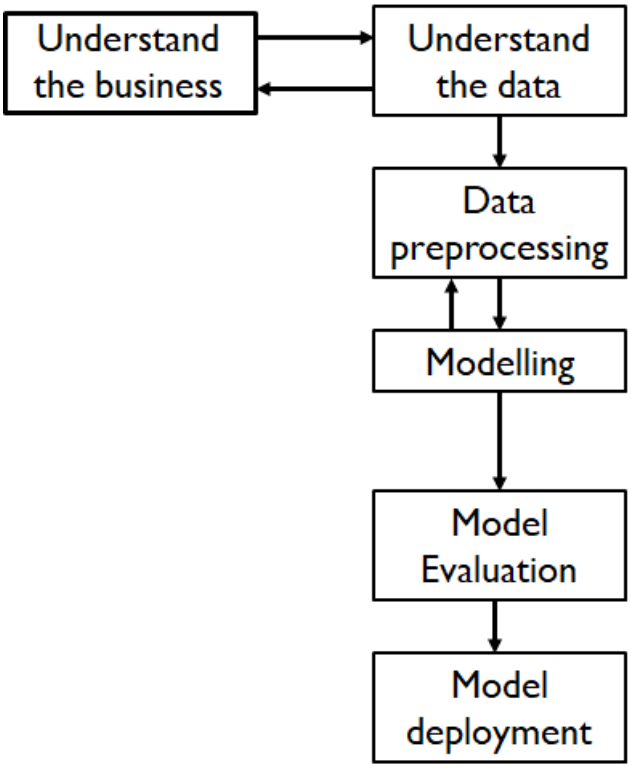
	Answer any FIVE FULL Questions	MARKS	CO	RBT
1	a. Explain the steps of machine learning and data mining process with a neat diagram.	6	CO2	1
	Ans: Understand the business, Understand the data, Data preprocessing, modelling, model evaluation, model deployment 			
	b. Apply the Z-score normalization in the following data and find the normalized values.: X= {8,10,15,20}	4	CO2	3
	Ans: mean 13.25 sd = 4.65 or 5.37 **New values [-1.127, -0.698, 0.375, 1.44]** **Or [-0.98, -0.6, 0.32, 1.25]**			
2	a. Find the outliers and 5-point summary for the following dataset and draw the box plot of that: {10,12,15,18,22,23,32,34,78,31,14}	5	CO2	3
	Ans: Q1=14 Q2=22, Q3=32 min =10, Max = 34, outlier 78,			
	b. Solve the following equations and find the value of x and y using the Gaussian Elimination Method. 3x+5y=9 2x+3y=5	5	CO2	3
	3 5 9 3R1-5R2 = -1 0 2 R2 + 2R1 = -1 0 2 **2 3 5 2 3 5 0 3 9** **X= -2, y = 3**			

3	<table><tr><th>Music Name</th><th>Number of downloads (in thousands)</th><th>User Review Rating</th><th>Number of User Reviews (in tens)</th><th>Whether you liked the Music?</th></tr><tr><td>A</td><td>2.35</td><td>7.5</td><td>2.3</td><td>Yes</td></tr><tr><td>B</td><td>9.87</td><td>5</td><td>0.5</td><td>No</td></tr><tr><td>C</td><td>0.87</td><td>4</td><td>7</td><td>Yes</td></tr><tr><td>D</td><td>1.2</td><td>7.1</td><td>5.6</td><td>Yes</td></tr><tr><td>E</td><td>5.5</td><td>6.2</td><td>4.3</td><td>No</td></tr><tr><td>F</td><td>2.3</td><td>1.9</td><td>2.8</td><td>No</td></tr></table> 4.8 downloads, user rating of 5.5 and has 3.1 user reviews.	10	CO3	3
	Ans: Distances [3.26, 5.71, 5.73, 4.66, 1.55, 4.39] **Nearest 3 are: A, E, F** **Inverse distance 0.306 0.642 0.227** **Normalised weight = 0.26 0.55 0.19** **Weight for YES = 0.26, Weight for NO= 0.74**			
4	Apply a) find S and b) Candidate Elimination on the following dataset and also draw the version spaces while applying candidate elimination algorithm. <table><tr><th>Origin</th><th>Manufacturer</th><th>Color</th><th>Decade</th><th>Type</th><th>Example Type</th></tr><tr><td>Japan</td><td>Honda</td><td>Blue</td><td>1980</td><td>Economy</td><td>Positive</td></tr><tr><td>Japan</td><td>Toyota</td><td>Green</td><td>1970</td><td>Sports</td><td>Negative</td></tr><tr><td>Japan</td><td>Toyota</td><td>Blue</td><td>1990</td><td>Economy</td><td>Positive</td></tr><tr><td>USA</td><td>Chrysler</td><td>Red</td><td>1980</td><td>Economy</td><td>Negative</td></tr><tr><td>Japan</td><td>Honda</td><td>White</td><td>1980</td><td>Economy</td><td>Positive</td></tr></table>	3+7	CO3	3
	Finds Japan, Honda, Blue, 1980 Economy (1) **Japan, ? Blue, ? Economy (3)** **Japan, ? ?, ? Economy (5)** **CE: S1: < Japan, Honda, Blue, 1980 Economy>** **G1: <? ? ? ? ?>** **S2: < Japan, Honda, Blue, 1980 Economy>** **G2: <? Honda ? ? ?> <? ? Blue ? ?> <? ? ? 1980 ?> <? ? ? ? Economy>** **S3: < Japan, ? Blue, ? Economy>** **G3: <? ? Blue ? ?> <? ? ? ? Economy>** **S4: < Japan, ? Blue, ? Economy>** **G4: <? ? Blue ? ?> <Japan ? ? ? Economy>** **S5: < Japan ? ? ? Economy>** **G5: <Japan ? ? ? Economy>**			
5	a. Explain the difference between lazy learning and Eager Learning with the help of an example. b. Describe various types of Performance Measure methods used to evaluate Regression.	[4+6]	CO3	2

Ans:

Instance-based Learning	Model-based Learning
Lazy Learners	Eager Learners
Processing of training instances is done only during testing phase	Processing of training instances is done during training phase
No model is built	Generalizes a model with the training instances before it receives a test instance
Predicts the class of the test instance directly from the training data	Predicts the class of the test instance using the model built
Slow in testing phase	Fast in testing phase
Learns by making many local approximations	Learns by creating global approximations

b) Standard error (residual), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative MSE, Coefficient of variation

6 | Apply the Linear Regression in Following Dataset: | 10 | CO3 | 3

X (Year)	Y (Expenditure)
1	12
2	19
3	29
4	37
5	45

Find the Expenditure of Company in 6th Year and 8th Year. Also plot the regression line.

$$a_1 = \frac{(\overline{xy}) - (\overline{x})(\overline{y})}{(\overline{x_i^2}) - (\overline{x})^2} \qquad a_0 = (\overline{y}) - a_1 \times \overline{x}$$

X_mean = 3, y_mean = 28.4, x2_mean = 11, XY_mean = 102

a1= 8.4, a0 = 3.2, 6th year value= 53.6, 8th year value = 70.4

CI CCI HoD

Detailed solution for numerical problems

Answer-1-(b)

1-(b) $X = \{0, 10, 15, 20\}$

$$\bar{x} = \frac{\Sigma x}{N}$$

$$= 0 + 10 + 15 + 20 \,/\, 5T4$$

$$= 13.25$$

$$\sigma = \sqrt{\frac{(x-\bar{x})^2}{N}} \quad \text{or} \quad \sqrt{\frac{(x-\bar{x})^2}{N-1}}$$

$$\sigma = \sqrt{\frac{(0-13.25)^2}{4} + \frac{(10-13.25)^2}{4} + \frac{(15-13.25)^2}{4} + \frac{(20-13.25)^2}{4}}$$

$$= \sqrt{27.5625 + 10.5625 + 3.0625 + 45.5625 \,/\, 4}$$

$$= \sqrt{21.6875}$$

$$= 4.65$$

$$\sigma = \sqrt{\frac{(0-13.25)^2}{3} + \frac{(10-13.25)^2}{3} + \frac{(15-13.25)^2}{3} + \frac{(20-13.25)^2}{3}}$$

$$= \sqrt{86.75 \,/\, 3}$$

$$= \sqrt{28.9166}$$

$$= 5.37$$

CMR

Z Score → $\frac{X-\mu}{\sigma}$

$Z_1 = \frac{8-13.25}{4.65}, \frac{8-13.25}{5.37} = -1.12, -0.977$

$Z_2 = \frac{10-13.25}{4.65}, \frac{10-13.25}{5.37} = -0.69, -0.60$

$Z_3 = \frac{15-13.25}{4.65}, \frac{15-13.25}{5.37} = 0.37, 0.32$

$Z_4 = \frac{20-13.25}{4.65}, \frac{20-13.25}{5.37} = 1.45, 1.25$

Answer.

Answer - 2 - (a)

$X = \{10, 12, 15, 18, 22, 23, 32, 34, 78, 31, 14\}$

sort in ascending order.

$= 10, 12, 14, 15, 18, 22, 23, 31, 32, 34, 78.$

$N = 11$

Median = $\left(\frac{N+1}{2}\right)$ th Term

$= \left(\frac{11+1}{2}\right)$ th Term

$= 6$ th Term

$= 22$

min = 10, max = ~~34 78 78~~ 34

$Q_1 = 10, 12, 14, 15, 18$

$Q_3 = 23, 31, 32, 34, 78$

$Q_1 = \left(\frac{N+1}{2}\right)$ th Term

$= \left(\frac{5+1}{2}\right)$ th Term

$= 3^{rd}$ Term

$= 14$

$Q_3 = \left(\frac{N+1}{2}\right)$ th Term

$= \left(\frac{5+1}{2}\right)$ th Term

$= 3^{rd}$ Term

$= 32$

$IQR = Q_3 - Q_1$

$= 32 - 14$

$= 18$

Upper Bound $\rightarrow Q_3 + 1.5 \times IQR$

$= 32 + 1.5 \times 18$

$= 59$

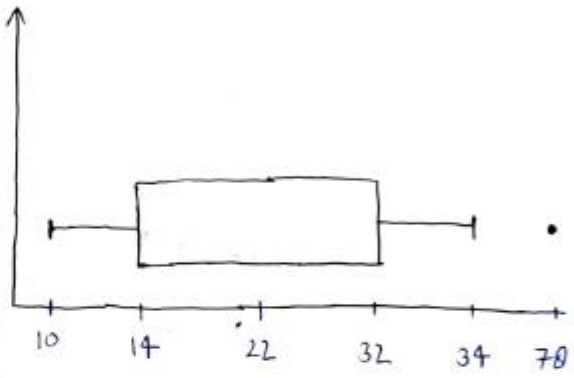
Lower Bound $= Q_1 - 1.5 \times IQR$

$= 14 - 1.5 \times 18$

$= -13$

Hence, $78 > 59$, so 78 is the outlier.

Box - Plot summary $\rightarrow$



Answer.

Answer - 2-(b)

$3x + 5y = 9$

$2x + 3y = 5$

$Ax = B$

$A = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix}$, $x = \begin{bmatrix} x \\ y \end{bmatrix}$, $B = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$

$= \left[\begin{array}{cc|c} 3 & 5 & 9 \\ 2 & 3 & 5 \end{array}\right]$

~~$R_1 \to 2R_1 - 3R_2$~~ $R_1 \to R_1/3$

$= \left[\begin{array}{cc|c} 0 & 1 & 3 \\ 2 & 3 & 5 \end{array}\right]$ $\left[\begin{array}{cc|c} 1 & 5/3 & 3 \\ 2 & 3 & 5 \end{array}\right]$

$R_2 \to R_2 - 2R_1$

$= \left[\begin{array}{cc|c} 1 & 5/3 & 3 \\ 0 & -1/3 & -1 \end{array}\right]$

$R_2 \to -3R_2$

$= \left[\begin{array}{cc|c} 1 & 5/3 & 3 \\ 0 & 1 & 3 \end{array}\right]$

$R_1 \to R_1 - \frac{5}{3} R_2$

$= \left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array}\right]$

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

$x = -2, \quad y = 3$

Answer.

Answer 2-(b)

Another way → using echleon form.

$3x+5y=9$

$2x+3y=5$

$A=\begin{bmatrix}3 & 5\\ 2 & 3\end{bmatrix}$, $x=\begin{bmatrix}x\\ y\end{bmatrix}$, $B=\begin{bmatrix}9\\ 5\end{bmatrix}$

$=\left[\begin{array}{cc:c} 3 & 5 & 9 \\ 2 & 3 & 5 \end{array}\right]$

$R_1 \to R_1/3$

$=\begin{bmatrix}1 & 5/3 & 3\\ 2 & 3 & 5\end{bmatrix}$

$R_2 \to R_2 - 2R_1$

$=\begin{bmatrix}1 & 5/3 & 3\\ 0 & -1/3 & -1\end{bmatrix}$

$\sim \begin{bmatrix}1 & 5/3 \\ 0 & -1/3 \end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}3\\ -1\end{bmatrix}$

$x+\frac{5}{3}y=3$, $-\frac{1}{3}y=-1 \Rightarrow y=3$

$x=3-\frac{5}{3}\times 3=-2 \Rightarrow x=-2,\ y=3$ Answer.

Answer - 3.

Music Name	No. of Downloads	User Review Rating	Number of user Reviews	Whether you liked the Music
A	2.35	7.5	2.3	Y
B	9.87	5	0.5	N
C	0.87	4	7	Y
D	1.2	7.1	5.6	Y
E	5.5	6.2	4.3	N
F	2.3	1.9	2.8	N

$$D_A = \sqrt{(4.8-2.35)^2+(5.5-7.5)^2+(3.1-2.3)^2}$$

$$= \sqrt{10.6425}$$

$$= 3.26$$

$$D_B = \sqrt{(4.8-9.87)^2+(5.5-5)^2+(3.1-0.5)^2}$$

$$= \sqrt{32.7149} \qquad = 5.71$$

$$D_C = \sqrt{(4.8-0.87)^2 + (5.5-4)^2 + (3.1-7)^2}$$

$$= \sqrt{32.9049}$$

$$= 5.73$$

$$D_D = \sqrt{(4.8-1.2)^2 + (5.5-7.1)^2 + (3.1-5.6)^2}$$

$$= \sqrt{21.77}$$

$$= 4.66$$

$$D_E = \sqrt{(4.8-5.5)^2 + (5.5-6.2)^2 + (3.1-4.3)^2}$$

$$= \sqrt{2.42}$$

$$= 1.55$$

$$D_F = \sqrt{(4.8-2.3)^2 + (5.5-1.9)^2 + (3.1-2.0)^2}$$

$$= \sqrt{19.3}$$

$$= 4.39$$

Now, sort the value in ascending order.

Music name	NOD	USR	NUR	WLM	ED
E ~~5~~	~~6.2~~ 5.5	~~4.3~~ 6.2	4.3	No	1.55
A	2.35	7.5	2.3	Yes.	3.26
F	2.3	19	2.8	No	4.39
D	1.2	7.1	5.6	Yes.	4.66
B	9.07	5	0.5	No	5.71
C	0.07	7.5	2.3	Yes.	5.73

K=3, select the first three values.

1.55	No
3.26	Yes.
4.39	No

calculated Inverse.

ED	Inverse	output
1.55	0.645 (1/1.55)	No
3.26	1/3.26 = 0.306	Yes
4.39	1/4.39 = 0.227	No

sum = 0.645 + 0.306 + 0.227

= 1.178.

ED	weights = Inverse/sum	output
1.55	0.547	No
3.26	0.259	Yes
4.09	0.192	No

Net weight for No = 0.547 + 0.192 = 0.739

" " " Yes = 0.259

0.739 > 0.259

so, class is No

Answer.

Answer - 6.

Linear Regression →

X (Year)	Y (Expenditure)	XY	X^2
1	12	12	1
2	19	38	4
3	29	87	9
4	37	148	16
5	45	225	25
$\Sigma\bar{X} = 15/5 = 3$	$\Sigma\bar{Y} = 28.4$	$\Sigma\overline{XY} = 101.2$	$\overline{X^2} = 11$

$$\bar{y} = \alpha x + \beta$$

$$\alpha = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$$= \frac{102 - 85.2}{11 - 3^2} \quad 3 \times 28.4$$

$$= \frac{102 - 85.2}{11 - 9}$$

$$= 8.4$$

$$\beta = \bar{y} - \alpha\bar{x}$$

$$= 28.4 - 8.4 \times 3$$

$$= 28.4 - 25.2$$

$$= 3.2$$

$$\bar{y} = 8.4x + 3.2$$

Expenditure in 6^{th} year $= 8.4 \times 6 + 3.2$

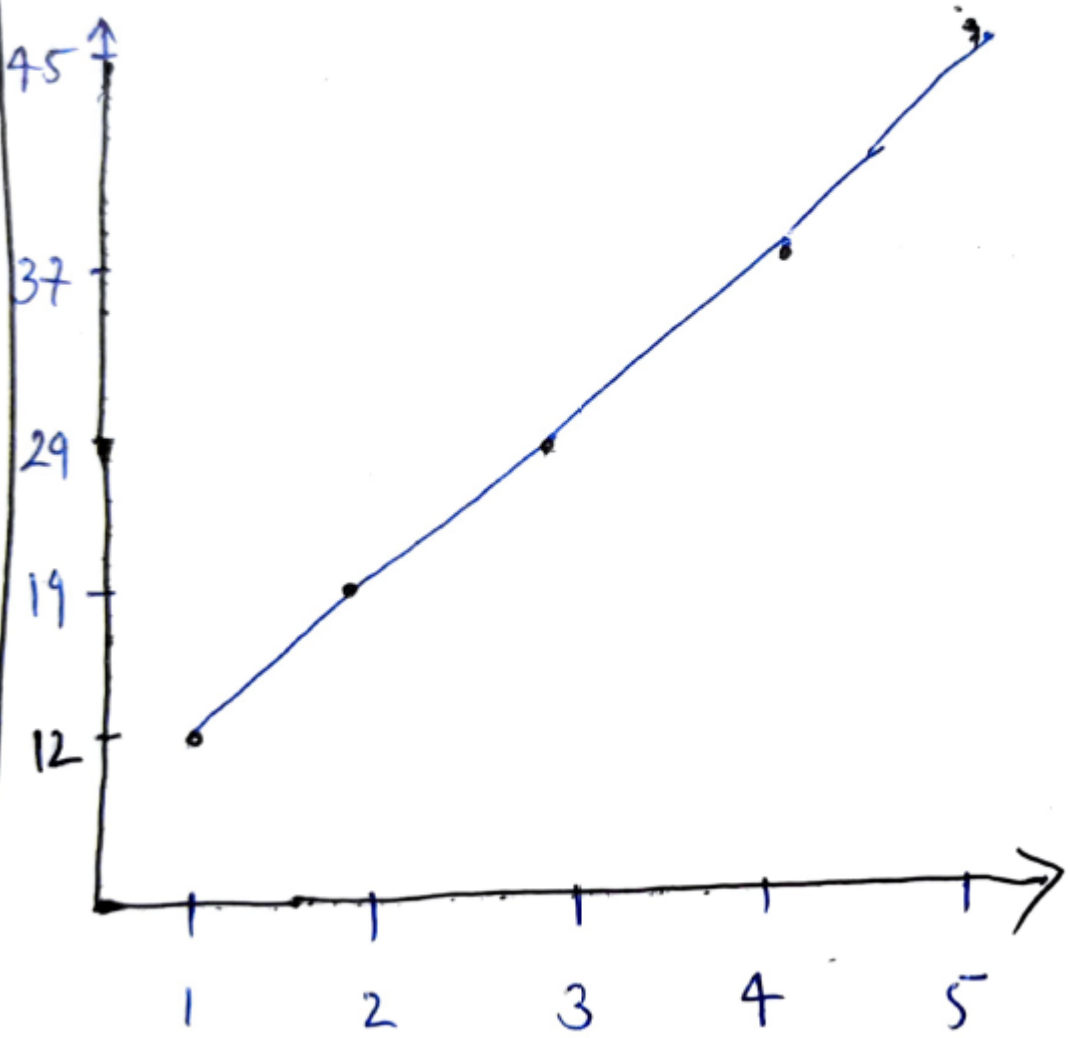
$= 53.6$

Expenditure in 8^{th} year $= 8.4 \times 8 + 3.2$

$= 70.4$

Answer.

Regression line



CO-PO and CO-PSO Mapping																				
Course Outcomes		**Blooms Level**	**Modules covered**	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**	**PSO3**	**PSO4**
CO1	Apply the knowledge of searching and reasoning techniques for different applications.	L2	M1	3	3	2	3	3	1	-	-	1	-	-	1	-	-	-	2
CO2	Have a good understanding of machine leaning in relation to other fields and fundamental issues and challenges of machine learning.	L3	M2	3	3	2	3	3	1	-	-	1	-	-	1	-	-	-	2
CO3	Apply the knowledge of classification algorithms on various dataset and compare results.	L3	M3	3	3	2	3	3	1	-	-	1	-	-	1	-	-	-	2
CO4	Model the neuron and Neural Network, and to analyze ANN learning and its applications.	L3	M4	3	3	2	3	3	1	-	-	1	-	-	1	-	-	-	2
CO5	Identifying the suitable clustering algorithm for different pattern.	L3	M5	3	3	2	3	3	1	-	-	1	-	-	1	-	-	-	2

COGNITIVE LEVEL	REVISED BLOOMS TAXONOMY KEYWORDS
L1	List, define, tell, describe, identify, show, label, collect, examine, tabulate, quote, name, who, when, where, etc.
L2	summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend
L3	Apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover.
L4	Analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer.
L5	Assess, decide, rank, grade, test, measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize.

PROGRAM OUTCOMES (PO), PROGRAM SPECIFIC OUTCOMES (PSO)				CORRELATION LEVELS	
PO1	Engineering knowledge	PO7	Environment and sustainability	0	No Correlation
PO2	Problem analysis	PO8	Ethics	1	Slight/Low
PO3	Design/development of solutions	PO9	Individual and team work	2	Moderate/ Medium
PO4	Conduct investigations of complex problems	PO10	Communication	3	Substantial/ High
PO5	Modern tool usage	PO11	Project management and finance		
PO6	The Engineer and society	PO12	Life-long learning		