

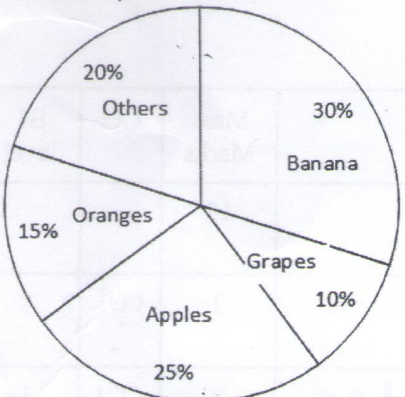
07/06/2023

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

Subject Code: HDSC501

Subject Name: Mathematics for Data Science/AIML

May-June 2023 (B.Tech / M.Tech.) Program: EXTC/IT/COMP/AIDS Examination: TY Semester: V				
Course Code: HDSC501		Course Name: Mathematics for Data Science/AIML		
Duration: 02 Hours 25		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)	Two of the eigenvalues of $3 \times 3$ matrix are -1,2. If the determinant of the matrix is 4. Find the third eigenvalue.	2	CO1	3
ii)	A continuous random variable X has the probability law $f(x) = kx^2, 0 \leq x \leq 2$ then find $P(0.2 \leq x \leq 0.5)$	2	CO2	3
iii)	Differentiate between Quantitative and Qualitative data.	2	CO3	3
iv)	Explain in short: What is imputation of data?	2	CO4	3
v)	Compute the roots of function $f(x) = x^2 - 3$ after two iterations using Bisection Method in the interval $[0,2]$ .	2	CO5	3
vi)	Differentiate between PCA and LDA	2	CO6	3
vii)	For the below stem and leaf plot what is the mode? Stem Leaf 1    0 1 1 5 9 2    1 3 6 7 8 3    0 1 3 3 3 6 7 4    2 3 5 5 7 9 5    3 4 8 7    1 7 9 8    4	2	CO3	3
viii)	Write in short, benefits of data cleaning.	2	CO4	3
Q.2	Solve any four questions out of six.	16		

i)	Solve the following equations by Gauss Jordan Method $x + 3y - 2z = 5, 2x + y - 3z = 1, 3x + 2y - z = 6$	4	CO1	3										
ii)	A random sample of size 16 from a normal population showed a mean of 103.75 cm. and sum of squares of deviations from the mean 843.75 $cm^2$ . Can we say that the population has a mean of 108.75 $cm^2$ ?	4	CO2	3										
iii)	<p>The below pie chart shows the sale of different fruits in a day for a shop:</p>  <p>Answer the following questions based on the pie chart: (i) If a total of 1200 kg of fruits were sold in a day, calculate the amount of bananas sold (in kg). (ii) Calculate the central angle for the category "oranges".</p>	4	CO3	3										
iv)	Explain the need of Exploratory data analysis.	4	CO4	3										
v)	Using Newton's Method find the optimum value of $f(x) = 4\sin x (1 + \cos x)$ .	4	CO5	3										
vi)	Explain steps involved in Linear Discriminant Analysis.	4	CO6	3										
Q.3	Solve any two questions out of three.	16												
i)	Find the dimension and basis for the four fundamental subspaces for $A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix}_{3 \times 4}$	8	CO1	3										
ii)	<p>To test the effect of a new drug, a controlled experiment was conducted. 300 patients were given the new drug while 200 patients were given no drug. On the basis of examination of these persons, the following results were obtained.</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th></th> <th>Cured</th> <th>Condition worsened</th> <th>No effect</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>New Drug</th> <td>200</td> <td>40</td> <td>60</td> <td>300</td> </tr> </tbody> </table>		Cured	Condition worsened	No effect	Total	New Drug	200	40	60	300	8	CO2	3
	Cured	Condition worsened	No effect	Total										
New Drug	200	40	60	300										

	<table border="1"> <tr> <td>given</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>No drug given</td> <td>120</td> <td>30</td> <td>50</td> <td>200</td> </tr> <tr> <td>Total</td> <td>320</td> <td>70</td> <td>110</td> <td>500</td> </tr> </table> <p>Use <math>\chi^2</math> test to find the effect of the new drug.</p>	given					No drug given	120	30	50	200	Total	320	70	110	500			
given																			
No drug given	120	30	50	200															
Total	320	70	110	500															
iii)	<p>Observe the graph below and answer the following question:</p> <p style="text-align: center;"><b>Result in a School</b></p> <p>(a) Which year has the smallest difference between the number of kids who passed and those who failed?                  (b) In the last five years, what was the average number of kids who failed in school?                  (c) How many times have the same numbers of kids failed?                  (d) What is the ratio of number of students failed in the years 1993-94 and 1995-96?</p>	8	CO3	3															
Q.4	Solve any two questions out of three.	16																	
i)	Explain all the steps involved in cleaning and preparing data.	8	CO4	3															
ii)	Use Steepest descent method for $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from $X = (0,0)$ .	8	CO5	3															
iii)	Explain all the steps involved in Principal Component Analysis.	8	CO6	3															

\*\*\*\*\*