U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: III**

Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 23AM3PCPSM **Max Marks: 100**

Course: Probability and Statistics for Machine Learning

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Explain using an example of how quantification of uncertainty through probabilistic forecast/models enhance decision making process in Machine learning Applications.	*CO1*	*PO1*	**06**
	b)	The Joint probability distribution of two random variables X and Y is given by: P (X=0, Y=1) = 1/3 P (X=1, Y=-1) = 1/3 and P (X=1, Y=1) = 1/3. Find Marginal distribution of X and Y.	*CO2*	*PO1*	**07**
	c)	A survey of 1,000 employees in a large organization revealed the following: • 65% participate in the company's wellness program. • 50% regularly attend skill-development workshops. • 40% participate in both the wellness program and skill-development workshops. Using this information, calculate: a. The percentage of employees who participate in either the wellness program or the skill-development workshops. b. The percentage of employees who do not participate in either activity. c. The percentage of employees who participate in the wellness program but do not attend skill-development workshops.	*CO3*	*PO2*	**07**
		OR			
2	a)	Comment on the following a. When do you say two events are independent? b. State the Multiplication theorem of Probability for independent events. c. Are two independent events mutually exclusive.	*CO1*	*PO1*	**06**
	b)	Among the Employees of a certain firm, 70% know Java, 60% know Python and 50% know both languages. What portion of programmers a. does not know Python? b. knows Java but not Python? c. does not know both Python and Java?	*CO2*	*PO2*	**07**

		c)	Two Computers A and B are to be marketed. A salesman who is assigned the job of finding customers for them has 60% and 40% chances respectively of succeeding in case of computer A and B. The two computers can be sold independently. Given that he was able to sell at least one computer, what is the probability that computer A has been sold?	*CO3*	*PO2*	**07**
			UNIT - II			
	3	a)	Comment on the following: a. Explain how Maximum a posteriori probability (MAP) can be used to estimate the unknown parameter from data b. Also justify how MAP estimates are better when compared to MLE estimates c. State the conditions when MAP estimate reduces to MLE d. How does Maximum Likelihood Estimation (MLE) differ from Maximum A Posteriori (MAP) Estimation in terms of their approach and focus?	*CO1*	*PO1*	**10**
		b)	Consider a dataset given below related to providing loan to customers based on their features such as income > 30K, Status of Employment, loan amount > 1 lakh Now given applicant whose income is < 30K and is unemployed has applied for a loan for 3K, will the bank give him / her a loan.	*CO2*	*PO3*	**10**
			OR			
	4	a)	a. Sate Bayes' Theorem and explain each of its terms **b.** The chances that a doctor 'A' will Diagnose a disease 'X' correctly is 60%. The chances that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor 'A', who had disease X, died. Use Bayes' theorem to compute the chance that his disease was diagnosed correctly?	*CO1*	*PO2*	**10**
		b)	Apply Naïve Bayes method to classify emails as either "spam" or "not spam" based on the frequency of words in the emails.	*CO2*	*PO3*	**10**

Applicant ID	Income (>30 K)	Employment Status	Loan amount (>5Lakh)	Was loan given
1.	Yes	Employed	Yes	No
2.	No	Unemployed	No	Yes
3.	Yes	Employed	Yes	No
4.	Yes	Employed	Yes	No
5.	No	Employed	No	Yes
6.	No	Employed	Yes	No
7.	No	Unemployed	No	Yes
8.	Yes	Employed	Yes	Yes

Email no.	"win"	"offer"	"Money"	Class(label)
1	2	1	1	spam
2	1	1	1	Spam
3	0	1	1	Not spam
4	0	0	1	Not spam

			UNIT - III			
	5	a)	Comment on the following a. Define random variable. How to calculate its Expectation and Variance. b. Differentiate between Discrete and continuous random variable.	*CO1*	*PO1*	**06**
		b)	An exciting computer game is released. Sixty percent of players complete all the levels. Thirty percent of them will then buy an advanced version of the game. Among 15 users, what is the expected number of people who will buy the advanced version? What is the probability that at least two people will buy it?	*CO2*	*PO2*	**07**
		c)	A computer user tries to recall her password. She knows it can be one of 4 possible pass words. She tries her passwords until she finds the right one. Let X be the number of wrong passwords she uses before she finds the right one. Find E(X) and Var(X).	*CO3*	*PO2*	**07**
			OR			
	6	a)	If X is a Poisson variable such that $P(X = 2) = 9P(X = 4)$. **i.** Compute the value of λ. **ii.** Compute $P(X = 0)$.	*CO1*	*PO1*	**06**
		b)	Two random variables X and Y have the joint distribution, $P(0,0) = 0.2, P(0,2) = 0.3, P(1,1) = 0.1, P(2,0) = 0.3, P(2,2) = 0.1$, and $P(x, y) = 0$ for all pairs (x, y). i. Find the probability mass function of $Z = X + Y$ ii. Find the probability mass function $U = X - Y$. **iii.** Find the probability mass function $V = XY$.	*CO2*	*PO2*	**07**
		c)	Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day then What is the probability that more than 8 new accounts will be initiated today?	*CO3*	*PO2*	**07**
			UNIT - IV			
	7	a)	The Life times of a certain kind of electronic devices have a mean of 300 hours and a S.D of 25 hours. Assuming that the distribution of these life times to be approximated closely to normal curve. a. Find the probability that any one of these electronic devices will have a life time of more than 350 hours b. What percentage will have lifetimes of 300 hours or less?	*CO2*	*PO1*	**06**
		b)	The installation time, in hours, for a certain software module has a probability density function $f(x) = k(1 - x^3)$ for 0 < x < 1. Find k and compute the probability that it takes less than 1/2 hour to install this module.	*CO2*	*PO2*	**07**
		c)	Let Z be a Standard Normal random variable. Compute i. $P(Z < 1.25)$ ii. $P(Z < 6.0)$ iii. With probability 0.8, variable Z does not exceed what value?	*CO3*	*PO2*	**07**
			OR			

	8	a)	Uploading a large file of 68 documents to a server requires sending packets of data sequentially. The time it takes to send each packet is random, but on an average, it takes 5 milliseconds with a variance of 2 milliseconds^2. a. What is the probability that the whole file is uploaded in less than 0.25 seconds? **b.** How much time it takes to upload 50% of the files	*CO1*	*PO2*	**06**
		b)	Lifetime of a certain hardware is a continuous random variable with density $$f(x) = \begin{cases} k - \frac{x}{50} \; for \; 0 < x < 10 \; years \\ 0 \quad for \; all \; other \; x \end{cases}$$ i. Find K. ii. What is the probability of a failure within the first 5 years?	*CO2*	*PO1*	**07**
		c)	The average height of professional basketball players is around 6 feet 7 inches, and the standard deviation is 3.89 inches. Assuming Normal distribution of heights within this group, i. What percent of professional basketball players are taller than 7 feet? ii. If a player is within the tallest 20% of all players, what can his height be?	*CO3*	*PO1*	**07**
			UNIT - V			
	9	a)	Consider the following nicotine levels of 40 smokers: 00 87 173 253 1 103 173 265 1 112 198 266 3 121 208 277 17 123 210 284 32 130 222 289 35 131 227 290 44 149 234 313 48 164 245 477 86 167 250 491 Find the quartiles and 40th percentile.	*CO2*	*PO2*	**10**
		b)	What are the advantages and disadvantages of using a sample instead of the entire population in statistical studies?	*CO1*	*PO1*	**10**
			OR			
	10	a)	Calculate Q_1, Q_3, D_6, P_{85} from the following data: x: 10 11 12 13 14 15 16 17 18 f: 3 4 5 12 10 7 5 2 1	*CO2*	*PO2*	**10**
		b)	Consider a hypothetical dataset of daily closing stock prices for a particular company over a month (in dollars):12.25, 13.50, 14.80, 15.10, 15.10,16.00, 16.50, 16.90, 17.50, 17.90, 32.00, 18.00. Use 1.5(IQR) rule to identify the potential outliers.	*CO1*	*PO3*	**10**

******