Program: F.Y.B.Sc IT **Semester:** I **Program Code:** UGIT01

Course: Quantitative Foundations for Data Analysis **Course Code:** NUIT107

Duration: 1 Hour **Examination Pattern:** NEP-Autonomous-External **Max. Marks:** 30

Instructions:

1. All questions are compulsory with internal choice.
2. Write clearly question and sub-question number.
3. Symbols have their usual meanings.
4. Scientific calculator fx 82 series or lower version is only permitted.
5. Graph paper will be provided, if required.
6. CO (Course Outcome), KL (Knowledge Level).

Examination:

REGULAR

Q. 1		**Attempt any THREE of the following.**	[15]	CO's	KL's
	(a)	Daily Transaction Log (in 1000s): 15, 20, 25, 30, 35 (i) Write Σ notation for total. (ii) Find average transactions. (iii) Identify sequence type. (iv) Find value on day 10. (v) Total for 10 days.		CO1	L2
	(b)	Let P = "Sensor active" and Q = "Threshold exceeded" (i) Construct Truth table for $P \rightarrow Q$ (ii) When will alarm trigger? (iii) Identify rows with false output. (iv) When is system safe? (v) Check whether $P \rightarrow Q \equiv \neg Q \rightarrow \neg P$		CO2	L2, L3
	(c)	Course enrollments for Artificial Intelligence (AI) and Data Science (DS) is as given. AI = {A, B, C, D, E}, DS = {C, D, F, G} (i) Find students enrolled in both. (ii) Find students enrolled only in AI. (iii) Total unique students. (iv) AI × DS pairs. (v) Map AI student to project score: A:85, B:80, C:90, D:95, E:88. State domain, range.		CO1	L3
	(d)	A bank grants loans if conditions on CIBIL score (P), income proof (Q), and no existing loan (R) are met. (i) Frame a compound proposition combining: "If P and Q or R, then loan granted." (ii) Construct the truth table for $(P \wedge Q) \vee R$. (iii) Identify logical equivalence for $\neg(P \vee Q)$ and $(\neg P \wedge \neg Q)$. (iv) Write the contrapositive and converse of: "If CIBIL score ≥ 750, then loan granted." (v) Give an example of a real-world data query using this rule in SQL-like syntax.		CO2	L4

	(e)	A stock market dashboard streams trade count: 100, 200, 400… (i) Identify sequence type. (ii) Compute trades in 15 minutes. (iii) Use aggregation function to find MAX, AVG. (iv) Frame a logical proposition for "If trade count doubles, notify admin." (v) Describe how propositional reasoning enhances real-time alert systems.		CO1, CO2, CO5	L3, L4, L5
Q. 2		**Attempt any THREE of the following.**	**[15]**	**CO's**	**KL's**
	(a)	Hospital records provide the following information: • Probability a random patient is COVID-positive: $P(COVID) = 0.10$ • False positive rate of the diagnostic test: $P(test + \mid no\ COVID) = 0.05$ • Optimistic sensitivity: $P(test + \mid COVID) = 0.95$ • Oxygen saturation (SpO_2) data for patients (in %): COVID-positive group: Mean=88, SD=4 COVID-negative group: Mean=95, SD=2 Answer the following: (i) Represent the means of oxygen levels as a 2×1 column vector, with the COVID-positive group first. (ii) Represent the standard deviations of oxygen levels as a 2×1 column vector, with the COVID-positive group first. (iii) Using Baye's Theorem, compute the probability that a patient actually has COVID given the test result is positive. (iv) Based on the given SD values, identify which group (COVID-positive or COVID-negative) shows lower variation in oxygen levels. Briefly explain your reasoning.		CO3, CO4	L3, L4
	(b)	A company is analyzing customer ratings (out of 10) for two products A and B from 4 users. Ratings are stored in the matrix $M = [8\ 7\ 6\ 9\ 5\ 8\ 7\ 6]$, where the first column represents ratings for product A and the second column for product B. Answer the following. (i) Write the scalar, vector and matrix present in the data set. (ii) If a customer is chosen at random, what is the probability that they rated product B atleast 8. (iii) Calculate the variance and ratings for product A. (iv) Draw boxplot to compare ratings of both products.		CO3, CO5	L3
	(c)	The following data represents the cloud response times (in milliseconds): 2, 4, 6, 8, 10, 12 (i) Construct a frequency polygon for the given data. (ii) Calculate the variance of the response times.		CO5	L3, L4

		(iii) Compute the deciles D_3. (iv) Find the standard Deviation. (v) Based on the data, identify the kurtosis (whether the distribution is flat, normal, or peaked).			
	(d)	An e-commerce platform uses a simple probabilistic model to decide whether to recommend a product. Based on past data: • Probability a visitor is a new user = 0.4 • Probability a visitor is a returning user = 0.6 • Probability a new user clicks on a recommendation = 0.3 • Probability a returning user clicks on a recommendation = 0.7 (i) Define the sample space for the type of user and click behavior (ii) What is the probability that a randomly chosen visitor is a returning user and clicks on a recommendation? (iii) What is the probability that a visitor clicks on a recommendation (regardless of type)? (iv) If a visitor is observed to click, what is the probability that they are returning user? (v) Explain in 2-3 lines how such a model helps improve recommendation systems in industry.		**CO4**	**L4**
	(e)	An image row (1×3) is represented by the vector $I = [2, 4, 6]$. A linear transformation (feature extractor) is given by the matrix $E = [1\ 0\ -\ 1\ 0\ 1\ 0\ ]$. (i) State the dimensions of I (as a column vector) and of E. (ii) Compute the matrix-vector product $E \times I^T$. (iii)Write the resulting transformed vector. (iv)How many features did the transformation reduce the data to? State the reduced dimension. (v) In 2-3 lines explain why such a linear transformation can be useful in AI image processing or compression (mention feature extraction or dimensionality reduction)		**CO3**	**L3, L4**

- - X - - X - -