National University of Computer and Emerging Sciences 0228

Lahore Campus

Probability & Statistics (MT2005)

Date: April 12, 2025

Course Instructor(s)

Ms. Sarah Ahmad

Ms. Kanwal Saleem

Ms. Huma Akbar

Dr. Nuzhat Aftab

MID-2 Exam

Total Time (Hrs):	**1**
Total Marks:	**25**
Total Questions:	**2**
ALL SECTIONS	**BCS,BSE**

[redacted] Roll No

[redacted] Section

[redacted] Student Signature

Do not write below this line

Strictly Attempt Q1 on the Answer Book, Solve Table on.2 on Question Paper and Attach It.

Lead Pencil work wouldn't be marked or claimed for rechecking. Use permanent ink pen.

If data is ambiguous, make an assumption and proceed—do not ask the invigilator.

Please read carefully and choose the correct variables and units before attempting your solution.

If you select the wrong variable(s), incorrect values and wrong cases etc., your entire answer will be awarded zero marks, even if the working steps are shown.

Q2/Table no.2 requires a direct answer. Avoid unnecessary details, cutting, or overwriting in a cell, as it will result in a zero score. Use a rough sheet for calculations.

02

***CLO 2:* Probability and Distribution Analysis**: Apply foundational principles of probability to analyze experiments, including Bayes' theorem, evaluate discrete and continuous distributions, and explore applications in machine learning.

Q1: **[Marks:15]**

A Case Study of Bayes for Sentiment Analysis: This task involves classifying tweets as positive or negative based on probabilistic reasoning. You are provided with a dataset of raw tweets, each labelled as Positive or Negative. Your objective is to apply Bayesian Classification to determine the sentiment of a new tweet.

Task: You are given a new tweet containing three words: ☞ Happy ☺ Learning ☺ NLP. Using Bayes' theorem, compute its probability of being Positive and Negative based on the given dataset. Determine whether the new tweet is classified as positive or negative and write your conclusion.

*Note: The words in a tweets are not case sensitive.

* Assuming emojis are not to be taken into account.

TABLE NO.1: TWEETS

Positive Tweets:	**Negative Tweets:**
1. I am feeling so happy today! ☺ 2. Learning new things makes me happy and excited! 3. Happy to be part of this amazing NLP workshop! 4. This course on learning NLP is fantastic! 😃 5. Happy moments happen when we keep learning every day! 6. Exploring new NLP techniques is truly exciting! ☺ 7. I love learning about AI and NLP applications! 8. Every day is a chance to grow. Keep learning and stay happy! 9. Applying NLP models has been a wonderful learning experience! 10. A happy journey into NLP and deep learning!	1. I am not happy with these difficult NLP concepts. 2. Machine learning is so confusing and not happy at all! 3. This NLP model keeps failing. What a bad day! 4. I hate learning new things when they are this complex. 5. Happy? Not when debugging this machine learning code. 6. This deep learning model takes forever to train, so frustrating! ☹ 7. NLP seems overrated and not very useful. 8. I'm struggling with NLP and Probability; it's making me really upset. ☹

Positive: 1.

6

CLO 2: **Probability and Distribution Analysis**: Apply foundational principles of probability to analyze experiments, including Bayes' theorem, evaluate discrete and continuous distributions, and explore applications in machine learning.

Q2: **[Marks:10]**

TABLE NO.2

Sr.	SCENARIOS	RESPONSE
1	The proportion of people who respond to a certain email is modeled as a continuous random variable with the probability density function: $f(x) = \frac{2(x+2)}{5}$, where $0 < x < 2$. Is this a valid pdf?	o Yes (o) No 2 $f(x) > 1$
2	A coffee shop tracks the number of loyalty card holders (X) who redeem their free coffee reward on Mondays. The probability distribution of X is $[x, f(x)]$: [2, 1/10], [3, 1/11], [4, 1/6], [5, 1/7]. The shop also offers an additional discount (in hundreds) on further purchases, calculated as $g(X) = X + 2$. Find the shop's expected total discount amount on Mondays.	$\$\frac{3140}{11} = \$285.4545$ 2
3	A software engineering team is analyzing the relationship between code compilation time (X) in seconds and model accuracy (Y) as a proportion for a machine learning system. The following statistics are given for a set of experiments: $\sigma_X = 2.5$, $\sigma_Y = 0.70$, $\sigma_{XY} = -1.20$. Calculate the correlation between X and Y.	-0.9071
4	A network administrator is monitoring the time (X) taken to transmit a data packet across different routers in a network. Due to network traffic variations, the transmission time X is a continuous random variable with a probability density function, $f(x) = \frac{1}{b-a}$ with $a \le x \le b$, for the interval [-2, 10]. Calculate and define its probability function.	$f(x) = \frac{1}{12}$ for $-2 \le x \le 10$
5	A database server processes read and write queries in parallel. Both query types are randomly distributed, representing the number of queries completed within a fixed 1-second interval. Based on historical logs, the system administrator aims to analyze the probability distribution of read and write queries to assess system performance and reliability under varying load conditions. How would you represent the distribution: Is this a univariate or joint distribution model? Also, classify it as discrete or continuous.	o Univariate (o) Bivariate 1 (o) Discrete 1 o Continuous

Q2 requires a direct answer. Avoid unnecessary details, cutting, or overwriting in a cell, as it will result in a zero score. Use a rough sheet for calculations.

*Strictly attempt here accordingly otherwise you will lose marks. *Roll no.*[redacted] *Sec*[redacted]