National University of Computer and Emerging Sciences

Lahore Campus

Rubric M2

MARKS: Working = 13 + 2 = Conclusion

TABLE NO.1: TWEETS	
Positive 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**!	**Negative Tweets:** 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. ☹
Pos 10 **Happy = 6** **Learning = 7** **NLP = 6**	Neg 8 **Happy = 3** **Learning = 4** **NLP = 4**

FORMULA or RELATED INFO: (1)

P(**Happy** | Pos) P(**Learning** | Pos) P(**NLP** | Pos) ***P(Pos)***

(6/10) (7/10) (6/10) (10/18)

0.6 x 0.7 x 0.6 x 0.56 = **0.14112**

(1) (1) (1) (1)

0.14112/0.18237 = **0.7738** (2)

Alternatively find Neg and 1-0.2262= 0.7738

P(**Happy** | Neg)P(**Learning** | Neg)P(**NLP** | Neg)P***(Neg)***

(3/8) (4/8) (4/8) (8/18)

0.375 x 0.5 x 0.5 x 0.44 = **0.04125**

(1) (1) (1) (1)

0.04125/0.18237 = **0.2262** (2)

Alternatively find Pos and 1-0.7738 = 0.2262

Conclusion: New Tweet is classified as Positive because it has higher probability 0.7738. (2)

TOTAL = P(**Happy** | Pos) P(**Learning** | Pos) P(**NLP** | Pos) **P(Pos)**+ P(**Happy** | Neg)P(**Learning** | Neg)P(**NLP** | Neg)P***(Neg)***

(6/10) (7/10) (6/10) (10/18) + *(3/8) (4/8) (4/8) (8/18)*

0.6 x 0.7 x 0.6 x 0.56 + ***0.375 x 0.5 x 0.5 x 0.44***

National University of Computer and Emerging Sciences

Lahore Campus

`0.14112 + 0.04125 = 0.18237`

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: 2 Each Scenario =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~~ ✓ No
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.	question error(probs not typed completely for all values of X)
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.	- 0.69
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 $a \leq x \leq b$, for the interval [-2, 10]. Calculate and define its probability function.	$f(x) = \begin{cases} \frac{1}{12} & -2 \leq x \leq 10 \\ 0 & Otherwise \end{cases}$
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~~ ✓ Bivariate ✓ Discrete ~~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.*_____ *Sec*;_____