

# Week-3 PSOSM 2023

1. Consider the below response:

“To this segment, consumer privacy is very important; they feel that they have been victims of privacy invasions, they are pessimistic about the future of privacy protection, and about a third of them favour creating a general federal regulatory agency on consumer privacy.”

To which of the three Westin Privacy categories is "they" in the above sentence most likely to fall?

- a. Fundamentalists
- b. Pragmatists
- c. Unconcerned
- d. None of the above

2. Which of the following features is used by the credibility model- TweetCred?

- a. Tweet links
- b. Tweet Author
- c. Tweet network
- d. Tweet meta data

3. Cronbach's alpha

- a. determines the statistical significance of a research study.
- b. assess the reliability and internal consistency of a questionnaire.
- c. analyze the effect size of an intervention.
- d. calculate the standard deviation of a sample population.

4. Suppose we have five tweets (TW1, TW2, TW3, TW4, and TW5) with their relevances based on user feedback or relevance labels:

TW1: Highly relevant tweet (Relevance = 2)

TW2: Moderately relevant tweet (Relevance = 1)

TW3: Not relevant tweet (Relevance = 0)

TW4: Relevant tweet (Relevance = 1)

TW5: Highly relevant tweet (Relevance = 2)

Now, let's assume the ranking algorithm generates a ranked list of tweets:

Rank 1: TW5 (Highly relevant tweet)

Rank 2: TW1 (Highly relevant tweet)

Rank 3: TW3 (Not relevant tweet)

Rank 4: TW2 (Moderately relevant tweet)

Rank 5: TW4 (Relevant tweet)

What would be NDCG@5 for this ranked list of tweets?

- a. 0.5
- b. .513
- c. .975
- d. None of the above

Solution:

Step 1: Calculate DCG@5 (Discounted Cumulative Gain at rank 5).

Step 2: Calculate IDCG@5 (Ideal Discounted Cumulative Gain at rank 5).

Step 3: Divide DCG@5 by IDCG@5 to get NDCG@5.

Step 1: Calculate DCG@5  $DCG@5 = rel_1 + rel_2/\log_2(2) + rel_3/\log_2(3) + rel_4/\log_2(4) + rel_5/\log_2(5)$

Given relevances:  $rel_1 = 2$  (for TW5)  $rel_2 = 2$  (for TW1)  $rel_3 = 0$  (for TW3)  $rel_4 = 1$  (for TW2)  $rel_5 = 1$  (for TW4)

$DCG@5 = 2 + 2/\log_2(2) + 0/\log_2(3) + 1/\log_2(4) + 1/\log_2(5)$   $DCG@5 = 2 + 2/1 + 0 + 1/2 + 1/2$   
 $DCG@5 = 4 + 0.5 + 0.5$   $DCG@5 = 5$

Step 2: Calculate IDCG@5 (Ideal DCG@5) To calculate IDCG@5, we consider the ideal ranking based on relevance:

Ideal ranking: TW5 (Highly relevant tweet) > TW1 (Highly relevant tweet) > TW2 (Moderately relevant tweet) > TW4 (Relevant tweet) > TW3 (Not relevant tweet)

$IDCG@5 = rel_1 + rel_2/\log_2(2) + rel_3/\log_2(3) + rel_4/\log_2(4) + rel_5/\log_2(5)$

For ideal ranking:  $rel_1 = 2$  (for TW5)  $rel_2 = 2$  (for TW1)  $rel_3 = 1$  (for TW2)  $rel_4 = 1$  (for TW4)  $rel_5 = 0$  (for TW3)

$IDCG@5 = 2 + 2/\log_2(2) + 1/\log_2(3) + 1/\log_2(4) + 0/\log_2(5)$   $IDCG@5 = 2 + 2/1 + 1/1.585 + 1/2 + 0$   $IDCG@5 = 2 + 2 + 0.63 + 0.5$   $IDCG@5 = 5.13$

Step 3: Calculate NDCG@5  $NDCG@5 = DCG@5 / IDCG@5$

$NDCG@5 = 5 / 5.13$

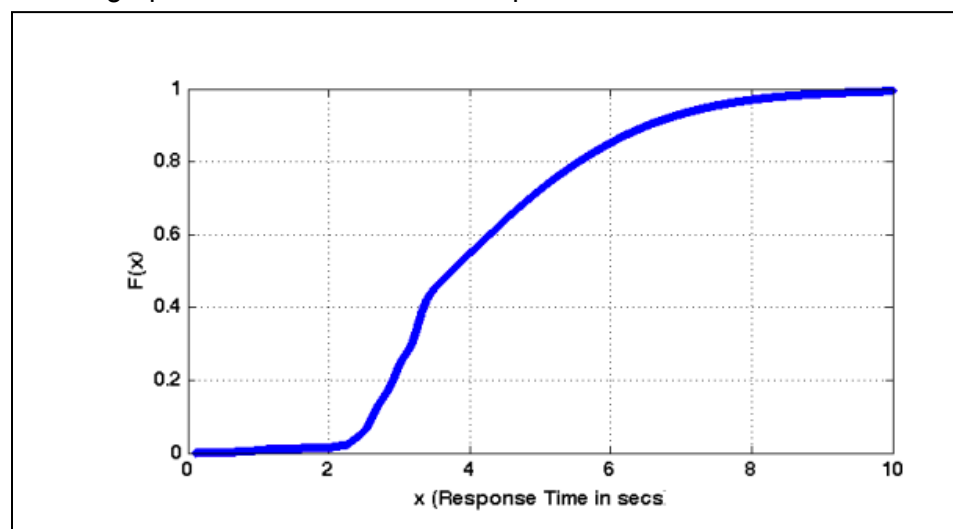
$NDCG@5 \approx 0.975$  (rounded to three decimal places)

So, the NDCG@5 for the ranked list of tweets is approximately 0.975.

Read through the report “Privacy in India: Attitudes and Awareness V 2.0” at [https://cdn.iiit.ac.in/cdn/precog.iiit.ac.in/research/privacyindia/PI\\_2012\\_Complete\\_Report.pdf](https://cdn.iiit.ac.in/cdn/precog.iiit.ac.in/research/privacyindia/PI_2012_Complete_Report.pdf) And answer the following questions [5-7].

5. How many participants have changed their default settings on their social network?
- a. 75%
  - b. 60%
  - c. 50%
  - d. 25%
6. What is the primary purpose of a Focus Group Discussion (FGD) in the above report?
- a. To conduct one-on-one interviews with participants.
  - b. To collect quantitative data through surveys and questionnaires.
  - c. To observe participants' behaviours in a controlled environment.
  - d. To facilitate group discussions and gather qualitative insights.
7. “When you hear the word privacy, what comes to mind”. Which is the most common respondent answer according to the above report?
- a. Information Privacy
  - b. Bodily privacy
  - c. Communication privacy
  - d. Territorial privacy

8. Consider the below graph and choose the correct option:



- a. For 82% of the users, the response time was less than 2 seconds, and for 99%, the response time was under 10 seconds.
- b. For 99% of the users, the response time was less than 6 seconds, and for 100%, the response time was under 10 seconds.
- c. For 82% of the users, the response time was less than 6 seconds, and for 99%, the response time was under 10 seconds.
- d. None of the above

9. What is the Likert scale?

- a. rating scale in surveys and questionnaires
- b. gives the score for any real-time tweets
- c. Interannotator agreement
- d. None of the above

10. TweetCred is

- a. a real-time web-based system to automatically evaluate the credibility of content on Twitter.
- b. a real-time app-based system to automatically evaluate the score of tweets on Twitter.
- c. a real-time app-based system to automatically assess the quality of content on Twitter.
- d. None of the above