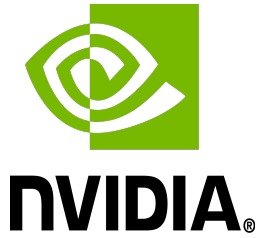


Data Challenge '21

The submission evaluation process is outlined in the following slides

Scoring Guidelines

Breakdown of the score per task



1 – Predict the correct fault

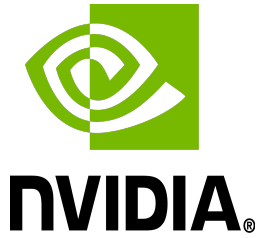


Overview

For the 1st part of the challenge, we use the classification *Accuracy* to score the capability to predict the correct fault.

- We compute the classification accuracy using **the Test dataset only**.
- To compute it, we use the ground truth labels and the labels produced by the **Log Performance only**.

$$0 \leq \text{Accuracy} \leq 1$$



1 – Predict the correct fault

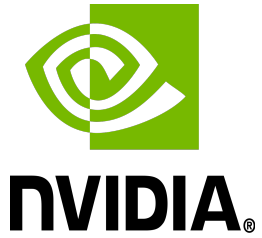


Evaluation Process

1. We compute the classification *Accuracy*

Accuracy = Accuracy (ground truth, Log Performance)

$$0 \leq \text{Accuracy} \leq 1$$



2 – Identify the signals having anomalous behaviour



Overview

For the 2nd part of the challenge, we define the *Average ranking score* to score the capability to identify the correct most important signal.

- We use **the Test dataset only**.
- To compute it, we use ranking produced by the **Log Performance function only**.

2 – Identify the signals having anomalous behaviour

Evaluation Process

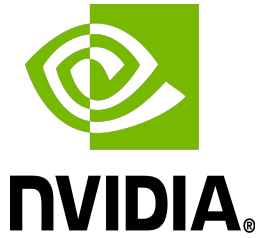
1. For each experiment e in *Test Dataset (TD)*, we check the presence and the rank of **most important signal** (identified according to the ground truth).

$x = \text{Ranking}(\text{most important signal})$

$$\text{Ranking score}(e) = \begin{cases} 1 - 0.15(x-1) & \text{if } x < 4 \\ 0.5 & \text{otherwise} \end{cases}$$

$$\text{Average ranking score} = \frac{1}{|TD|} \sum_{e \in TD} \text{Ranking score}(e)$$

$$0.5 \leq \text{Average ranking score} \leq 1$$



3 – Predict the correct fault in the shortest time



Overview

For the 3rd part of the challenge, we define the *Average relative time* to score the identify the fault in the shortest possible time.

- We use **the Test dataset only**.
- To compute it, we use ranking produced by the **Log Performance function only**.

3 – Predict the correct fault in the shortest time

Evaluation Process

1. For each experiment e in *Test Dataset (TD)*, for each team i , we get *the time for prediction* ($T(e,i)$).
2. From $T(e,*)$ we select the subset $\hat{T}(e)$ as the times from the team i predicting the correct label for e
 - i.e., $\text{Label}(e,i) = \text{ground truth}(e)$
3. We get the *Shortest Time* ($ST(e)$) in $\hat{T}(e)$
 - If no team predicts the correct label, the shortest time is set to the experiment duration

$$ST(e) = \begin{cases} \min(\hat{T}(e)) & |\hat{T}(e)| > 0 \\ |e| & \text{otherwise} \end{cases}$$

3 – Predict the correct fault in the shortest time

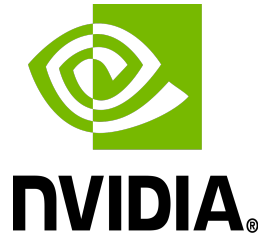
4. For each experiment e in *Test Dataset* (TD), for each team i , we get the time for *prediction* ($T(e,i)$), and we compute the relative time to classification

$$\text{Relative Time}(e,i) = \frac{ST(e)}{T(e,i)}$$

5. For each team we compute the *average relative time* to classification

$$\text{Average relative time} = \frac{1}{|TD|} \sum_{e \in TD} \text{Relative Time}(e)$$

$$0 < \text{Average relative time} \leq 1$$



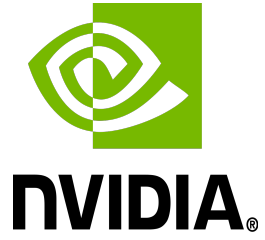
4 BONUS – Identification of system parameter configuration



Overview

For the 4th BONUS part of the challenge, we define the *Bonus score* to score the capability to split the experiments based on the system parameter correctly.

- We compute the *Bonus score* using **all no-fault experiments available**.
- To compute it, we use the ground truth system parameter labels and the labels produced by the **Log Performance only**.



4 BONUS – Identification of system parameter configuration



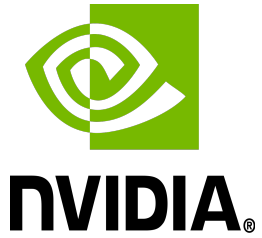
Evaluation Process

1. We compute the *adjusted rand index*
 - If the team did not submit the bonus point the adjusted rand index=0

$$\text{Bonus score} = \begin{cases} 1.3 & \text{if adjusted rand index} \geq 0.8 \\ 1.2 & \text{if } 0.7 \leq \text{adjusted rand index} < 0.8 \\ 1.1 & \text{if } 0.5 \leq \text{adjusted rand index} < 0.7 \\ 1.0 & \text{otherwise} \end{cases}$$

$$1 \leq \text{Bonus Score} \leq 1.3$$

Final Score



Data challenge Score



For each team, the data challenge score is computed as the combination of each task score.

$$\textit{Data challenge} = \textit{Accuracy} * \textit{Average ranking score} * \textit{Average relative time} * \textit{Bonus score}$$

$$0 \leq \textit{Data challenge} \leq 1.3$$