



Indian Institute of Technology Guwahati

KRITI '24

Automated Research Paper Categorization

(Mid Prep - 400 Points)

Organizer

IITG.ai



Automated Research Paper Categorization

400 Points

Start : 28/01

End : 07/02

Problem Statement

Contemporary paper submission platforms necessitate users to upload paper titles and abstracts, followed by the selection of appropriate categories for their submissions. However, the multitude of available categories poses a challenge for authors seeking optimal classification. Imagine a submission system that not only streamlines this process but also enhances user experience by offering intelligent category suggestions based on the paper's content.

In this competition, participants will be tasked with developing a solution that, given the paper title and abstract, accurately identifies the most suitable categories to which the paper belongs. This challenge aims to foster innovation in automated categorization, paving the way for more efficient and user-friendly paper submission platforms.

Link to Kaggle Competition: <https://www.kaggle.com/competitions/kriti-2024>



Invitation Link: <https://www.kaggle.com/t/853ddfaba5db45dbbedad4a7ea412a04> (Only to be used by registered team members)

Team Formation

1. A team can consist of a maximum of 6 members.
2. There should be at least 1 first year student.

Deliverables

1. Inference Code
 - a. Internet Access Disabled
 - b. Runtime ≤ 9 hours
 - c. No external data for training.
 - d. Submission file must be named submission.csv
2. Presentation
 - a. Tentatively scheduled on 10/02/2024
 - b. Max Limit: 7 minutes

Points Distribution

1. Final Leaderboard Ranking - 300
2. Presentation - 100



Rules

1. You'll be provided with a **Training** as well as **Testing** Data Set.
2. There will be 3 Leaderboards:
 - a. **Public Leaderboard** (50% of test.csv)
 - b. **Private Leaderboard** (50% of test.csv)
 - c. **Final Leaderboard** : This leaderboard will take into account scores from the private leaderboard as well as a **hidden test set**. The final score will be a weighted average of the normalized scores of both the leaderboards.

$$\text{Final Score} = (\text{norm}(\text{private}) + 2 * \text{norm}(\text{hidden})) / 3$$

where, $\text{norm}(x) = (x - \min(x)) / (\max(x) - \min(x))$

3. You are allowed to make only **5 submissions in a day**.
4. You are supposed to **choose one submission** before the end of competition.
5. All teams must keep their **team name** on Kaggle as it appears on the GC Leaderboard. All team members must join the team before making any submissions.
6. You're strictly advised to **not publish** your work on any public platform before the end of the competition. If found, 50% points will be deducted from final leaderboard ranking points.