

## Guidelines for KDD Cup 2019 Proposal format

Questions : [kddcup2019@kdd.org](mailto:kddcup2019@kdd.org)

SIGKDD-2019 will take place in Anchorage, Alaska, US from August 4<sup>th</sup>-8<sup>th</sup> 2019. The KDD Cup competition is anticipated to last for 2-4 months, and the winners will be notified by mid-July 2019. The winners will be honored at the KDD conference opening ceremony and will present their solutions at the KDD Cup workshop during the conference. The winners are expected to be monetarily rewarded, with the first prize being in the ballpark of ten to twenty thousand dollars.

We are looking for strong proposals in two distinct areas :

1. Proposals for Hosting Data for KDD Competition or
2. Proposals for Hosting Platform for KDD Competition

### 1. Proposals for Hosting Data & Problem for KDD Competition

Hosting data and organizing KDD Competition provides a unique opportunity to solve challenging problems utilizing crowd sourcing strategies. KDD has long pioneered the space of data science competitions and more than thousand entries are made every year. In order to select the best possible problem we would like the proposals to have following sections:

1. Problem Statement
2. Why is this an important problem? (Optional)
3. Auto-ml : Why this problem is relevant for auto-ml? (Optional)
4. Data Source
5. Data Description & Details (Size of data in GB - Train, Test )
6. Evaluation Metric \*
7. Submission file format
8. Time lines
9. Summary table \*

*Example Summary table \**

Summary table provides a snapshot of details of data :

Topic	Train- rows	Train-cols	Test-r ows	Test- cols	Metric(s )	#of phases (ph)	Auto-ml
Predicting example	10000	400	6000	399	MAE	1	No

Additional documents such as endorsement letters, relevant papers could be provided to improve the strength of the proposal.

## 2. Proposals for Hosting Platform for KDD Competition

As data science industry grows there are a lot of companies specializing in providing platforms for hosting the data science competition. We would like to encourage proposals in for the platform this year and look for the following sections for the proposal.

1. Platform Background & Overview
2. Key Capabilities - feature summary of the platform \*
3. Ease of hosting competition
  - a. Ability of platform to handle multi label problems.
  - b. Ability of platform to add a metric if required by the problem.
4. Conclusion - Why should we select the platform?

Key Capabilities - feature summary of the platform \*

No	Feature	Platform Host
1	Platform Name	XYZ
2	Ability to host regular predictive analytics competition with multiple metrics	Yes
3	Ability to host auto-ml competition with multiple phases on the platform with multiple metrics	No
4	Ability for competitors to discuss solution strategies using built in discussion forum	Yes
5	Ability to provide hosting ability to data sponsors or chair team to manage.	Yes
6	Ability for competition hosts to write blog sessions for content marketing	Yes
7	Interpretability on platform	No
8	Ability to view results & basic charts (ROC, etc) of the submission.	Yes

Note : Additional documents such as endorsement letters, relevant papers/documents could be provided to improve the strength of the proposal.