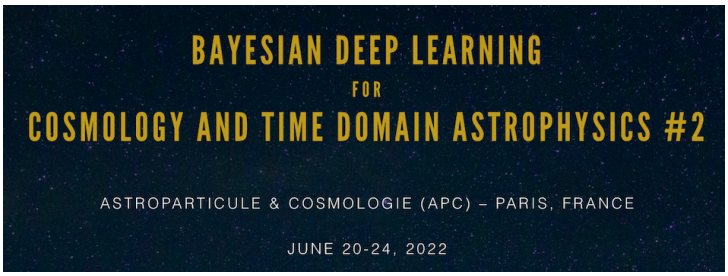

ID de Contribution: **58** Code de contribution: **ethics1** Type: **Talk**

Goodness-of-fit Evaluation: Toward Trustworthy Probabilistic Machine Learning

Likelihood-based and machine learning-based probabilistic classifiers are utilized by astronomers to classify astronomical objects and events. Simultaneously, these models are increasingly deployed to guide consequential decision-making tasks in our society (resource allocation, admission recommendation, etc). While the applications of these models have been rapidly increasing, we have not seen a corresponding leap in developing validation methodologies that can verify the trustworthiness of these models. Typically, the performance of a probabilistic classifier is assessed with metrics, such as accuracy, precision, and Area under the ROC curve, among others. Except in trivial cases, where a model can perfectly identify class members, neither of these performance metrics can evaluate the goodness-of-fit. To fill this gap, we have developed a novel kernel-based, distribution-free method that (1) evaluates the goodness-of-fit of a probabilistic classifier and (2) quantifies its distance from the oracle classifier. This tool, that is initially developed to evaluate the trustworthiness of classifiers in decision-support systems, has ample applications in astronomical settings. In this talk, I will introduce this novel tool, present theoretical guarantees, and its application to trustworthy ML. Then, I will illustrate the performance of the proposed test statistic by applying it to a set of controlled simulated data; and finally, demonstrate its application to astronomical data analysis.

Auteur principal: FARAHI, Arya (University of Texas at Austin)

Orateur: FARAHI, Arya (University of Texas at Austin)

Classification de Session: Ethics of large scale machine learning

Classification de thématique: Ethics of large scale ML