

# Knowledge Discovery Interest Group

Raffaele D'Abrusco

CENTER FOR

ASTROPHYSICS

HARVARD & SMITHSONIAN

## ⊕ Three interesting, different talks

- ⊙ R. Martinez-Galarza: *X-ray datasets: A Machine Learning Perspective*  
Application of ML methods to both tabulated properties of *Chandra* X-ray sources and data product to classify and select interesting/rare.
- ⊙ P. Skoda: *The Role of VO Technology in Astronomical Machine Learning*  
A comprehensive review of how the VO facilitate ML-based research and the areas where improvements is needed.
- ⊙ Y. Wu: *Classification of Galaxy Spectra based on Convolutional Neural Network*  
Application of CNN to the problem of the spectroscopic classification of galaxies.

## Future topics of discussion from KD-IG session

- ⊕ **What needs to be done to make existing and future science platforms ML-ready?** Accessing both tabular data and data products is a key requirement for most KD applications, especially in high-energy astrophysics.
- ⊕ **Should the standardization and normalization of astronomical data be standardized?** Is it worth pursuing for massive data access tasks? If so, what methods should be implemented? What level of documentation should be provided for reproducibility purpose?
- ⊕ **Training sets, training sets.... training sets.** The potential of supervised methods depends critically on availability and quality of labels for training sets.

# Getting involved

- ⊕ Chair: R. D'Abrusco (rdabrusc@cfa.harvard.edu)
- ⊕ Vice-chair: -
- ⊕ IG webpage: <https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaKDD>

## Staying in touch

- ⊕ E-mail: [kdd@ivoa.net](mailto:kdd@ivoa.net)
- ⊕ Slack: *IVOA#kdd*
- ⊕ In the inter-InterOps periods, we'll call running thematic meetings, involving presence of representative from other WGs/IGs as needed, on specific themes chosen by the IG.