# Applications of unsupervised machine learning techniques for data exploration and discovery in ISM science

## Dalya BARON

## (Stanford University, USA)

Tuesday, July 22, 2025, 16:15–18:00

In this lecture, we will explore several classes of algorithms that belong to the family of unsupervised machine learning. These algorithms are called *unsupervised* because they do not require "ground truth" labels or target variables for training. Instead, they operate directly on the data and are used for clustering, component separation, dimensionality reduction, data visualization, and outlier detection. These methods are particularly useful for exploring the complex and heterogeneous datasets common in ISM science and can facilitate new discoveries.

We will begin with a broad discussion of the motivation for using data science and machine learning techniques in the context of data exploration and discovery. From there, we will work together to understand how to apply these tools effectively to astronomical datasets. Since all of these methods rely on some notion of distance or similarity to relate astronomical objects, we will first look at different ways to represent data and consider the tradeoffs of each approach. We will then survey dimensionality reduction, clustering, and outlier detection techniques, and discuss how to interpret their outputs meaningfully. Finally, we will go through a set of guidelines for incorporating unsupervised machine learning into our own research in a safe and constructive way.

## Bibliography

- *Machine Learning in Astronomy: a practical overview*, Baron, D., 2019, Arxiv