

Dustin Burnham
1/22/2019
Data Science 400
Milestone Project 1: Data Flow Diagram

Task:

In astronomy, one common problem is there are not enough people to analyze the massive amount of data coming from modern telescopes. One project I have worked with through Galaxy Zoo, is to classify galaxies as either elliptical, spiral, or irregular. The website relies on “citizen scientists” to do this simple classification to help them train their galaxy classifier. Hopefully once enough data is analyzed, the classifier will be good enough to analyze images of galaxies and classify them without the need for human eyes. With the updated classifications, science can be done in multiple subfields of galactic astronomy.

Data Flow Summary:

The data itself is generated using a CCD in a telescope to count photons. These photon counts are then converted into an image. This image is then evaluated by a “citizen scientist” and classified. This classification is used to train a machine learning classifier, that when applied to new galaxy images will be able to pick out the galaxies and identify what type they are. This information is updated to galaxy catalogues where astronomers can query for data relevant to their research in galactic astronomy.

The data flow diagram for this process is on the following page.

