

# Supervised Learning for Star Cluster Classification

This project has the objective to demonstrate the application of supervised machine learning algorithms to predict “In\_cluster” and “Isolated” star classifications from an N-Body star cluster dataset.

## Installation

The project’s Python code file is “**Capstone\_Project\_N\_Body\_Classification\_Rev\_C.py**” Before executing the Python code, the following two (2) CSV datasets must be saved to same folder as the Python file:

**N\_Body\_Star\_Classification\_Capstone\_Dataset\_Rev\_A.csv**

**N\_Body\_Star\_Classification\_Pre\_Processing.csv**

## Usage

The Python code imports N-Body star cluster training/test datasets; accomplishes dataset pre-processing and dimensional analysis; benchmarks the classifier (Decision Tree) performance with the training/test dataset; trains and tests SVM and logistic regression classifiers to compare with benchmark performance.

The **N\_Body\_Star\_Classification\_Capstone\_Dataset\_Rev\_A.csv** file is the training/test dataset while the **N\_Body\_Star\_Classification\_Pre\_Processing.csv** file is used as the test dataset to validate the selected classifier choice.

If further data exploration is desired, the following CSV file names are the source initial and final time epoch N-Body star cluster datasets:

**c\_0000.csv**

**c\_1800.csv**

This file can be found at the following URL:

<https://www.kaggle.com/mariopasquato/star-cluster-simulations>