



# DATA 601 – AI Classification and Clustering Example Demo Discussion

Felix Gonzalez, P.E.  
Adjunct Instructor  
Division of Professional Studies  
Computer Science and Electrical Engineering  
University of Maryland Baltimore County

# Exercise: Image Classification (1/7)



- We want to develop a ML classification model (supervised ML) that classifies new images as either “Space object” or “Not a space object”
- Data: NASA space images and non-space images
- The data set has 22 labeled images.
  - Seems to be balanced (11 space, 11 non-space labels)
  - 18 for training (~80% of data)
  - 4 for testing (~20% of data)
- We will then classify 9 new images as either “Space Object” or “Not a space object” given the boundaries of the images in the training data.
  - Think about confidence when we are classifying new images

# Exercise: Image Classification (2/7)



## Space Object

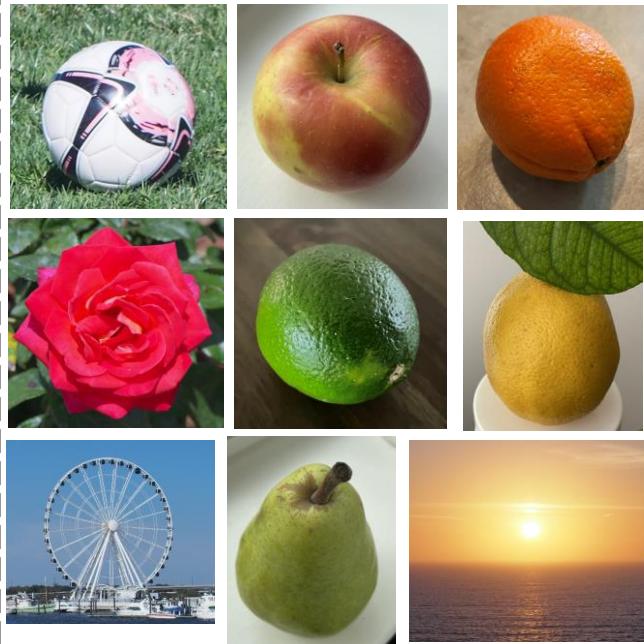
Pictures from: <https://images.nasa.gov/>; [www.nasa.gov/](http://www.nasa.gov/);  
<https://flickr.com/photos/nasawebbtelescope/>; [www.jpl.nasa.gov](http://www.jpl.nasa.gov)



## Training Data

## Not a Space Object

Pictures by: Felix Gonzalez



# Exercise: Image Classification (3/7)



Testing Data

Image A



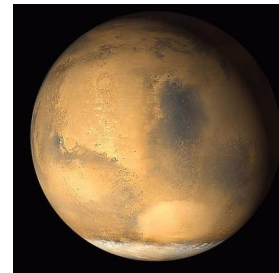
Image B



Image C



Image D



# Exercise: Image Classification (4/7)



Model Deployment: New Data

Image E



Image F



Image G



# Exercise: Image Classification (5/7)



Model Deployment: New Data

Image H



Image I



Image J





# Exercise: Image Classification (6/7)



Model Deployment: New Data

Image K



Image L



Image M



# Exercise: Image Classification (7/7)



- Is the data used for training/testing enough?
  - Some images (testing and new) were too different to the training data
  - Underrepresented images like vehicles which indicate a data balance issue.
- Biases in the data?
  - Did we evaluate the training data before deployment?
  - Cognitive bias? Confirmation bias? Other biases?
  - Would the model be biased on some non-space objects with dark background?
- Was the training data appropriate for the images we were classifying? In all cases or in some cases?
- What can we do to address the shortcomings?
  - Increase number of labeled data: Non-space with dark background, and vehicles
  - Create a new class with corresponding data like space/non-space vehicles.
  - Should we increase the number of space objects with different background colors?
  - Use a different model
  - Use unsupervised ML algorithm if possible

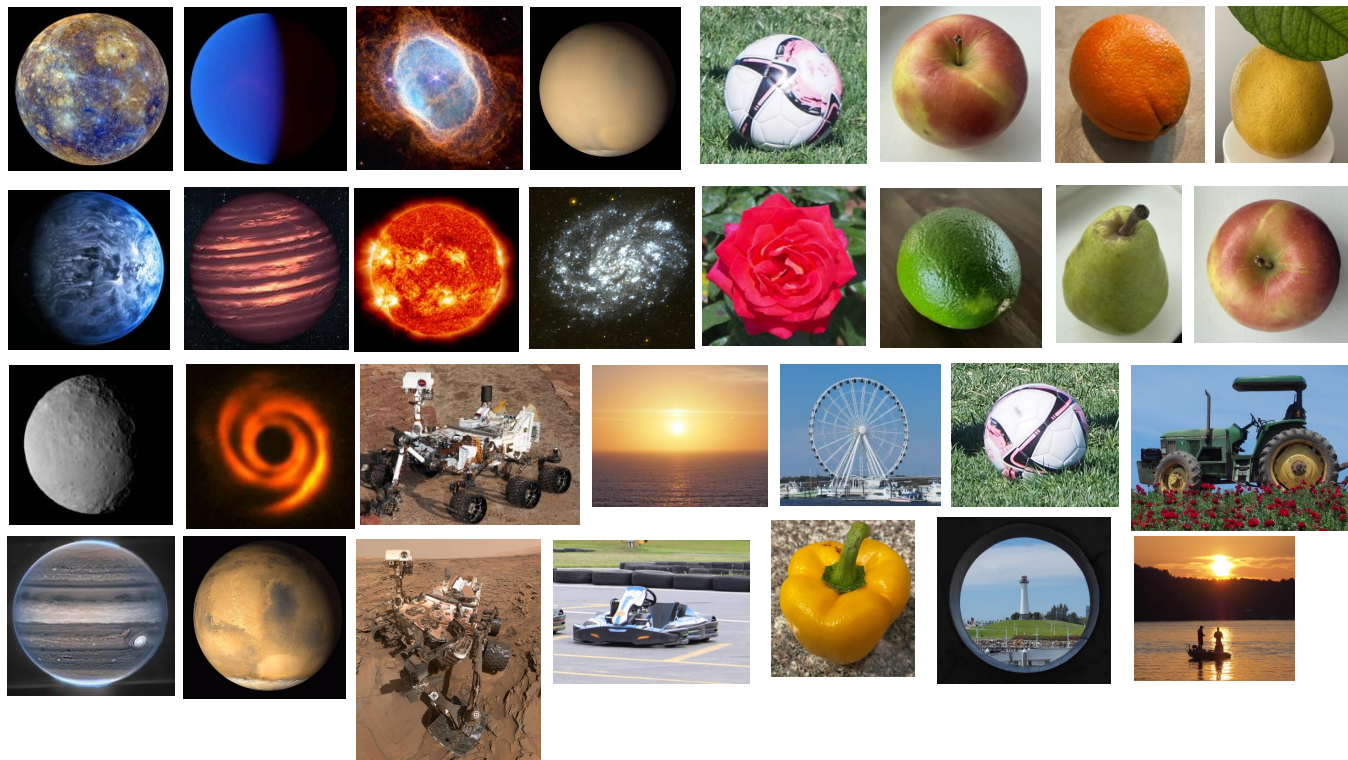


# Exercise: Image Clustering (1/4)



- We want to develop a ML clustering model (unsupervised ML) that creates clusters (e.g., groups) for all-previous images (30 images)
- What would be features that we could compare to create significant clusters?
- Features can be any we can think of that we can collect data, such as color, shape, type, category, weight, etc.
- Note that we could also have clusters for outliers.
- In this case let's consider a significant cluster as one that has at least 4 images.
- An image can be part of one cluster.

# Exercise: Image Clustering (2/4)



# Exercise: Image Clustering (3/4)



- Potential solution 1:
  - Cluster 1: Fruits/Vegetables
  - Cluster 2: Planets
  - Cluster 3: Vehicles
  - Cluster 4: Everything else (e.g., non-planets, sunsets, flower, ball)
- Potential solution 2:
  - Cluster 1: Red/Orange dominant color
  - Cluster 2: Blue dominant color
  - Cluster 3: Sports related
  - Cluster 4: Everything else

# Exercise: Image Clustering (4/4)



- Clustering algorithms work in a similar way.
- They find patterns in the data and depending on the algorithm and parameters specified will create different clusters.
- Different algorithms may group the data very differently depending on the shape
- Example Clustering Algorithms
  - Kmeans: Clusters the data given a specified number of clusters (i.e,  $k$ ). There are methods to calculate the optimal number of clusters can be calculated.
  - DBSCAN: Clusters the data given a specified min samples in a cluster and a specified density (i.e., epsilon)



Questions?