Categorization Challenge

Members:

Christopher Roman.

Luis Enrique Navarro.

Uriel Benitez.

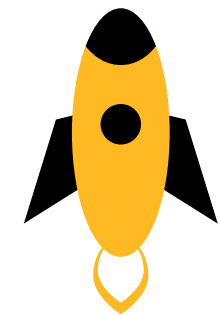# AGENDA

Baker Hughes

Challenge Description
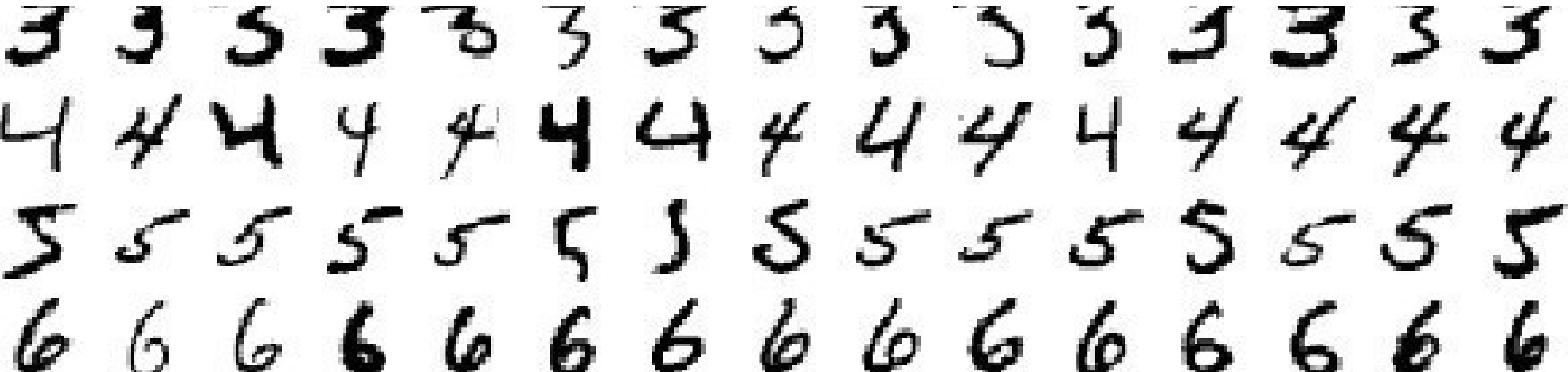
Data processing.

Feature selection.

Model selection.

Build a **classifier** or **categorization model** for handwritten digits in the MNIST database.

DATA PROCESSING

Baker Hughes™

We have ?

60,000 color images.
10,000 Gray scale images.
1 csv file with image labels.

What we did ?

Step 1: **Look for NAN values** in the csv label file.

step 2: **Normalize** both sets of images in a **single color scale.**

step 3: **vectorize the matrix** to imput the data in the model.
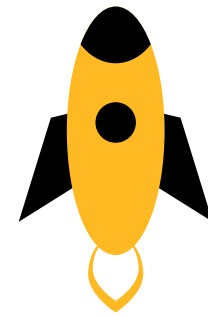
What we tried ?

Work with binary pictures.

Reduce the number of pixels in a image.

Rotate a specific set of images to uniform the data.

Feature selection.

Model selection.