HOPHACKS

DATA SCIENCE CHALLENGE:
DISCOVERING NEW CATEGORIES FROM IMAGES

INTRODUCTION

Modern society is awash with data. It is estimated that 136,000 photos are uploaded to Facebook every minute. The sheer volume of the available data is complicated by the fact that these images usually do not have labels associated with them. For example, the photos a user uploads to Google Photos often fall into categories such as pictures of friends & family, food, and travel. However, it is difficult to automatically discover these categories and organize the gallery accordingly. Fortunately, recent advances in data science and machine learning can help us achieve this goal of finding new categories in unlabeled data.

This competition, hosted in partnership with MINDS, challenges you to use unsupervised learning to discover new categories within a large set of images. To simplify the problem, the datasets we provide to you are preprocessed so that each image contains a single object. We will also provide common features, precomputed on the dataset. You will apply ***clustering*** algorithms which generate a label for each image. Images with the same label are considered to contain objects of the same category. Before and during the competition, there will be short data science workshops available to help you get started.

PRIZES

Prizes will be given to the teams who attain the top three highest clustering accuracies on the test set. The prizes are as follows:

First place - $1024
Second place - $512
Third place - $256

ABOUT MINDS

Founded in 2017, the mission of the Johns Hopkins Mathematical Institute for Data Science is to establish the mathematical principles behind the analysis and interpretation of massive amounts of complex data. In a message from René Vidal, the director of MINDS: “[MINDS] brings together a multidisciplinary team of mathematicians, statisticians, computer scientists, and engineers to develop the fundamental mathematical, statistical, and computational principles for the analysis and interpretation of massive amounts of complex high-dimensional data. The institute will also help develop multidisciplinary educational programs on the foundations of data science, and foster interactions among data scientists through a series of collaborative research workshops and summer schools.”

HAPPY HACKING!
THE HOPHACKS TEAM