**Datasets Available:**

[CIFAR-10](https://www.cs.toronto.edu/~kriz/cifar.html)

Benchmark set of 60000 32x32 images across 10 categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck. Downloadable in forms suitable for python, Matlab, and C

[CIFAR-100](https://www.cs.toronto.edu/~kriz/cifar.html)

Extended version of CIFAR-10, but with 100 classes with 600 images each. Each image is labeled by a “fine” class and a “coarse” superclass (ex. Apple class and fruit and vegetables superclass)

[dotabuff](https://www.dotabuff.com/)

Data for Dota 2 players. Dotabuff doesn’t have its own API so it might be slightly complicated to get the data into usable formats, but there are relevant forum posts on the topic.

[Reddit posts](https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/?utm_source=reddit)

Data on 1.7 billion Reddit comments (includes subreddit ID, upvotes, post contents)

[FIF19 Player Stats](https://www.kaggle.com/karangadiya/fifa19)

Contains player statistics on people in FIFA19

[MUCT Face Database](http://www.milbo.org/muct/)

Collection of 3755 faces with 76 manual landmarks each1

Google images?

Chrome extension to download several images at a time. Can sort through by hand to some extent.

**Ideas:**

Fruit ripeness:

Classify images of fruit as ripe, overripe, or underripe

Train on a variety of different fruits. Classifier recognizes the fruit first, then its ripeness. Users could maybe correct fruit classification to help ripeness classification?

Start small with a single fruit with lots of data available. Apples are available in the CIFAR-100 dataset.

Predicting FIFA19 player market value:

Use FIFA19 player stats dataset to predict a player’s market value (or wages)

There might not be much depth, as higher player stats are likely correlated with higher wages (not surprising)