Capstone Three: Project Ideas

I deliberately chose project ideas and datasets that related to business problems. I excluded other feasible project ideas that did not relate to business problems.

Idea One:

Use convolutional neural nets (CNN) with tensorflow to classify wheat as having blast (a disease) or not, or possibly categorize degree of blast severity.

This relates to a relevant business problem. Many farmers around the world struggle with the issue of having their wheat damaged by blast disease. Sometimes this disease may be hard to diagnose or detect with an untrained eye. If there were an app that could be linked up with a phone camera (or other camera) that could easily diagnose whether wheat has blast (and maybe the severity) or not, this could save farmers money.

Downsides: It appears there's already an entire scientific paper written (on the dataset's website) that categorizes the wheat blast using CNNs. Would I be able to almost repeat what they've done if I modify the methods slightly, or is this a no go?

Another downside is that the dataset zip file is 3 gigabytes. This quantity of data may be hard to work with on a laptop unless I use cloud resources or only use a portion of the dataset.

Wheat blast dataset:
https://purr.purdue.edu/publications/3772/1

Idea Two:

Use CNNs with tensorflow to try to predict whether a restaurant is expensive or not based on a photo. This would have business applications for yelp, because building this automatic prediction functionality into their site/app could help categorize restaurants that haven't yet been labeled as expensive or not by users. Accurately categorizing more restaurants as expensive or not would help yelp's users and thus improve yelp's business.

Downsides: Messy, complicated dataset?

Yelp restaurant classification dataset:
https://www.kaggle.com/c/yelp-restaurant-photo-classification/overview

Idea Three:

Use CNNs with tensorflow to accurately classify lesions as their correct disease. This could have business applications because a mobile app could be created that allows users to take a photo of their lesion and then the app classifies it as a disease. Such an app could potentially be profitable if it was accurate and FDA approved.

Downside: The zip file is 3 gigabytes-large dataset.

Skin cancer dataset:
https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000