Springboard Capstone 2 Ideas

Tzu-Ying Chen

1. Captcha letter identification:

Using machine learning model to identify numbers and characters from captcha images to help a friend to build a open source mobile phone application to book railroad tickets in Taiwan – Data source: friends provided a good amount of captcha images for training

* Could use neutral network and deep learning method to identify the letters. Potentially using some unsupervised algorithm to clusters similar images of the characters and label the group by telling what the character is

1. Spam Email detection

Identify content of emails to identify whether an email is spam no not. The data set is provided by Kaggle: <https://www.kaggle.com/venky73/spam-mails-dataset>

* Could use classification algorithm to identify which terms appears frequently in the email and hence determine whether it is a spam or not

1. Breast Cancer Wisconsin

Identify whether the patient has breast cancer or not by different symptoms and statistics. The data set is provided by uci machine learning repository: <http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

* Could use classification algorithm to identify whether a patient has breast cancer by different symptoms the patient has.