**Capstone Ideas by Kausik Chattapadhyay**

**Idea 1**: **Artist Identification with Convolutional Neural Networks.**

Artist identification of fine art paintings is a challenging problem primarily handled by art historians with extensive training and expertise. Using Convolutional Neural Networks (CNNs) with the goal of identifying the artist of a painting as accurately and precisely as possible. My dataset consists of 300 paintings per artist from 57 well-known artists.

**Dataset:** A large dataset of art compiled by Kaggle that is based on the WikiArt dataset [13]. This dataset contains roughly 100,000 paintings by 2,300 artists spanning a variety of time periods and styles.

**Idea2:** IMPROVING YELP RESTAURANT RECOMMENDATIONS

Current reviews on Yelp are predicated on an overall score for a given establishment with a very rough "star" system and a textual review. Star ratings do not incorporate the abundance of valuable information in the textual review itself. Our objectives are to evaluate whether topics generated by LDA conditioned on star ratings can be used to automatically generate user preference profiles and restaurant characteristic profiles that could be useful in matching customers to restaurants, as well as to build a K-means model which may accurately predict a given users rating for establishments.

**Dataset:** Yelp Dataset Challenge

**Idea3: "Fake news" classification model**

There are a lot of "fake news" targeting a group of community and spreading rumors. People has the right to know nothing but the truth. Since the U.S. election, a lot of political experts claimed that the election was influenced by a significant amount by the spread of biased and/or untrue news. These fake news posts exploit Facebook users’ feeds to propagate information throughout the internet. These news deliberately deliver false information to the public in order to propagate information. There was a Kaggle competition called the "Fake news Challenge" and "Facebook is employing AI" to filter out the fake news stories from the users feeds. The challenge is to build a text classification project that can differentiate between "Real News and Fake News" .

Here is the article that you can look at if you are interested to know about it.

<https://opendatascience.com/blog/how-to-build-a-fake-news-classification-model/>

The Kaggle challenge could also be found over here in this link:

<http://www.fakenewschallenge.org/>

**Dataset:** <https://www.kaggle.com/mrisdal/fake-news>

I think classifying "Fake News" would be very much beneficial to Facebook and other social Medias as they would want to distinguish from real news vs fake news and it would contribute a lot to their trending suggestions.

**Idea4:**  **Airbnb New User Bookings**

This is a classic Kaggle competition that Airbnb hosted a year back in order to recruit Data Scientists in their company. This got me interested since it is already addressing a issue faced by a company and this is real world domain problems. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast [demand. In](http://demand.in/)this recruiting competition, Airbnb challenges you to predict in which country a new user will make his or her first booking.

The challenge can be found in this link : <https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings>

Data Sets: <https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings/data>