Carnegie Mellon Data Science Club (DSC)

Mentorship Groups

Raaka

- Grace
- Brand
- Veronica
- Mia
- Saahil

Idris + Anisha

- Ayush
- Yeonji
- Mihir
- Amy
- Ayden
- Jenny

Al + DK

- Claire
- Jei
- Daniel
- Owen
- Shiyang

Introductions!

MENTORS

- Year, Major
- Favorite Applications of Data Science
- Favorite CMU Class
- Other CMU club involvements
- Hobbies
- Favorite Album/Artist

MENTEES

- Year, Major
- Current Classes
- Why DSC?
- Other CMU club involvements
- Hobbies
- Favorite Album/Artist

Within Groups: what's a data science project you have worked on/want to work on?

Mentorship Kaggle Competition

- online community platform for data scientists and machine learning enthusiasts
- mini-courses on a variety of data science topics
- find datasets
- use GPU integrated notebooks
- compete with other data scientists to solve challenges
- using Kaggle to have each mentorship group compete against each other in a data science competition over the course of the semester!

Kaggle Competition Option #1

Getting Started Prediction Competition

House Prices – Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting

Kaggle · 3,897 teams · Ongoing

- Field: Machine Learning (ML)
- In this competition, you are challenged to predict the sales price of a home using 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa.

Kaggle Competition Option #2

- Field: Computer Vision (CV)
- In this competition, your goal is to correctly identify digits from the MNIST dataset, a famous dataset that contains of tens of thousands of handwritten images.
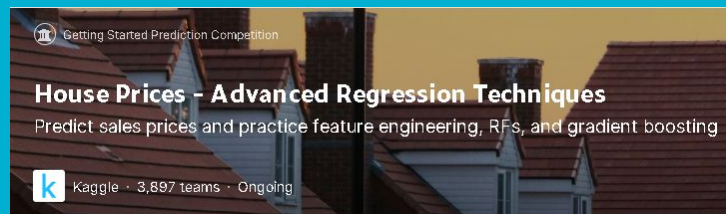
Kaggle Competition Option #3

Getting Started Prediction Competition

Natural Language Processing with Disaster Tweets
Predict which Tweets are about real disasters and which ones are not

Kaggle · 861 teams · Ongoing

- Field: Natural Language Processing (NLP)
- In this competition, you're challenged to build a machine learning model that predicts which Tweets are about real disasters and which one's aren't. You'll have access to a dataset of 10,000 tweets that were hand classified.

Voting on Kaggle Competition

- **1: House Prices (ML)**

- **2: Digit Recognizer (CV)**

- **3: Disaster Tweets (NLP)**

Getting Started Prediction Competition

House Prices - Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting

Kaggle · 3,897 teams · Ongoing

Getting Started Prediction Competition

Digit Recognizer
Learn computer vision fundamentals with the famous MNIST data

Kaggle · 1,194 teams · Ongoing

Getting Started Prediction Competition

Natural Language Processing with Disaster Tweets
Predict which Tweets are about real disasters and which ones are not

Kaggle · 861 teams · Ongoing

Within Groups: What skills do you want to build this semester?