

---

---

# Capstone Ideas

— Caleb Stephenson —

---

---

# Idea #1: Book/Movie/TV NLP recommender

- Make a recommender to take a paragraph of text (“I want to read something with x, y, z”) and return recommendations, using actual human questions/answers as a database.
- Data source: Reddit
  - /r/suggestmeabook
  - /r/moviesuggestions
  - Etc
- Audience:
  - Anyone who wants a new book/movie/tv show to enjoy
  - Content providers who want to connect subscribers to content
- Metrics:
  - F1 score (true negatives aren't important here)

## Idea #2: Exoplanet transit identification

- Train a model to find possible exoplanets from “light curves” of observed stars. Identifying likely false positives and true positives.
- Data Source: Kepler mission data available to the public through NASA.
  - lightkurve API
- Audience: Astronomers wanting to identify targets for further observation (“exoplanet candidates”), to try to confirm the existence of exoplanets.
- Metrics:
  - Accuracy. We want to identify the most likely True Positives and False Negatives before committing telescope resources to observations.

## Idea #3: Restaurant closings and Reviews

- Try to predict the probability of restaurant closing within the next 1/3/6 months using online reviews.
- Data: Yelp, Google, Etc.
- Audience: Small business owners who want to use data to identify problems with how their business is being perceived by their customers.
- Metrics:
  - Time series metrics?