

## Getting Started

---

DS 6011: Capstone Part I / Capstone Prep  
School of Data Science  
University of Virginia

Last updated: September 13, 2024

# How to Get Started

Tips for getting started:

- Complete onboarding (e.g., training)
- Take inventory (tools, code, data, environment)
  - > What have they tried?
  - > What works well / does not work well?
- Understand the big picture.
  - > What is the goal?
  - > How will your work be used?

# How to Get Started – Breaking things down

- Look for intermediate milestones:
  - > Forming training set / validation set
  - > Decide on metrics
  - > Coding data pipelines

# How to Get Started – Breaking things down

- Thinking about functionality

- > What needs to happen? (build lists, embed objects, build classifier, ...)

- > Are there modules for the functionality?

May not need to build things from scratch

Can check for built-ins, papers w code, Hugging Face, ...

What do you need to code / build?

Break out large tasks and add to roadmap

# How to Get Started – Breaking things down

- Look for techniques (NLP, time series, ...)

The type of data will help guide

Helpful approaches:

> Literature review

Has this problem / similar problem been solved? How?