Getting Started

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

Last updated: September 7, 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?

- Look for intermediate milestones:

  > Forming training set / valuation set

  > Decide on metrics

  > Coding data pipelines

- 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

- 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?