Machine Learning Engineering Bootcamp Capstone

Step 1: Start Planning for Your Capstone

Summary

Time Estimate: 2 Hours

In this step, you will submit three capstone project ideas and start planning your capstone path based on your chosen area of focus.

Brainstorm Your Project Ideas

For your initial project ideas, you'll want to:
- Include a short blurb for each of your ideas.
  - The blurb should, at a high level, describe the problem and the data you'll be using to solve it. At this point, there's no need to talk about specific methods and techniques.
- Post your idea (title and blurb) on the community and solicit feedback from both mentors and other students. Pick one idea to work on based on the feedback you get. Discuss the idea with your mentor to ensure they're on board.

Please note: The goal of a project is NOT to do something novel — it's to demonstrate your competence as a machine learning engineer. It's perfectly acceptable to work on a dataset that's been worked on before and even answer a question that's been answered before, as long as the work is your own.

Examples For Inspiration

This is one of the project ideas by one of our alumni, Siri Surab.

Here are a few more ideas to spark inspiration for your capstone project. Many of these ideas come from natural language processing or computer vision since they're two of the hottest fields in AI. Your project doesn't need to be in one of these specializations.

- **Inventory tracking and compliance using object recognition:** A company wants to automatically track inventory in its warehouses using a camera with an object recognition algorithm. A similar model could be used in a home application, such as a smart fridge that recognizes what's placed in it.
- **Language translation:** Also called machine translation, this technique uses AI to translate one human language to another, in text or speech. You can work between the two formats like speech-to-text transcription or text-to-speech generation.
- **QA systems and chatbots:** Increasingly, companies are using automated chatbots to address their customer service workloads. These bots produce human-like responses to questions and are getting better every day.
- **Text summarization:** Imagine an application that can digest the daily news and produce a coherent summary tailored to a user. You can apply summarization to different domains, such as an application that can automatically produce a

personalized summary for a student who's trying to research a large amount of material.

- **Fraud/spam detection:** Detect "bad" transactions or items in a dataset. This could take the form of detecting fraudulent credit card transactions, fake news on social media, spam in email, doctored images or video, or abusive behavior on Twitter. Depending on the problem, you can use a variety of techniques, ranging from "traditional" machine learning to the latest in deep learning.

Former Capstone Project

This is a project by one of our alumni, Siri Surab. She used the Quora Duplicate Questions dataset from Kaggle, and applied NLP techniques from both "old-school" Machine Learning as well as Deep Learning to identify duplicate questions on Quora.

- [Ideation](#)
- [Blog Post](#)
- [GitHub repository](#)

We don't expect you to understand all of these techniques at the beginning of the course, but we've presented this here as an example of what your final project will look like. You'll go much deeper into this specific project in a later unit on NLP.

Plan For Your Project Based On Your Focus Area

In total you will spend ~100 hours on your project, a bit more if you choose to complete an Advanced level Capstone and include an advanced topic. In the last resource, you just learned about four areas of focus. Be aware that each of these focuses would influence how you spend time on each of the milestone steps of your capstone project.

1.) Think about which area you want to focus on based on your strengths and what you want to showcase for your future employers.
2.) Here is a potential plan for you. It contains a breakdown of suggested time allocation for each main part based on your focus and some areas where you need to budget extra time. This is only a reference, please discuss concrete details with your mentors. Don't feel that you need to commit to this plan. Your plan may likely change as you proceed through the course.

	Data Collection & Processing	Models Prototype & Scaling	Deployment/ Engineering Architecture
	Step 1~5	Step 6~9	Step 10~12
Data -focused	50%~60% You may spend extra time on data collection, cleaning, and processing. And if you need to process real-time data, you will need to build ETL infrastructure.	20~30%	20~30%
Model -focused	25%	50% You will try different models and make ensemble models to enhance performance. If you are developing DL models, please budget extra time.	25% If you are short on time, feel free to leverage existing deployment methods like Algorithmia.
Architecture -focused	30~40% You can shine by building automation in the data pipeline to support analysis at scale.	20~30%	40%% You can build automation to train/test data or design your own engineering architecture. We suggest this focus on students with a strong SWE engineer background.
Product -focused	30%	20%	50% You might spend extra time developing the app UI/UX and streaming data.

Project Submission Steps

1. Write a description of your three capstone project ideas.
Your ideas can be broad and high-level. The descriptions should address the problem and identify the data you'll use to solve it. You do not need to talk about specific methods and techniques.
- Write at least 3 to 4 sentences explaining your idea and identifying the data you'd use to solve it.

2. Submit a Google Doc link to the submission box.
- Remember to enable sharing permissions to "comment."
- Please do not submit .pdfs, .ppts, or markdowns.

3. Review your ideas and your tentative decision for your focus area with your mentor during your next call.

4. Post your idea (including a title and description) to your student community to receive peer feedback.