

AI Technician Capstone

Cohort 3

April 10, 2023

Contents

0	Overview	2
1	Exploratory Data Analysis	4
2	Machine Learning Prediction	5
3	Web Application Development	6
4	Final Presentation	8

Phase 0

Overview

Phases

The AI Technician capstone will be conducted in 4 phases over 18 weeks. Each phase will begin with a 1-hour kickoff meeting that will orient you to the phase and discuss important considerations for the phase. The phases include:

- Phase 1: Exploratory Data Analysis (2 weeks)
- Phase 2: Machine Learning Modeling (6 weeks)
- Phase 3: Web Application Development (8 weeks)
- Phase 4: Final presentation (2 weeks)

Capstone Deliverables

The deliverables from this capstone will be:

1. 1x Design document
2. 1x Hosted web application
3. 2-3x AI models (1 per team member)
4. 1x Presentation

Workload

You will have at least one full day per week to work on your capstone. Each week has an associated deliverable that must be sent to your AI Professional mentor by **Friday at 1200**.

Mentorship and Progress Checks

Each capstone team will be assigned an AI Professional mentor. The mentor will formally meet with the team at least once every 3 weeks to review progress and provide guidance. Additionally, the deliverables specified for each week will be delivered to the team mentor for weekly review.

Seeking Help

Your assigned AI Professional mentor will be available to provide help on an as-needed basis. When meeting with the mentor, for each question you have, bring a written document that answers the following questions:

1. What is the issue?
2. Why is it not working?
3. What could be the problem?

4. What have you tried to solve the problem?

The purpose of answering these questions is to facilitate troubleshooting on your own and to help clearly identify the problem you are encountering.

Individual and Group Tasks

In each phase, there are individual and group tasks. The individual tasks are denoted by [Ind], and the group tasks are denoted by [Grp].

Teams

Capstone teams will consist of 2-3 AI Technicians and are based on current portfolio assignments.

Phase 1

Exploratory Data Analysis

Learning Objectives

LO1. Get familiar with a given dataset

LO2. Produce useful visualizations

LO3. Practice cleaning and formatting data for a data science task

Scenario

You are a data scientist at a small bank that is trying to get more customers to subscribe to term deposits. In an effort to increase the number of subscriptions, the bank has been running a telemarketing campaign. You have been asked to analyze the data collected so far in the campaign and use it to inform which customers the bank should call next week to maximize the number of subscriptions.

Introduction and Deliverables

In this phase, you will analyze the provided Bank Marketing dataset. The data was collected from direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. The classification goal is to predict if a client will subscribe to a term deposit (this is denoted by a “yes” or a “no” in the last column of the dataset). The bank wants clients to subscribe to their term deposit.

Week	Task	Deliverable
1	Download and clean the dataset, understand the features, produce visualizations of the data, identify key features, construct new features	Jupyter notebook
2	Write-up of project overview and Phase I and review with AI Professional mentor	Design document

Table 1.1: Deliverables in Phase I. Progress review weeks are highlighted.

Phase 2

Machine Learning Prediction

Learning Objectives

- LO1. Become familiar with machine learning capabilities and tasks
- LO2. Tune hyperparameters to improve model performance
- LO3. Perform model selection under constraints for an applied problem

Scenario

After analyzing the telemarketing campaign data, you have been asked to use the data to help the bank prioritize the customers to call next week. The `train.csv` data is the historical data so far in the campaign, and `test_X.csv` are data on potential customers to call. You will build a machine learning tool to decide in which order the bank should call those potential customers to maximize the number of term subscriptions gathered by the end of the week. The bank has only allocated 120 call-hours for telemarketing next week.

Introduction and Deliverables

Now that you have prepared the data, you are going to use it to provide predictive analytics for the bank telemarketing campaign. There are many questions we could try to answer with the dataset, but we are going to focus on predicting whether or not someone is going to subscribe to a term deposit. Specifically, the bank has a list of approximately 8,000 clients and wants to know in what order they should be called given that there are a limited number of labor-hours available for the telemarketing campaign.

This phase is scheduled to take 6 weeks with tasks and deliverables as shown in Table 2.1.

There are three main deliverables for this portion of the capstone.

1. An organized Jupyter notebook with all specified tasks. This notebook should contain all the steps needed to train your models and provide inferences on the test set for submission.
2. Inferences on the test set. The inferences should be a single column in the same order as the test data set. You will submit your results to your group mentor in CSV format.
3. A well-structured write-up of the phase. This document will include the questions you answer throughout the Phase as well as intermediate results such as plots.

Week	Task	Deliverable
3	Build model 0	Jupyter notebook
4	Fine-tune hyperparameters of model 0 and submit inference results	Inference results
5	Review with AI Professional mentor	Document
6	Build candidate models with hyperparameter tuning	Jupyter notebook
7	Model selection and final inference results	Inference Results
8	Write-up of Phase II and review with AI Professional mentor	Document

Table 2.1: Deliverables in Phase II. Progress review weeks are highlighted.

Phase 3

Web Application Development

Scenario

Introduction and Deliverables

In this phase, you will create a Django web app. The web app will not have much functionality at first, but you will add functionality to it throughout the rest of the capstone.

Week	Task	Deliverable
9	Create new Django web application, add model to Django web application, and perform inference on unseen data	Django App
10	Deploy on DSE and review with AI Professional mentor	Endpoint
11	Leave/flex week	Document
12	Leave/flex week	Docker container
13	Create Postgres database, record inference data in database, and generate summary statistics	Endpoint
14	Add querying capability and interactively display query results on Django web application	Endpoint
15	Implement a front-end design and review with AI Professional mentor	Endpoint
16	Implement a front-end design and write-up of Phase III	Design document

Table 3.1: Deliverables in Phase III. Progress review weeks are highlighted.

Week 9

Learning Objectives

- LO1. Practice creating a simple web application
- LO2. Add model inference functionality to a web application
- LO3. Understand how to read package documentation

Week 10

Learning Objectives

- LO1. Understand how, why, and when to Dockerize web applications
- LO2. Become familiar with the Data Science Environment
- LO3. Understand how to monitor an application in production

Weeks 11-12

Use these weeks to get caught up, take leave, or get ahead if you are taking leave later in the capstone.

Weeks 13-14

Learning Objectives

LO1. Configure a web application with a database

LO2. Store and use data in a database

LO3. Understand how databases can be used to improve the functionality of web applications

Weeks 15-16

Learning Objectives

LO1. Improve the look and feel of your web application

LO2. Present an analytic in a clear way

LO3. Practice adding functionality to deployed web applications

Phase 4

Final Presentation

Learning Objectives

LO1. Practice synthesizing your work into a coherent presentation

LO2. Practice presenting your work

LO3. Practice learning from other teams' presentations

Scenario

Introduction and Deliverables

Week	Task	Deliverable
17	Prepare presentation and review with AI Professional mentor	Presentation
18	Final presentation	Presentation

Table 4.1: Deliverables in Phase IV. Progress review weeks are highlighted.

Tasks

1. [Grp] Create a presentation summarizing your work during the capstone. Your presentation should cover the following:
 - (a) Introduction
 - i. What did you do during the capstone?
 - (b) Motivation
 - i. What makes the project or your analysis interesting?
 - (c) EDA Outcomes
 - i. What techniques did you use during EDA?
 - ii. What visualizations did you produce during EDA?
 - iii. How did EDA inform your understanding of the dataset?
 - iv. What did you learn during the EDA phase?
 - v. Show figures and outcomes from your Jupyter notebook.
 - (d) Modeling Outcomes
 - i. What models did you examine?
 - ii. How did each of your models perform?
 - iii. How did you measure how your models performed?
 - iv. How did you select which model to use?
 - v. What did you learn during the modeling phase?
 - (e) Web Application Development
 - i. What analytics did you develop?

- ii. How did you design your web application?
- iii. Demonstrate the functionality of your web application
 - A. Demonstrate performing live inference with your web app
 - B. Demonstrate the functionality of your database
 - C. Demonstrate the analytics you developed
- iv. What did you learn during the web application development phase?
- (f) Reflection
 - i. What did you learn during the capstone?
 - ii. What skills were new to you? What skills were review?
- 2. [Grp] Schedule a rehearsal with your AI Professional mentor
- 3. [Grp] Give your presentation to an audience including AI2C leadership and CMU faculty