MSDS 626 Final Project

Robert Clements

MLOps Proof-of-Concept Final Project

A big part of working in the data science field is dealing with the pace of new technology. You'll quickly see in this course that, for each stage of the MLOps pipeline, there are many available tools to choose from, from open source python libraries, to fully-supported paid platforms. Realistically, there are no adequate one-size-fits-all platforms that can do everything, and you will most likely be forced to use different tools for different purposes.

At some point in your career you may be asked to do a proof-of-concept (POC) to determine if the latest cool new application, library, or programming language should be adopted by your team or organization. These POCs can go from simple (a few days of work testing out a piece of software) to complex (more than a year of work testing new infrastructure). For example, you might be asked to do a POC on:

- checking if a particular python unit testing library will improve the data science team's productivity;

- checking the viability of switching from AWS to GCP or using a multi-cloud strategy, in terms of cost, security, future-proofing, etc.;

- testing if a different orchestration engine will work in your existing infrastructure;

- deciding if a python library for generating Shapley values is adequate for explainability;

- testing if an expensive feature platform is more cost-effective than rolling your own;

- the list goes on and on.

For this project you will pretend to be part of a brand new MLOps team at a company that has been tasked with building the ML infrastructure from the ground up. As part of a group of 3-4 students, you will need to design the architecture, and select the tools, and then give justification to convince your leadership team (which is me, I guess) to adopt the infrastructure you've selected.

Some questions to consider when selecting each tool

- Should you use an expensive paid platform and risk vendor lock-in?

- Should you go with a fully open-source (free) infrastructure, and risk lack of stability and/or support?

- Should you do some combination of the two?

- Can the company afford to hire a team big enough to support a complex infrastructure?

- Does each tool play nice with each other?

Important Note: you absolutely must consider **at least two** different options for each stage of the pipeline (described below in more detail). For example, when considering what tools to use for Experiment Tracking, you should be comparing at least two options, one of which should be MLFlow (because we will be learning MLFlow in the labs), and the other option can be one of Weights & Biases, Metaflow, DVC, or any other tools you may have heard of or used before. You are also free to compare more than two tools if you want to, and you can use the same tools for more than one stage of the pipeline - e.g. DVC does both data versioning and artifact tracking and so it can be included in both the Experiment and Artifact Tracking and the Data Versioning stages.

Project Check-In Deadlines & Deliverables

Your team will be required to provide me with ~weekly project check-ins. I will create a folder in Piazza related to each deliverable. By the deadline, for each deliverable please post in Piazza your answers to the following.

Tue 3/21: Group Details.

- Form a group of 3-4 people of your choosing and decide who will be responsible for posting updates on Piazza, then have that person post the members of the group, the group name, and answers to the below.

- What kind of company you are part of:

  - Big and established.
  - Mid-size.
  - Startup.

- What kind of data science team you are supporting:

  - Experimentation.
  - Traditional (core) ML team.
  - R&D ML team

- What kind of service or product your team is creating/supporting/offering.

On each of the given dates below, let me know which tools you will be assessing for the specified stage, and which team members will be the ones assessing them. It is fine if more than one team member, or even if ALL team members are assessing each tool.

- **Tue 3/28**: Experiment and artifact tracking, model registry

- **Tue 4/4**: Data versioning

- **Tue 4/18**: ML pipeline orchestration

- **Tue 4/25**: CI/CD, linting, styling, branching

- **Tue 5/2**: Web app prototyping

- **Tue 5/9**: Model monitoring

Final Submission

In lieu of presenting your project to the whole class, you will instead present your proposal to me through Zoom on either Thur May 11 or Tue May 16. Your presentation should be **less than 5 minutes**, and I will need 2 minutes afterwards to ask questions! Just pretend I am a very (self-) important CTO too busy and distracted to sit through

more than 5 minutes of anything important. If you exceed 5 minutes and I am unable to ask my questions, your grade will be impacted. I **will have to** end the presentation promptly at the 7 minute mark to let the next group into the meeting to start their presentation.

Your presentation should be a slide deck containing:

- an architecture diagram.

- which tools you compared for each stage.

- how you compared them (with suitable metrics, tables, and visuals).

- which tools you are recommending and why.

Note: each team member should speak for at least one minute of the presentation. I recommend you designate one person to navigate the slides for the entire duration, instead of each speaker sharing their screen when it's their turn to speak.

Grading

Your Project grade will be based on (out of a total of 100 points):

- The seven project check-in deliverables in the above Project Check-In Deadlines & Deliverables section (each is worth 1 point).

- Slide deck is free of errors, looks professional, contains necessary information (10 points).

- Presentation under 5 minutes (10 points).

- Each group member speaks (7 points).

- Six tool comparisons with convincing evidence (each is worth 11 points).

To earn a grade, you must also grade the other members of your group, with your reasoning for each grade. I will use the grades you assign, as well as my own evaluation of each group member's presentation, to assign final participation grades.