ML Engineering Interview Process

Table of Contents

[The Process 1](#_Toc164954421)

[General Interview 2](#_Toc164954422)

[Introductions 2](#_Toc164954423)

[Generic Questions 3](#_Toc164954424)

[Technical Questions 3](#_Toc164954425)

[Final Questions 4](#_Toc164954426)

[Candidate Questions 4](#_Toc164954427)

[Engineering Challenge 4](#_Toc164954428)

# The Process

1. Screening Interview
   * To be conducted by HR.
2. General Interview
   * To be conducted by members of the ML Engineering Guild
   * Introduce the Lab and the interviewers to the candidate and vice versa.
   * Ask general questions about the person, their work experience and professional competencies.
   * Ask technical questions about their experience with different technologies and tools.
3. Coding Challenge
   * Provide the candidate with a coding challenge to complete for their second interview.
   * The challenge should be appropriate for the level of seniority and challenge the candidate to use the competencies we are looking for.
   * Ideally the challenge should allow the candidate to be able to exhibit their own style.
   * This should be submitted in advance of the technical interview to allow for preparation time for the interviewers.
4. Technical Interview
   * To be conducted by members of the ML Engineering Guild.
   * Review the outcome of the coding challenge by asking the candidate to explain what they have done.
   * Ask them why they made certain choices and how they approached the problem.
   * Ask them to make a change live on the call.
   * Show the candidate some code and ask them to review it as if they were reviewing a pull request.
   * This code should include some notable mistakes / talking points.
5. Personal Fit Interview
   * To be conducted by members of the DataLabs Leadership Team.

## General Interview

### Introductions

1. Introduce the interviewers.
2. Ask the candidate to introduce themselves.
3. Short background on Experian.
4. Introduction to the Lab:
   * Independence to rest of the business
   * Size and make-up of the Lab (Product Managers, Data Scientists and Engineers)
   * How we find work (finding opportunities inside and outside the business)
   * How we work (i.e. projects, MTPs, guilds)
5. Explain the role:
   * Mix of Data Science and Software Engineering
   * Solution architecture
   * Cloud infrastructure
   * Good coding practices
   * Introducing new practices
   * Maintaining infrastructure
   * Project leadership
   * ML Ops
   * Productisation

### Generic Questions

For example:

* Personality:
  + Why are you looking for a new job/role?
  + Why do you want to leave your current organisation?
  + Why do you want to work at Experian?
  + Why do you want to be an ML Engineer?
  + What are your long/short term career aspirations/goals?
  + What motivates you in a job?
  + How would others describe you?
  + What are your biggest strengths/weaknesses?
  + What’s something you’re proud of?
* Previous work experience:
  + What’s an example of a mistake you’ve made at work?
  + Do you have an example of a project that hasn’t gone well? What would you do differently if given another chance?
  + Do you have an example of when you’ve disagreed with a decision someone has made at work? How did you react?
* Work ethic:
  + What do you do when you’re stuck?
  + How would you approach leading a project?
  + How do you handle a situation where you believe your project manager is ineffective?
  + How would you handle handing over a project to someone who is uncollaborative and obstructive? Has that happened to you before?
  + Do you consider yourself to be a pragmatist or a purist?
  + Have you ever introduced new tech to your team/project? What was it? How did you introduce it?

### Technical Questions

For example:

* Are there any software patterns you particularly like and use? (We like DI, DDD, TDD)
* Do you have experience with IaC?
* What experience do you have with GenAI / RAG? Pain points?
* What is your experience with containerisation? Docker? Kubernetes?
* Do you have experience with CICD pipelines? What is their purpose?
* What is your cloud compute experience? What tools have you used? AWS – Sagemaker, Lambda, ECR? Azure – AI Search, Container Apps, App Service?
* Do you have experience with Dask or Spark? Pros and cons?
* What are your first thoughts when considering how to improve performance of code? Specifically, python?
* Have you built APIs?

### Final Questions

For example:

* When could you start?
* What timeframe do you need to hear by?
* What is your availability for coding challenge and next interview?
* Salary expectations and holidays done by HR?

### Candidate Questions

## Engineering Challenge

The brief below will set out a challenge and the expectation is you shouldn’t spend more than 4 hours on it – don’t worry if you don’t have time to finish.

### Brief

A Business Unit has a training pipeline that creates an XGBoost model monthly to identify potential fraudsters based on their credit data stored in Snowflake. They would like you to create an inference pipeline that allows them to call the model using a REST API. All the features will be provided in the request, and they would like the fraud probability predicted by the model returned as an output.

The training pipeline isn’t fully complete. The business logic is mostly finished but they have not come to a decision on the cloud platform to run the training pipeline in and how to store the training artifacts.

They are looking for support in designing the production training pipeline and for you to build an inference pipeline that can be deployed alongside it. Later on, the Business Unit would like you to review the training pipeline code and guide them on how to follow your suggestions for getting production-ready.

The Business Unit are all on holiday so you can’t ask them any questions before you present your work, so note down any assumptions you make and they can be addressed when they return.

Deliverables:

* Inference pipeline for model json provided.
* Design for production inference and training pipelines.
* Plan for integrating the inference pipeline with the training pipeline.