You'll often encounter the Machine Learning Case Study interview during an onsite round for a Machine Learning Engineer (MLE), Data Scientist - Machine Learning (DS-ML) or Software Engineer-Machine Learning (SE-ML) position.

What to expect in the interview

The interviewer is evaluating how you approach a real-world machine learning problem. The interview is usually a technical discussion of an open-ended question. There is no exact solution to the problem; it's your thought process that the interviewer is evaluating.

Here's a list of questions you might be asked:

- How would you build a trigger word detection algorithm to spot the word "activate" in a 10 second long audio clip?
- An e-commerce company is trying to minimize the time it takes customers to purchase their selected items. As a machine learning engineer, what can you do to help them?
- You are given a data set of credit card purchases information. Each record is labeled as fraudulent or safe. You are asked to build a fraud detection algorithm. How would you proceed?
- You are provided with data from a music streaming platform. Each of the 100,000 records indicates the songs a user has listened to in the past month. How would you build a music recommendation system?

Resources

Useful content to prepare for this interview:

- deeplearning.ai:
  - Structuring your Machine Learning Project (Coursera)
- Stanford Deep Learning class (CS230):
  - Deep learning intuition (video)
  - Full-cycle deep learning projects (video)
  - AI+healthcare case studies (video)
  - Deep learning project strategy (video)
  - Case study on conversational assistants (video)
- Search for case studies from the companies in the same industry as the ones you're interviewing with. Here are examples of company ML case studies:
  - Machine learning-powered search ranking of airbnb experiences (blog post)
  - Machine learning at facebook: understanding inference at the edge (paper)
  - Empowering personalized marketing with machine learning - Lyft Engineering (blog post)
  - Learning a personalized homepage - Netflix Tech Blog (blog post)

Recommended Plan

All interviews are different, but the *ASPER framework* is applicable to a variety of case studies:

1. **Ask**. Ask questions to uncover details that were kept hidden by the interviewer. Specifically, you absolutely want to answer the following questions: "what are the product requirements and evaluation metrics?", "what data do I have access to?", "how will the learning algorithm be used at test time, and does it need to be regularly re-trained?"
2. **Suppose**. Make justified assumptions to simplify the problem. Example of assumptions are: "we are in small data regime", "Human-level error is 7%", "the data distribution won't change over time", etc.
3. **Plan**. Break down the problem into sub-problems. A common plan in the machine learning case study interview is: (i) Data collection, labelling and preprocessing, (ii) Modelling and (iii) Deployment.
4. **Execute**. Announce your overall plan, and tackle the sub-problems one-by one. In this step, the interviewer might ask you to write code or explain the maths behind your proposed algorithms.
5. **Recap**. At the end of the interview, summarize your answer and mention the tools and frameworks you would use to perform the work. It is also a good time to give your ideas on how the problem can be extended.

Here are useful rules of thumb to follow:

- In the Machine Learning Case Study, the interviewer will also evaluate your excitement for the company's product. If you are interested, make sure to show your curiosity, creativity and enthusiasm.
- Listen to the hints given by your interviewer. **Example**: In an imbalanced clinical data case study, you are using logistic regression to classify if a patient's health is at risk (1) or not (0). You focus on explaining the algorithm, while your interviewer asks you "what are you optimizing for?". You know logistic regression and answer "we're minimizing the logistic loss function (or binary cross-entropy)". Your interviewer follows up with "what are some properties of the data set?" In this scenario, the interviewer is probably expecting you to connect the dots between your loss function and the imbalanced data set. In fact, you might want to weigh the terms in your loss function to account for the data imbalance.
- Don't focus solely on the modeling unless the interviewer asks you to do so. Many candidates are only interested in what model they will use and how to train it, while most of the problem lies elsewhere, i.e. in the data pipeline, deployment, or problem-specific challenges.
- Because case studies are often open-ended and can have multiple valid solutions, avoid making strong statements such as "the correct approach is …" It might offend the interviewer if the approach they are using is different from what you describe. It's also better to show your flexibility with and understanding of the pros and cons of different approaches.
- Read about the company before this interview. The Machine Learning Case Study is often inspired by an in-house project. Is the team working on a domain-specific application? **Example 1**: If the team is working on a face verification product, take a look at the Face Recognition lessons of the Coursera Deep Learning Specialization (Course 4), as well as the DeepFace (Taigman et al.) and FaceNet (Schroff et al.) papers prior to the onsite. **Example 2**: If the team is building an autonomous car, you might want to read about topics such as object detection, path planning, safety, or edge deployment.
- When out of ideas or stuck, think out loud rather than staying silent. Talking through your thought process will help the interviewer correct you and point you in the right direction.