

Part 3

# MACHINE LEARNING INTERVIEW QUESTIONS

SWIPE →



## Q1: Name an example where ensemble techniques might be useful.

---

Ensemble techniques use a combination of learning algorithms to optimize better predictive performance.

They typically reduce overfitting in models and make the model more robust (unlikely to be influenced by small changes in the training data).

You could list some examples of ensemble methods (bagging, boosting, the “bucket of models” method) and demonstrate how they could increase predictive power.

## Q2: How do you handle missing or corrupted data in a dataset?

---

You could find missing/corrupted data in a dataset and either drop those rows or columns, or decide to replace them with another value.

In Pandas, there are two very useful methods: `isnull()` and `dropna()` that will help you find columns of data with missing or corrupted data and drop those values.

If you want to fill the invalid values with a placeholder value (for example, 0), you could use the `fillna()` method.

## Q3: What do you think is the most valuable data in our business?

---

These questions try to test you on two dimensions. The first is your knowledge of the business and the industry itself, as well as your understanding of the business model.

The second is whether you can pick how correlated data is to business outcomes in general, and then how you apply that thinking to your context about the company. You'll want to research the business model and ask good questions to your recruiter—and start thinking about what business problems they probably want to solve most with their data.

## Q4: Where do you usually source datasets?

---

Machine learning interview questions like these try to get at the heart of your machine learning interest.

Somebody who is truly passionate about machine learning will have gone off and done side projects on their own, and have a good idea of what great datasets are out there.

If you're missing any, check out Quandl for economic and financial data, and Kaggle's Datasets collection for another great list.

## Q5: What's the “kernel trick” and how is it useful?

---

The Kernel trick involves kernel functions that can enable in higher-dimension spaces without explicitly calculating the coordinates of points within that dimension: instead, kernel functions compute the inner products between the images of all pairs of data in a feature space.

This allows them the very useful attribute of calculating the coordinates of higher dimensions while being computationally cheaper than the explicit calculation of said coordinates. Using the kernel trick enables us effectively run algorithms in a high-dimensional space with lower-dimensional data.

## Q6: How do you ensure you're not overfitting with a model?

---

This is a simple restatement of a fundamental problem in machine learning: the possibility of overfitting training data and carrying the noise of that data through to the test set, thereby providing inaccurate generalizations.

- Keep the model simpler: reduce variance by taking into account fewer variables and parameters, thereby removing some of the noise in the training data
- Use cross-validation techniques such as k-folds cross-validation.
- Use regularization techniques such as LASSO that penalize certain model parameters if they're likely to cause overfitting.

## **Q7: What's your favourite algorithm, and can you explain it to me in less than a minute?**

---

Interviewers ask such machine learning interview questions to test your understanding of how to communicate complex and technical nuances with poise and the ability to summarize quickly and efficiently.

While answering such questions, make sure you have a choice and ensure you can explain different algorithms so simply and effectively that a five-year-old could grasp the basics!



## Q8: Do you have research experience in machine learning?

---

Most organizations hiring for machine learning positions will look for your formal experience in the field.

Research papers, co-authored or supervised by leaders in the field, can make the difference between you being hired and not.

Make sure you have a summary of your research experience and papers ready—and an explanation for your background and lack of formal research experience if you don't.



**FOUND IT HELPFUL?**

Let us know in the comments

-----

At DDS, we will help you build a career in Data Science

[LINK IN BIO](#)

-----