Gojek Data Science Platform

About the team

Thank you for your interest in Gojek and in our team. We are the Data Science Platform team and we are part of the Marketplace organization in Gojek.

Gojek Data Science works on some of the most interesting problems in transport, logistics, and economics. We leverage machine learning to build data products for ride-hailing, logistics, food delivery, and payments. From selecting the right driver to dispatch, to dynamically setting prices, to serving food recommendations, to forecasting real-world events, to detecting fraud and preserving trust, we process hundreds of millions of orders per month, across more than 18 products, in four countries. All are driven by machine learning.

The Data Science Platform team is relatively small (about 14 people as of February 2021) and mostly located in Singapore but with some members in Jakarta. We serve the needs of many dozens of data scientists and analysts throughout Gojek in all the various countries we operate. Most of these data scientists are located in Singapore, but there are also others in our Jakarta HQ, Bangalore development office, and the remote offices in Thailand and Vietnam.

We have two engagement models. Traditionally, we have been working closely with the data scientists in developing and enhancing their systems to ensure they are production-ready. This “consulting” approach has provided a lot of learnings but is getting difficult to scale. Now, we are looking at building a self-service platform that data scientists can easily leverage for most of their work. We will continue to develop this platform to cover all of the data science needs of Gojek and also support our old engagement model for special cases.

You can find some examples of our work and team members online:

- A series of blog posts about our platform vision.
- Feast: a feature store we developed together with Google
- DSSG talks
- A Meetup talk about MLflow usage at Gojek
- A KubeCon talk about Kubeflow
- PyData talks
- Some posts on our official blog

We also work closely with other teams in the broader Data organization. Specifically, we depend on Data Engineering (located in Bangalore) for a lot of our raw data streams such as clickstream, GPS logs, etc, and Business Intelligence (located in Jakarta) for our business metrics.

We also rely heavily on managed services either provided by Google Cloud (Kubernetes, BigQuery, Dataproc, Dataflow) or the Gojek systems team (DNS, security). We also work closely with Google on various open-source projects such as Tensorflow and Kubeflow.

We use a large variety of modern technologies but our primary ones include Google Cloud Platform (GCP), Kubernetes, Docker, Terraform, Golang, Python, Java. We are one of the biggest users of GCP and Kubernetes in the region.

About the role

In this role, you will work with data scientists to build end-to-end machine learning frameworks that directly impact our products. You will use your experience in software engineering, data systems, distributed systems, machine learning, artificial intelligence, and deep learning to develop these systems and take them into production.

Our team members have a mixture of different skills but fall broadly into two archetypes: *infrastructure engineer* or *machine learning engineer*.

Infrastructure engineers are comfortable with cloud platforms (GCP preferred, but AWS and Azure are great too), networking (including network security), relational or non-relational databases (optimization, high availability), and “Big Data” processing systems such as Spark or Flink. They are fluent in Golang, Python and shell scripting. They understand operational concerns such as service level agreements and monitoring.

Machine learning engineers work more closely with data scientists. They are comfortable with the tools that data scientists prefer: Python with Pandas and Scikit, Tensorflow/Keras/PyTorch, xgboost, and others. They are comfortable with ML concepts though they do not need to be completely proficient (that is for the data scientists!) and also extremely proficient with software engineering concepts such as software design, testing, deployment, and SRE.

Our daily work is a mixture of these two extremes so an ideal candidate will have at least some skills from both ends of the spectrum. We also need to work with the more open-ended R&D process that is typical of data science.

We are looking to grow our team rapidly so while we will consider candidates at all levels, we are also preferring those who are more senior. These seniors will help lay the foundation for our team’s practices and culture by bringing in their own experiences and providing mentorship for more junior team members. They will help drive the numerous projects we are undertaking.

About the interview process

The interview process has two principles:

- It is not just for us to find out more about you, but for you to find out more about us.
- We are trying to find people to expand the capabilities of the team, not just to check off line items on the job description.

The process has three stages, each one building upon the previous stages to provide both sides more information about each other and refined to each of your particular experience and skills.

Take-home test

The initial stage is a take-home test. There are two versions of the test, one that is more infrastructure and another that is more software engineering. It is intended as a realistic example of the kind of work we do daily. To that end, it is neither trivial nor is it the most exciting.

You will be given one week to complete the test but we believe it can be completed in about 8-10 hours if you are already familiar with the technologies. Unfortunately, there is a bit of Google Cloud usage in order to access the data, but beyond that, most of the technology choices are up to you.

We are looking for a “production-ready” submission, so all the usual concerns apply: deployability, maintainability, documentation, testing, etc. We are reviewing the quality of the submission, not just the correctness of the output.

Design and coding interview

If we like your take-home submission, the next round is a design and coding interview. This will take about two hours and you will meet two members of our team.

This starts with you walking us through your submission. We want to see how you present your work. We will then pose a new requirement and you will have to walk us through the process to get this incorporated into your submission. Finally, in the remaining time, we will pair program with you to code this extended requirement. Here, we are evaluating how you would work with us as a team, your fluency with tools of the trade, and also seeing if your submission can be easily understood and extended.

Resume deep dive

The final stage is a resume deep dive, which will feel more like a “regular” interview. This is another two hours with another two members of the team.

We will talk about everything relevant that is on your resume and, based on your performance in the earlier stages, will want to focus on certain areas. The goal here is to understand what skills and experiences you could bring to the team. Feel free to ask us questions too! There may be some whiteboard design/coding as well.

Next steps

Our recruitment coordinator will send you the take-home test. Please do reach out if you have any questions about the test.