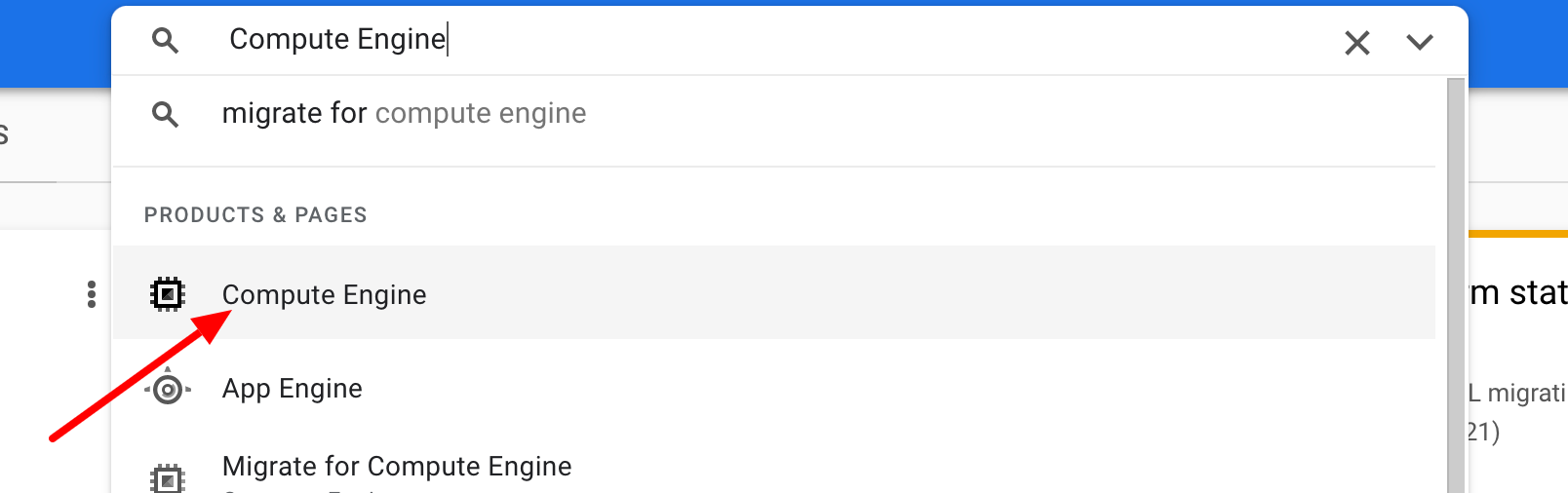
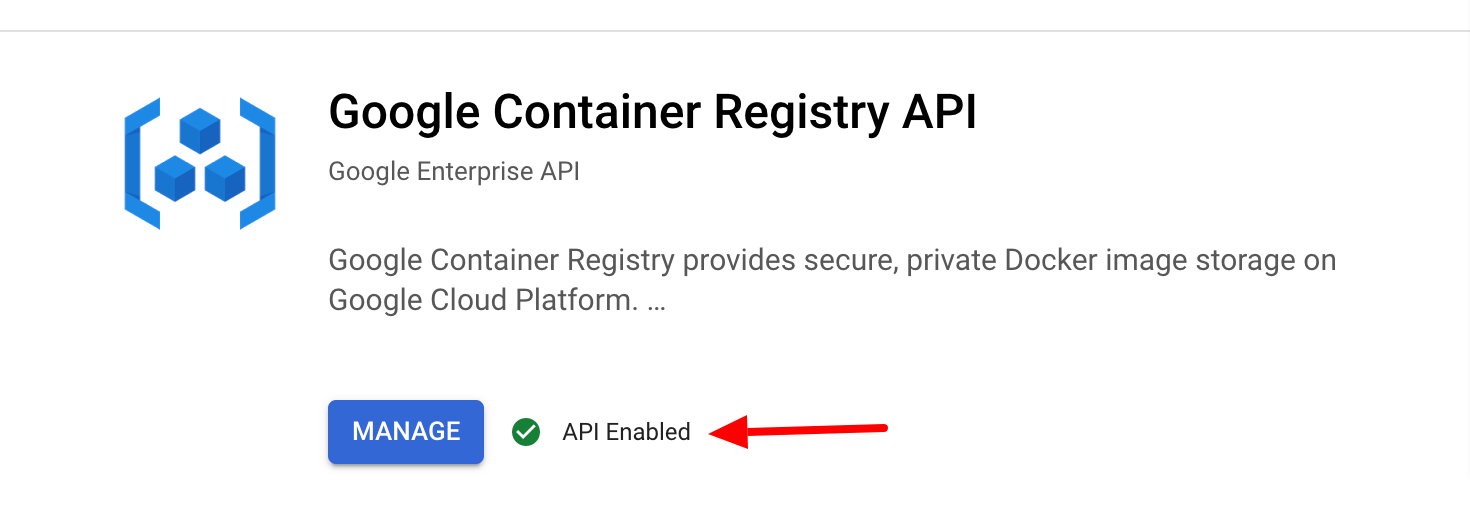
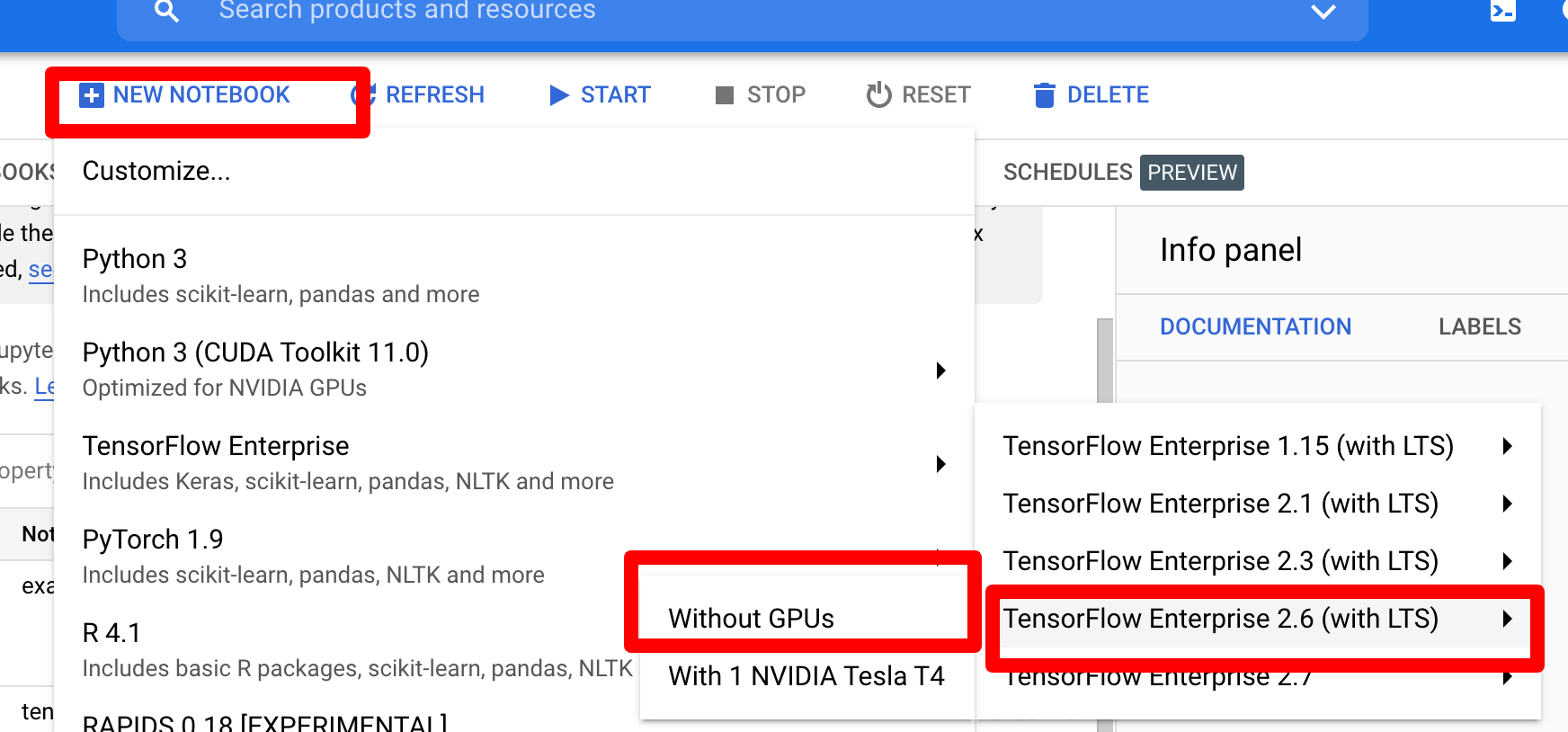
1. Enabling the Compute Engine API, Vertex AI API and the Container Registry API.

To enable navigate to each API from your google console dashboard as seen below.





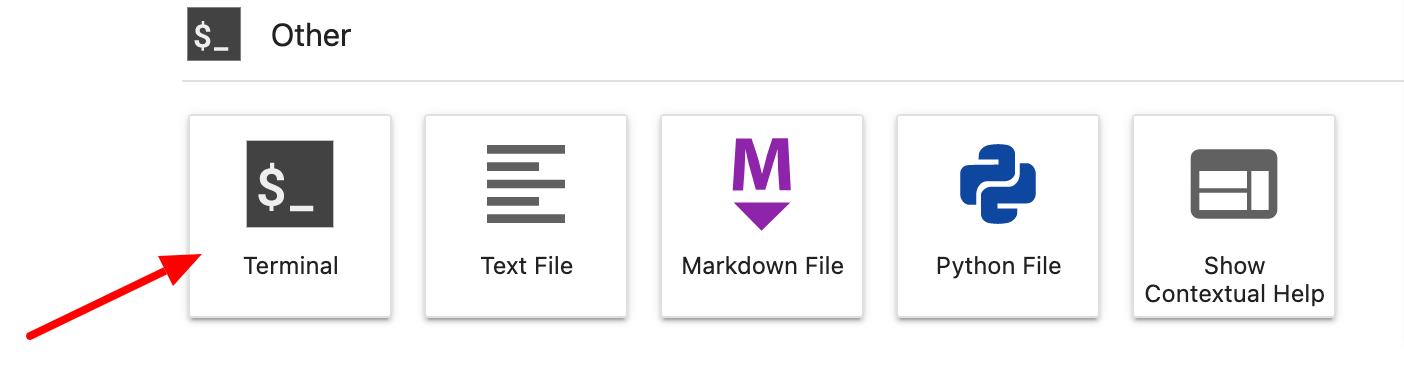
1. Creating a Vertex AI Workbench instance



Click New Notebook and select the following option, then click create with defaults selected to create a new notebook.

1. Containerizing training code

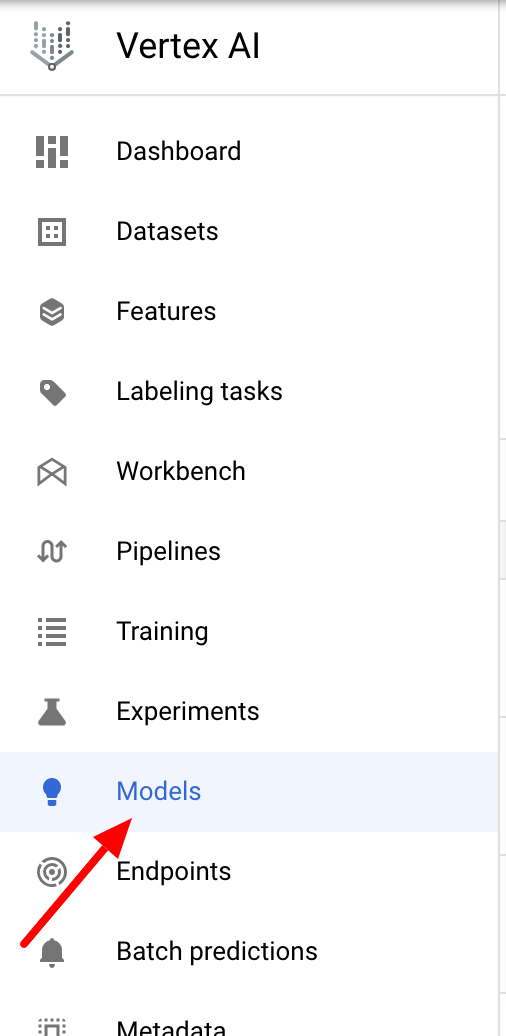
From the created notebook in Step 2 click to open Jupyterlab and create a new terminal instance.



In this tutorial we will go over submitting a training job to Vertex by putting our training code in a Docker container and pushing this container to Google Container Registry. Using this approach, we can train a model built with any framework. The complete code can be found here.

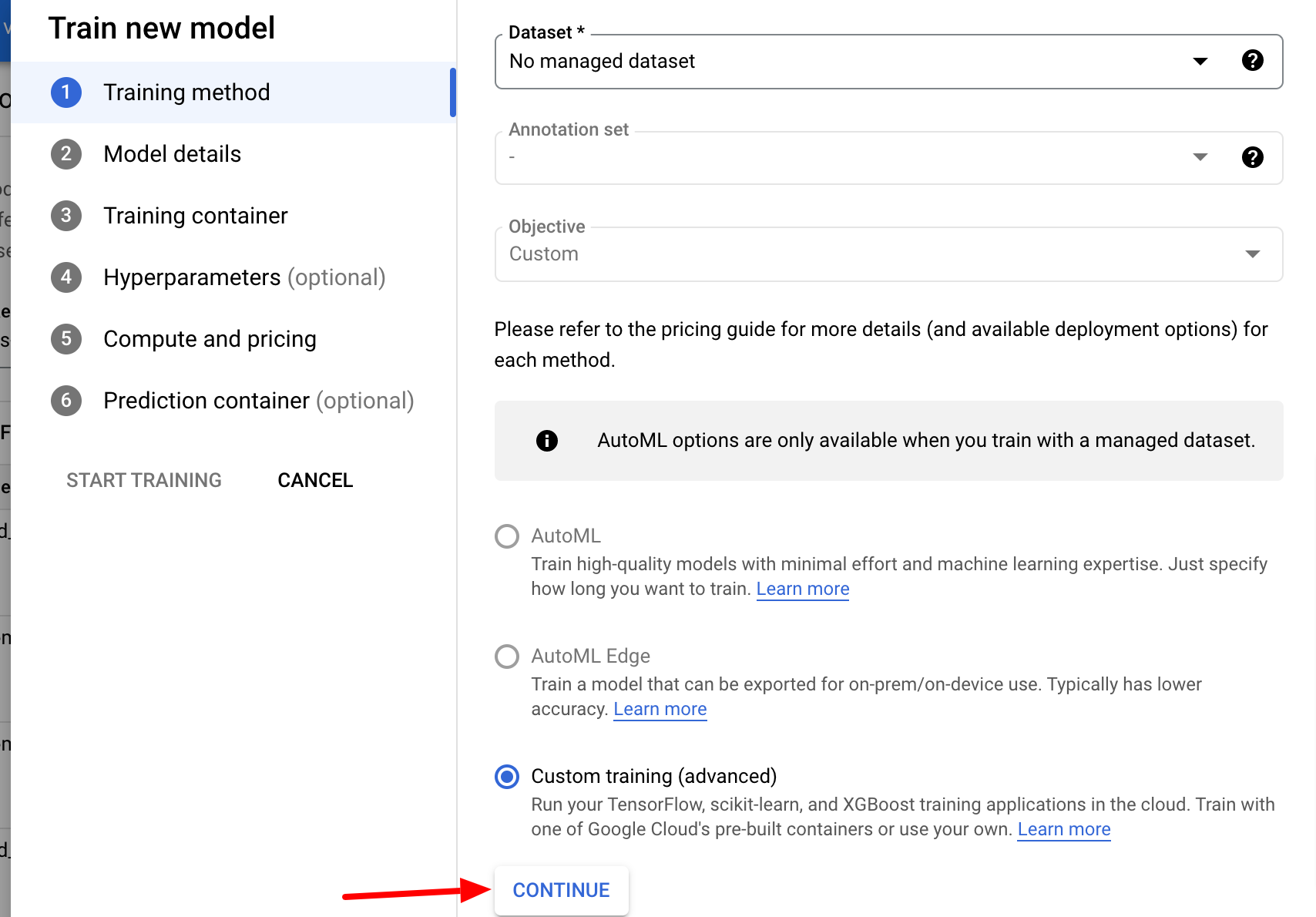
1. Custom training via our own custom container on Google Container Registry:

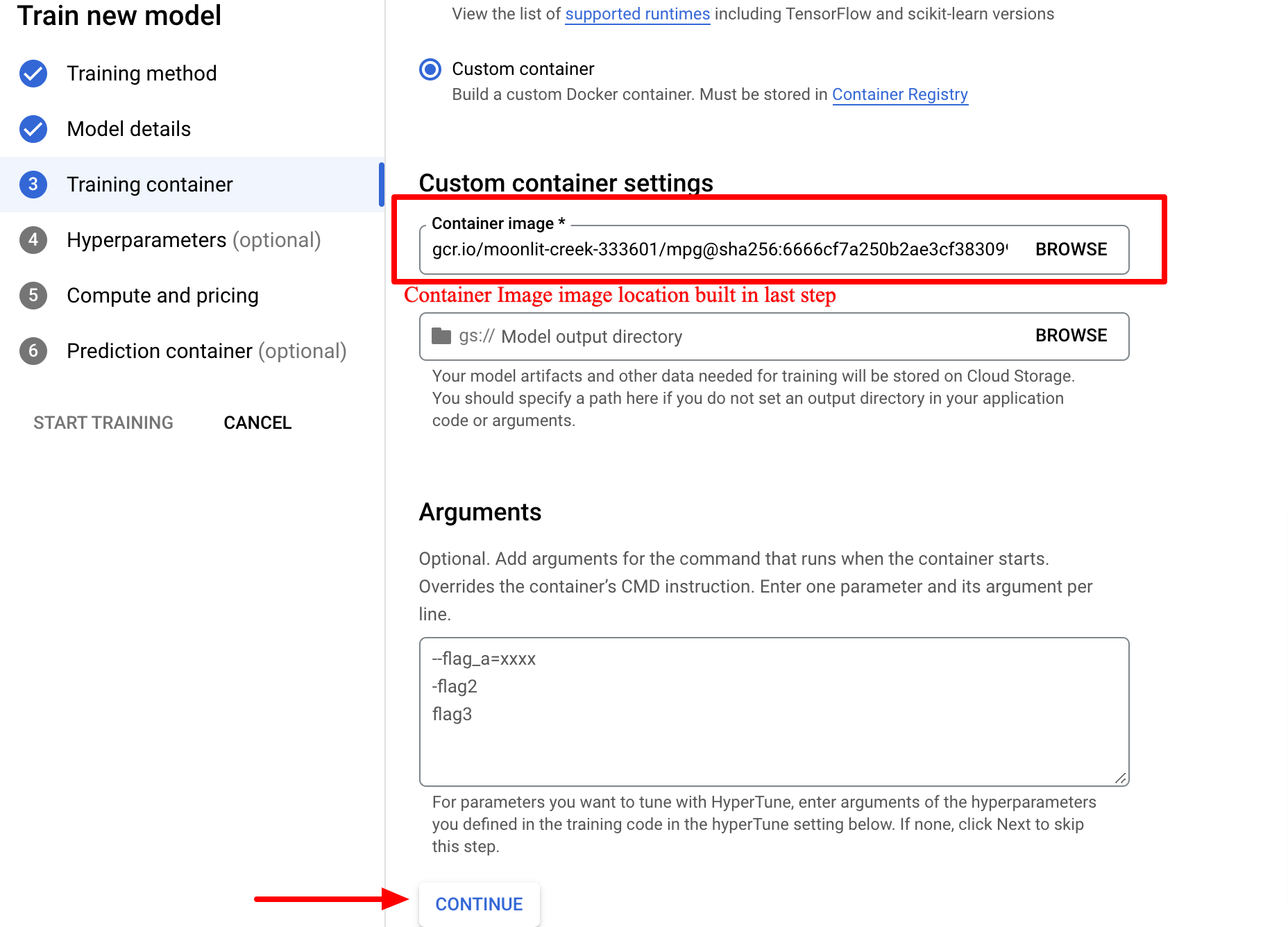
Navigate to models in Vertex AI API dashboard:



On the Models page click Create for a Custom Training job.

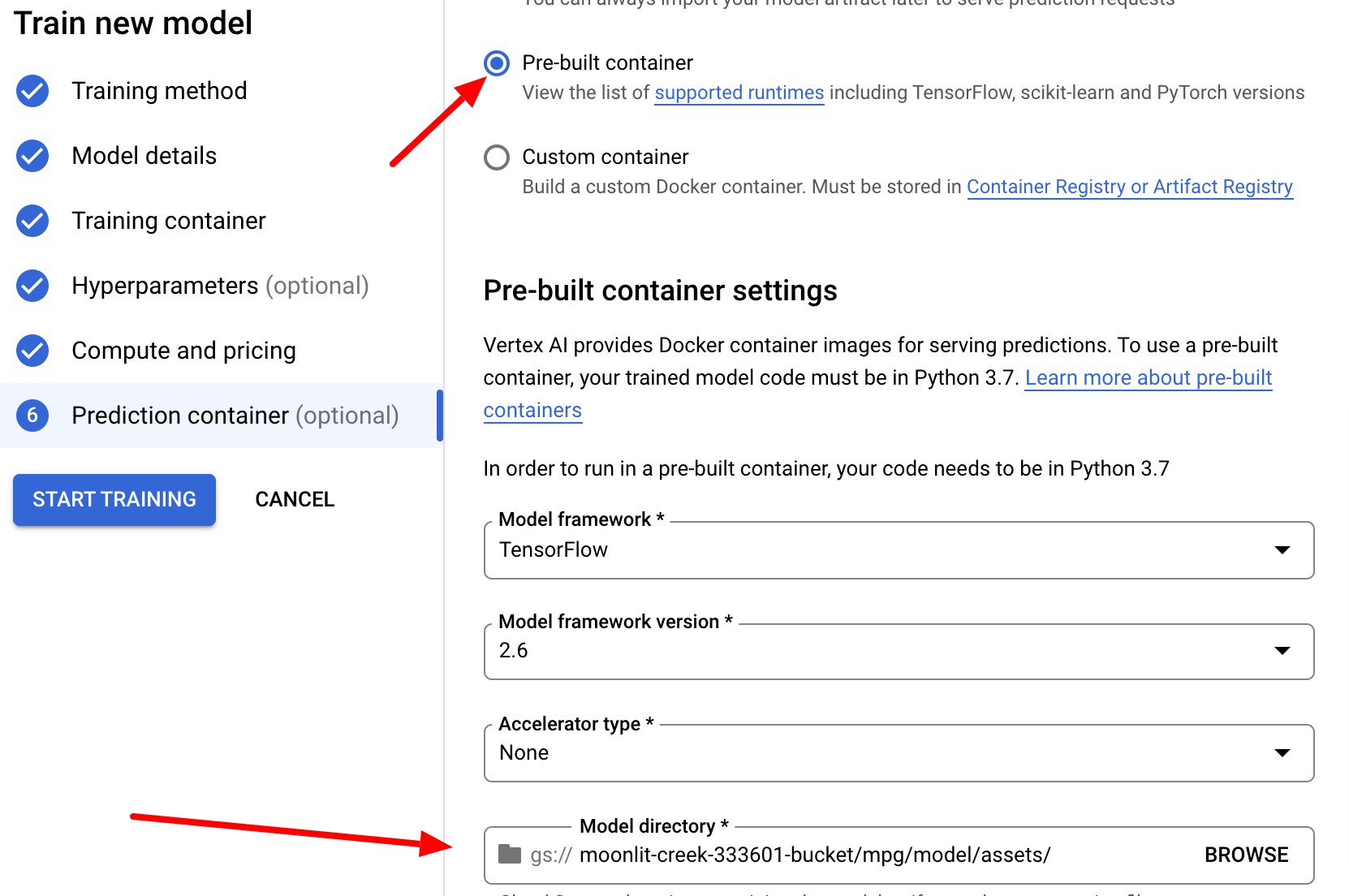
1. Parameters for starting custom training:





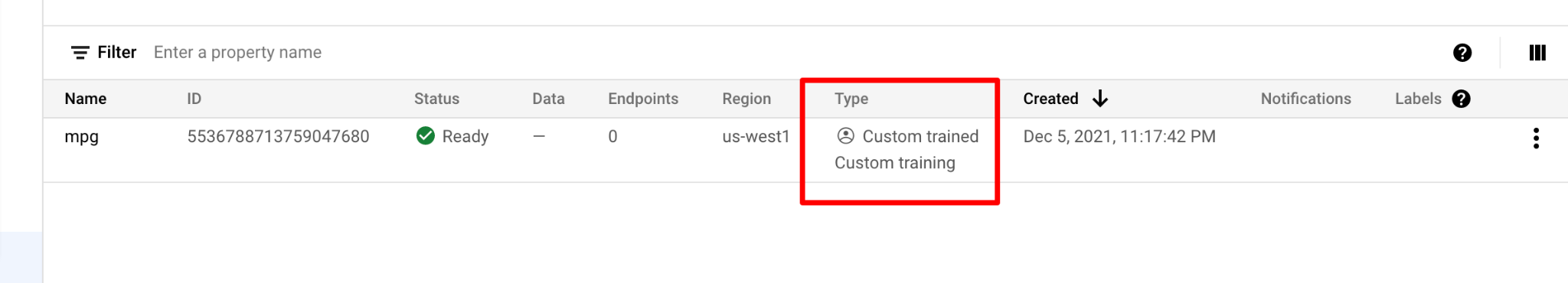
After clicking Continue leave Hyperparameter tuning unchecked as we are not doing that right now.

For Prediction select Pre-built container and click Start Training. The bucket name is where trained model will be exported and should be correct.

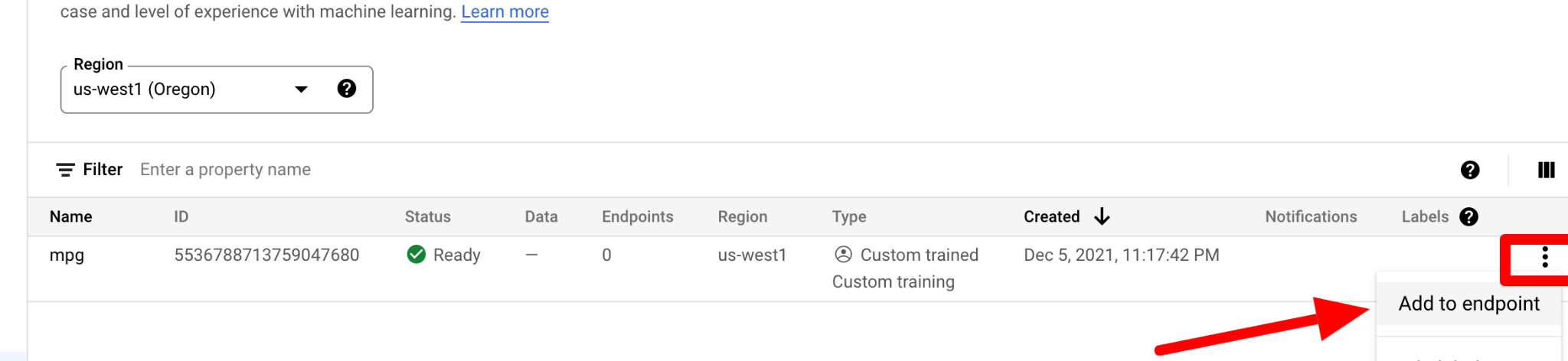


1. Deploying a model to endpoint:

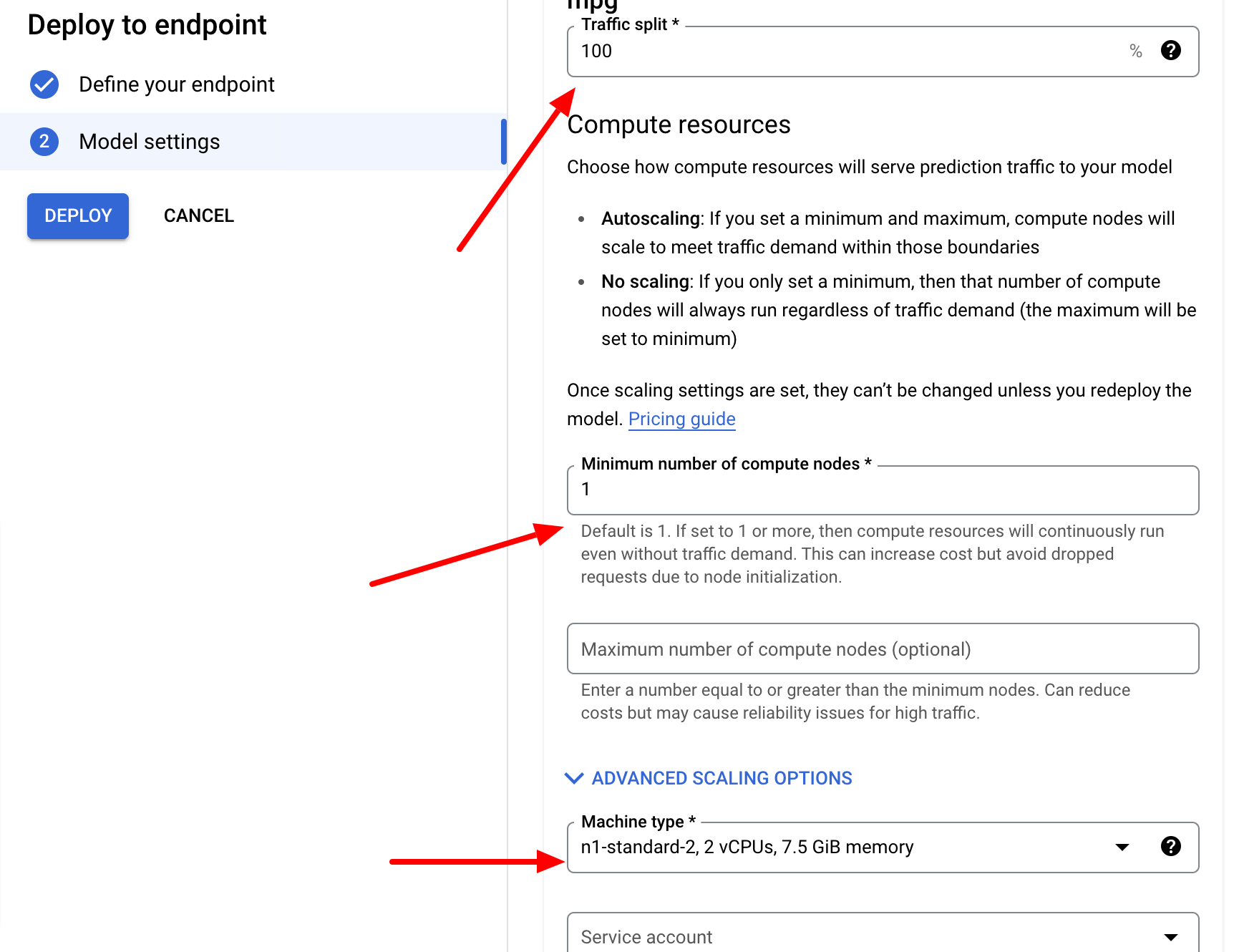
Navigating to models we can see trained model with custom training



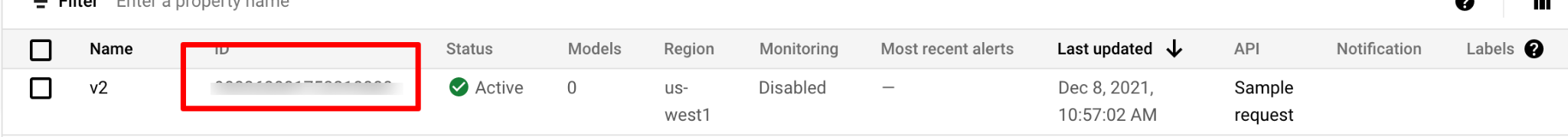
Deploying to endpoint:



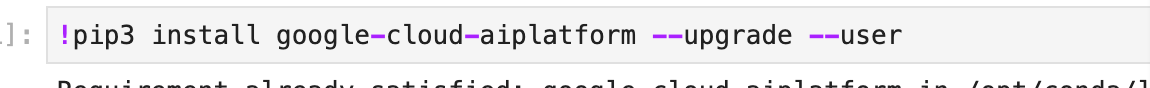
Then create a new endpoint with following parameters:



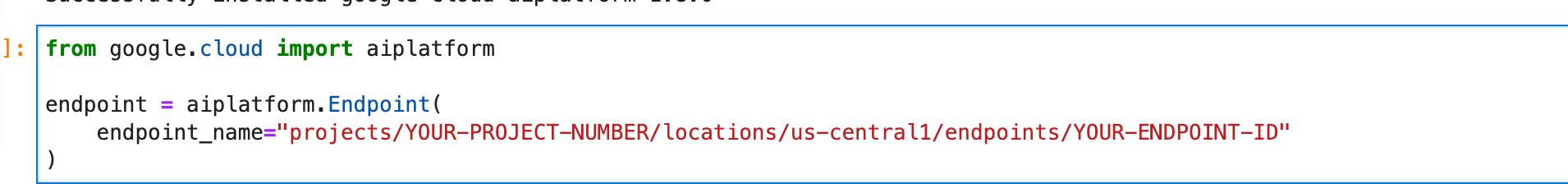
Click Deploy to push the model to the endpoint. After the endpoint is ready get hold of the Endpoint ID and navigate to the project dashboard to get the project number. We will need these details to make a prediction on our custom trained model.



From the workbench instance click the launcher to create a new Python notebook in the jupyterlab. Install the Vertex AI SDK using the following command in a new notebook cell :



Then run these commands to refer to the newly created endpoint, replacing the values with your project values:



Running a prediction using the deployed endpoint:



1. Cleanup:

Ndeploy model from endpoint and delete storage bucket.

