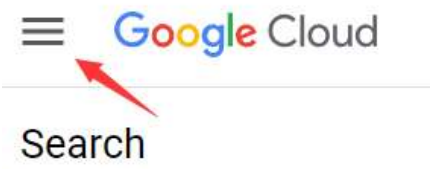
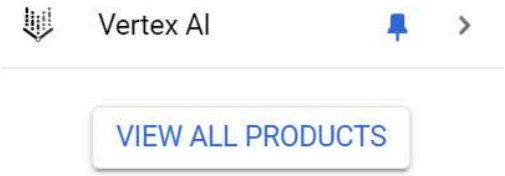
Appendix A
A Simple Guide on GCP Tools

How to Find the Tools

Sign up or Sign in to your Google Cloud Platform. Create a project or select an existing project first (nyc-ddc is our project name here). Then in your Google Cloud Platform Console, go to the right-upper corner, click the sidebar.

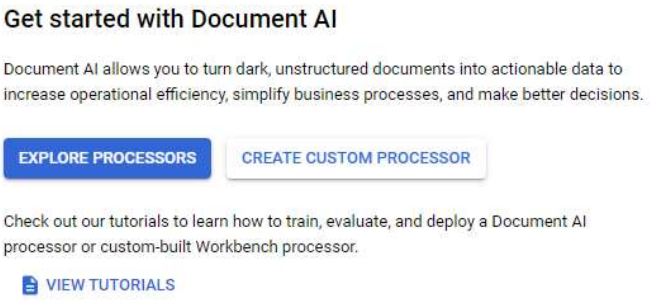


If you have already enabled the products/APIs you want, you can see them in the sidebar. Or click “View All Products” to search and enable them.

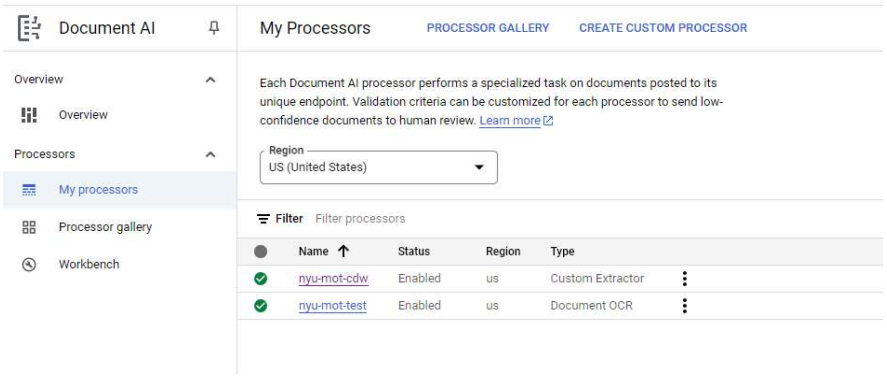


Document AI - Get Started

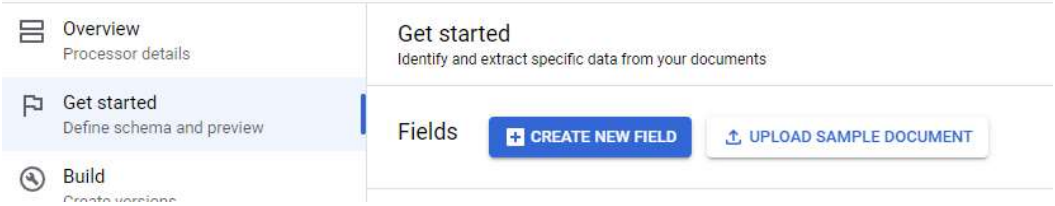
Find and enable the Document AI feature of GCP. Click “Create Custom Processor” to create a new processor, or you can click “Explore Processors” to use the existing models you created.



Select My processors in the sidebar, and select the corresponding processor in the filter. In this project, "nyu-mot-cdw'' is our processor. You can create your own processor in the “Processor Gallery” tab in the sidebar.

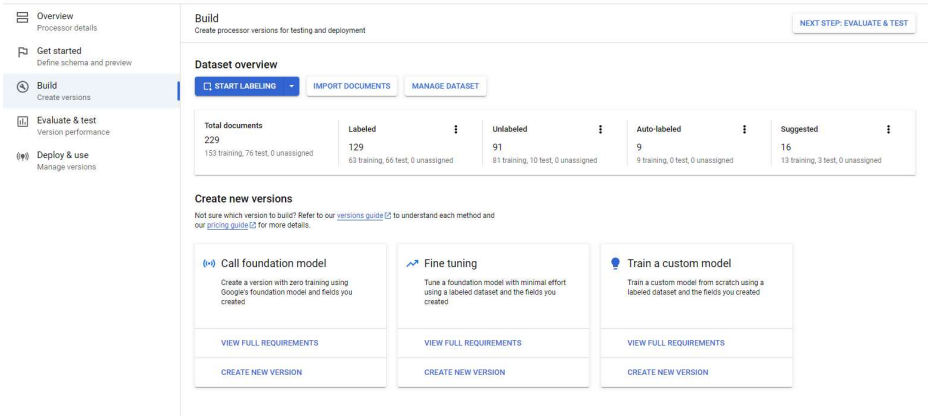


Find Get Started in the sidebar. You can click “Create New Field'' to specify your fields, which are the information you want to extract from the documents. After you define the fields, you can upload the sample file to check your setting/schema.

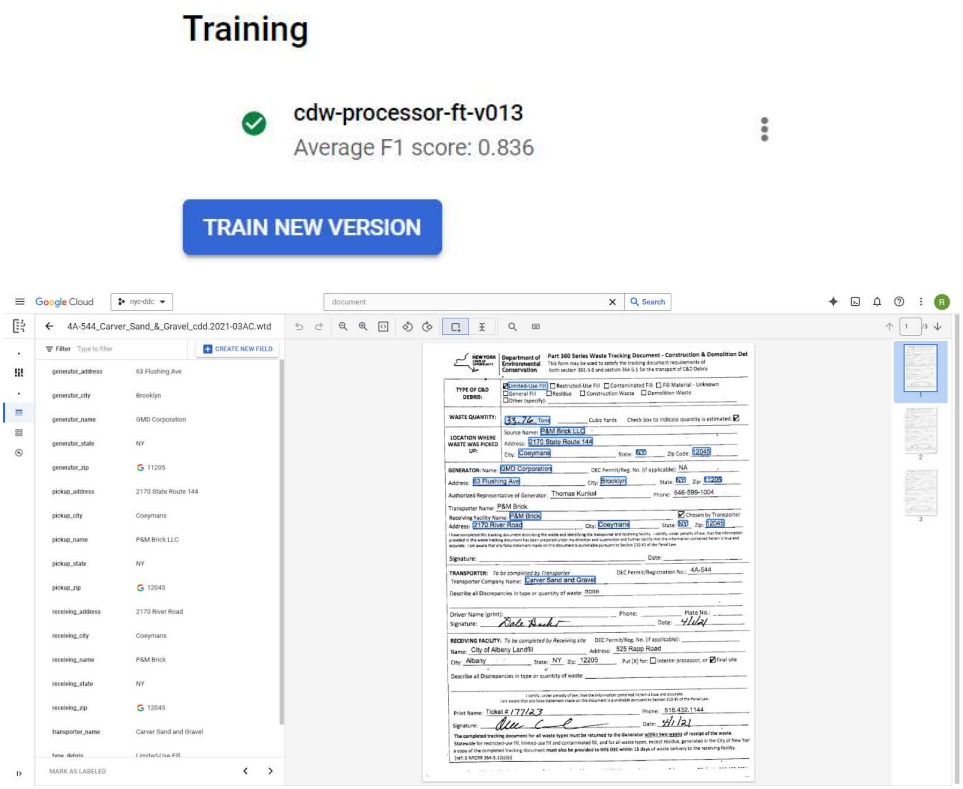


Document AI - Build

The Build page is used to create our model for testing and deployment. You can upload your training data, which should be some documents here. And like most supervised models, you should do the labeling.

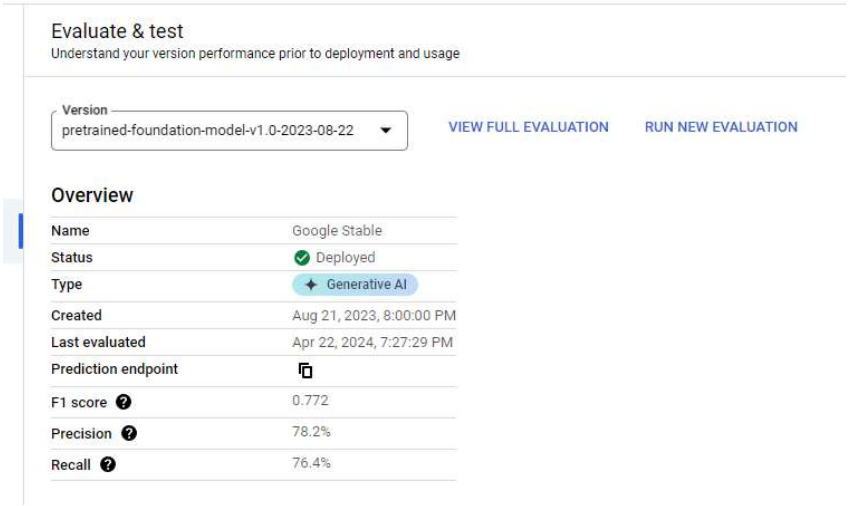


In the “MANAGE DATASET”, you can manage your dataset, do the labeling, upload more documents, and train your model.



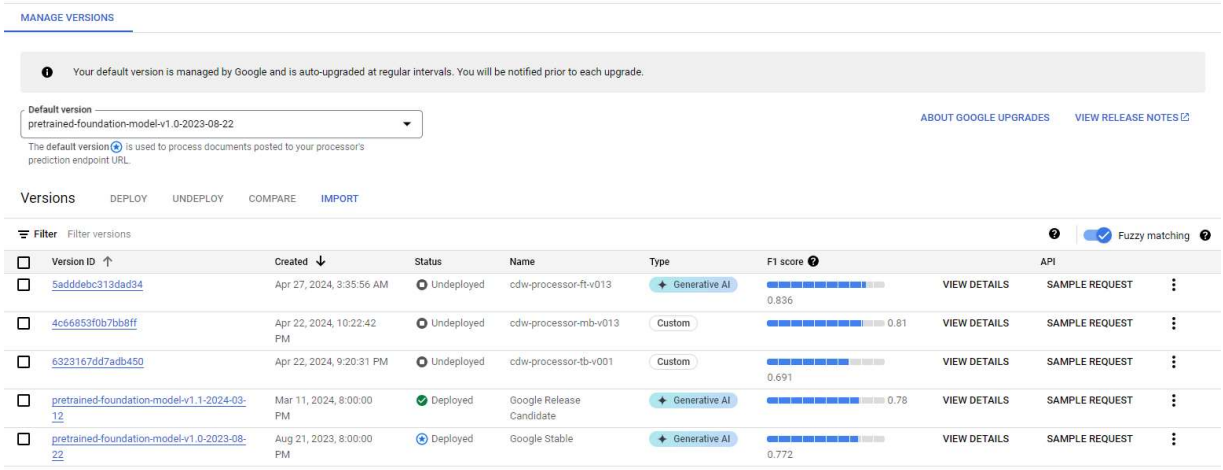
Document AI - Evaluate & Test

You can get an idea of the performance of models before deployment, including the F1 score, Precision, and Recall, all these shows the accuracy of your models.



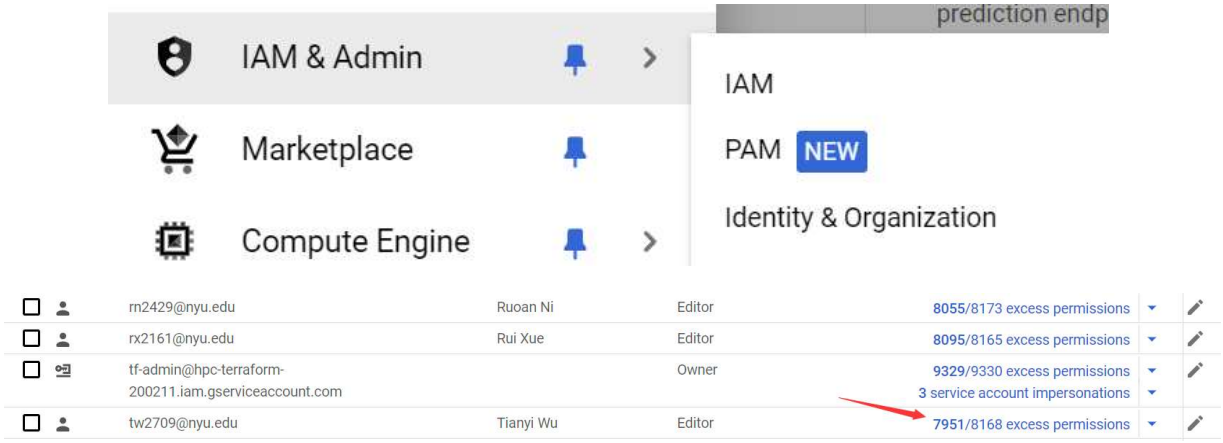
Document AI - Deploy & Use

You can train several versions for models. In this page, you can choose which version is your default version, and which you want to deploy for use, and which version you want to discard.



Document AI - How to Call it via API

First, check your access permission. In most cases, if you have the editor permission in your project, you will also have access to the products/APIs used in the project.



Permissions for tw2709@nyu.edu

Current permissions for Editor role

```
documentai.processors.create
documentai.processors.delete
documentai.processors.fetchHumanReviewDetails
documentai.processors.get
documentai.processors.list
documentai.processors.processBatch
documentai.processors.processOnline
documentai.processors.update
documentai.processorTypes.get
documentai.processorTypes.list
documentai.processorVersions.create
documentai.processorVersions.delete
documentai.processorVersions.get
documentai.processorVersions.list
documentai.processorVersions.processBatch
documentai.processorVersions.processOnline
documentai.processorVersions.update
errorreporting.groups.list
iam.roles.list
```

Once you have confirmed that you have the permission to use Document AI, you should set up authentication in your work environment. The easiest way to do this is use Google Cloud CLI. You can download it here: link.

Then, run the following command in your Command Prompt/Anaconda Prompt/Powershell Prompt :

```
gcloud init
```

And run this command to create local authentication credentials for your Google Account. You will be asked to enter your GCP account and password.

```
gcloud auth application-default login
```

Check these locations for the json file. If the json file exists, you successfully set up authentication.

Linux, macOS: $HOME/.config/gcloud/application_default_credentials.json

Windows: %APPDATA%\gcloud\application_default_credentials.json

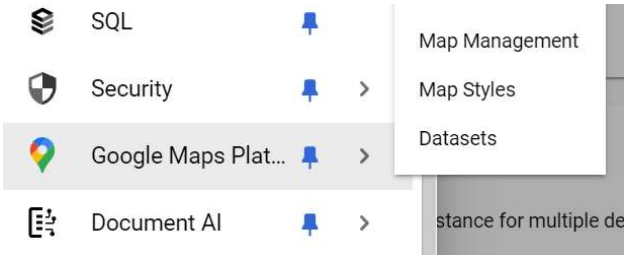
Now you can call Document AI, send it a request, use it to process your documents in your work environment, like your PC.

For code example, check :https://github.com/NYU-Tandon-TMI/cdw/blob/main/toolbox/document_extractor_one_folder.ipynb

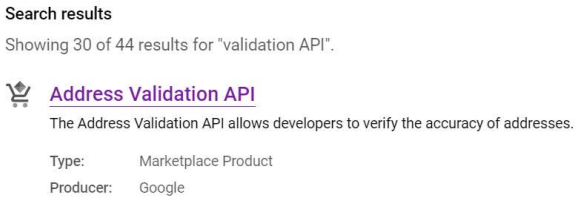
For more details, check: https://cloud.google.com/document-ai/docs/

Validation API

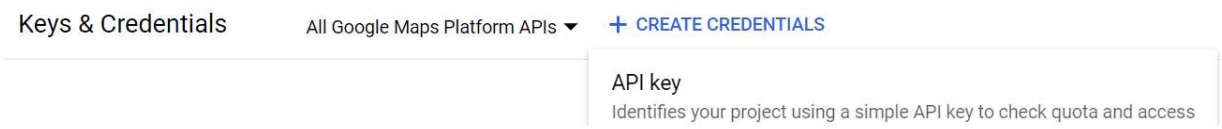
Validation API is among the numerous APIs & Services provided by Google Maps Platform. Similarly, find and enable Google Maps Platforms in the sidebar of your GCP console.



And go to the APIs & Services, find and enable Validation API



This API is easy to use. What you should do is create API keys, and put it in your request code.



For code example, check: https://github.com/NYU-Tandon-TMI/cdw/blob/main/toolbox/document_preprocess.ipynb

For more details, check: https://developers.google.com/maps/documentation/Address-validation