

Training the machine learning model

Last updated 2020-08-03

In IBM Watson™ Knowledge Studio , the creation of the machine learning model involves training the machine learning model and evaluating how well the model performed when annotating test data and blind data.

Creating a machine learning model

When you create a machine learning model, you select the document sets that you want to use to train the model and specify the percentage of documents that are to be used as training data, test data, and blind data.

About this task

By exploring the performance metrics, you can identify ways to improve the model's accuracy.

Restriction: Only three machine learning models can be trained at a time per Knowledge Studio instance. If your instance contains multiple workspaces and the number of machine learning models that are being trained in other workspaces totals 3 already, then your request to train the machine learning model in your workspace will be queued until the other training processes are done.

model

When you create a machine learning model, you select the document sets that you want to use to train the model and specify the percentage of documents that are to be used as training data, test data, and blind data.

About this task

By exploring the performance metrics, you can identify ways to improve the model's accuracy.

Restriction: Only three machine learning models can be trained at a time per Knowledge Studio instance. If your instance contains multiple workspaces and the number of machine learning models that are being trained in other workspaces totals 3 already, then your request to train the machine learning model in your workspace will be queued until the other training processes are done.

Procedure

To create a machine learning model:

- ① Log in as a Knowledge Studio administrator and select your workspace.
- ② Select **Machine Learning Model > Performance**.
- ③ Verify that all of the document sets have been approved and that all annotation conflicts have been resolved through adjudication. Only documents that have become ground truth through adjudication or approval can be used to train the model.
- ④ Click **Train and evaluate**.
- ⑤ Optional: To specify how you want to allocate documents from your document sets to be used by the system-level training, test, or blind sets, click **Edit settings**.

See [Document set management](#) for help determining which ratios to apply.

- ⑥ Click **Train** to train the model, or click **Train & Evaluate** to train the model, evaluate annotations added by the machine learning model, and analyze the performance statistics.

Important: Training a machine learning model can take several minutes or several hours, depending on the number of human annotations that exist and the total number of words across all documents.

- ⑦ Select the document sets that you want to use for training the model.

Note: The document sets must contain at least 10 annotated documents.

- ⑧ After the model is created, select one of the following actions:

Option	Description
Log	View the log file to see whether any problems occurred.
Details	View the annotation performance statistics, change the document sets that you want to use for training and testing the model, and create snapshot versions of the model artifacts.
Export	If you have a Standard plan or a Premium plan, you can export a <code>ZIP</code> file to your local system that contains the components that are required for the model to run in a machine learning runtime environment, such as Watson Explorer.

Evaluating annotations added by the model

You can compare the ground truth view for annotations added by human annotators to the annotations added by the model.

Procedure

To evaluate the annotations added by the model:

- ① Select **Machine Learning Model > Performance > Train and evaluate**. The Training/Test/Blind Sets page is displayed.
- ② Click **View Ground Truth** for the training set or test set to see the annotations that were added through pre-annotation and by human annotators. The ground truth editor opens. Click to open individual documents and see how the mentions, relations, and coreferenced mentions were annotated.

truth editor opens. Click to open individual documents and see how the mentions, relations, and coreferenced mentions were annotated.

- ③ On the **Performance** page, click **View Decoding Results** to see the annotations that the machine learning model added to documents in the test set. This button is available only after you evaluate the model. By viewing results, you can see how well the machine learning model labeled mentions, relations, and coreferenced mentions in the test data.
- ④ If you want to change how the documents are divided between training, test, and blind data sets, click **Performance > Train and evaluate > Edit Settings**. For example, if initial results seem acceptable, you might want to increase the number of documents in the test set to further test the machine learning model's results. You can change the ratio for how documents are automatically divided for different purposes, or you can select specific document sets to use as training data, test data, and blind data.
- ⑤ If you made any changes, click **Train & Evaluate** to retrain the model and re-evaluate the annotations.

Deleting a machine learning model

You cannot delete a machine learning model.

You can delete the workspace that was used to develop the model, but you cannot delete the model itself. Deleting a model is not the best approach. Instead, update or replace the artifacts that are used to train the model. Even if the model is not producing the results you expect, you can continue to refine it. Each time you create a new version, the model is built anew. You can edit artifacts like dictionaries and the type system, and choose to use different annotation sets when you train the next version.