

Webinar: Trainable Classifiers

Sign up for the MIP Preview Program: <https://aka.ms/MIP-Preview>

Follow us on twitter: twitter.com/MIPnews

FAQ developed from Q&A session: March 2020

Q: Where can I read more about classifiers?

A: We have three great support documents that trainable classifiers. Get started on trainable classifiers [here](#).

Q: For pre-created classifiers (resumes), is it possible to review the documents which were classified and mark/exclude false positives and train the system better?

A: We are currently working on a feedback capability so that users can identify false positives and our model learns from the feedback.

Q: How can I improve accuracy of Microsoft pre-created classifiers? More specifically resumes. I receive a lot of false positives with the sets I'm currently testing.

A: We are working on a feedback capability that will allow customers to report true and false positives to improve all classifiers (both pre-created and custom). As part of bringing a feedback capability, we are also working on allowing users to customize pre-created classifiers.

Q: What is the difference between seeding data and testing data?

A: Seeding data should all be positive samples of the category and for trainable classifiers, are provided as the first step. Testing data is provided afterwards and should contain a reasonable ratio (50:50, 25:75, etc.) of both positive and negative data.

Q: Must the data be on SharePoint or is OneDrive for business possible too?

A: Seed data must be supplied through SharePoint, but test data may come from either SharePoint or OneDrive.

Q: What happens if we have less than 50 example files?

A: We recommend at least a minimum of 50 positive samples at the beginning for seeding for best results, but you can still create a classifier and see how it performs on test data with fewer than 50.

Q: What happens when there is a collision between the trainable classifier and the sensitivity labels?

A: Sensitivity labels will act based on whether documents are detected as a match for a trainable classifier. To understand label priority between sensitivity labels, please refer [here](#).

Q: What specifically needs to be enabled before the custom trainable classifiers can be used? Where exactly does scanning process needs to be enabled by Global Admin? The message I receive from system is: 'To create your own trainable classifier, we first need to scan your content locations to generate analytics that will help us learn what type of content is in your organization. Only global admins can start the scanning process.'

A: Yes – this is the opt-in experience and it will show on the “Trainable Classifiers” pivot in “Data Classification”. To create custom classifiers, we will do a scanning of the tenant’s SharePoint data to generate analytics required to build custom models. You will need to be a Global Admin in order to start the scanning process.

Q: Why does it take one week of deployment time to Publish? Is there a way to reduce this time so that the feature deploys quicker?

A: To clarify, after you publish, you can use the trainable classifier immediately in sensitivity labels, retention labels, and Communication Compliance policies. There may be some time before those policies take effect and that is what the one-week accounts for. That timeframe will vary depending on the solution.

Q: Will there be any improvement to reviews? If you have a large file set, it takes a long time to go through. Also, you would need to review multiple times to improve the rating.

A: We are currently not considering improving the review process, but we understand if review may take time with a large file set. Our current design involves grouping documents into sets of 30 to review iteration by iteration. The review set helps to ensure that the classifier can make the most improvements with the fewest amount of review iterations.

Q: When initiating the tenant scanning to opt in, is there any expectation of a performance degradation?

A: Regarding SharePoint performance degradation, no you should not experience that.

Q: And this opt-in scan only sans SPO site collections?

A: Yes, only SharePoint site collections.

Q: Is there any tenant service degradation whatsoever? We are a very large enterprise tenant and need to be aware of any negative impacts for anything I do as a global/tenant admin.

A: We only scan SPO, there will be no degradation.

Q: Trainable classifiers, from the docs, apply to retention labels, but not yet with sensitivity labels - is that correct?

A: There is a private preview with sensitivity labels. Let us know if you'd like to be added:

<https://aka.ms/MIP-Preview>

Q: Do these classifiers work with AIP Scanner and Office 365 DLP Policies?

A: *It doesn't work with AIP yet, but we are working towards that. There is no ETA currently. You can use trainable classifiers through a sensitivity label condition in Office 365 DLP policies.*

Q: Tenant Admin is required to enable the feature but once enabled, what Roles are needed to create the classifiers?

A: *The compliance admin role (not role group) will be needed to create classifiers.*

Q: What licensing is needed to use trainable classifiers and to which users does it apply (i.e. individual that creates the classifier, owner of the document, etc.)?

A: *You will need an M365 E5 license for your organization's admins to use trainable classifiers. As for which users it applies to, it will depend on how the sensitivity, retention, or Communication Compliance policies are set up. For instance, if you create an auto-apply retention label for resumes in all Exchange accounts, all users with an E5 mailbox will be affected. If you specify that all SharePoint sites should be covered in the retention label policy as well, all users with access to those sites will be impacted by the policy.*

Q: How are false positives handled?

A: *We are working on a feedback capability so that users can identify false positives and our model learns from the feedback.*

Q: Can you specify a confidence level?

A: *We are looking into providing this capability for our customers, but it does not exist today.*

Q: Does this work with an Office 365 E3 + AIP P2 license?

A: *Unfortunately, no. Trainable classifiers will only be available through an M365 E5 license.*