|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | | |
|  |  |  |  | |
| Microsoft Viva:  Syntex Document Understanding Model |  | Syntex Document Understanding |  |  |
|  |  |  |  |
|  |  |  | Virtual Lab |
|  |  |  |  | |

## Module Overview

This lab creates a Syntex Document Understanding Model that extracts metadata from Web Contracts.

The video contained in this blog provides an overview of what you are going to build in this lab:

<http://myteams.blog/2021/07/12/building-a-document-understand-model/>

# Lab Guide

### Estimated time to complete this lab

45 minutes

### Objectives

After completing this lab, you will be able to:

* Build a basic Document Understanding model
* Review the option for applying Retention and/or Sensitivity labels

### Prerequisites

* Microsoft 365 Viva Tenant from: [Modern Workplace Digital Experiences (microsoft.com)](https://cdx.transform.microsoft.com/)
* Document Understanding Training Files from:

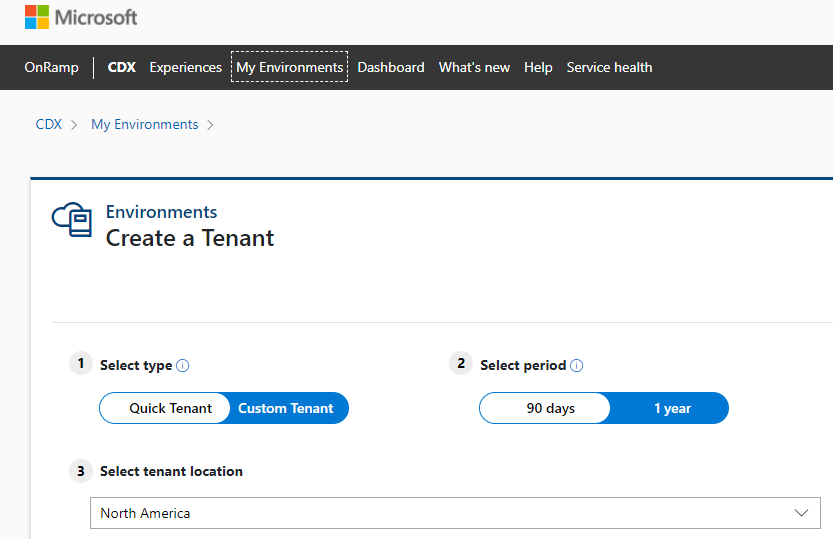
<https://github.com/karlkocar/SyntexDocUnderstanding-Training>

* Document Understanding Testing Files from:

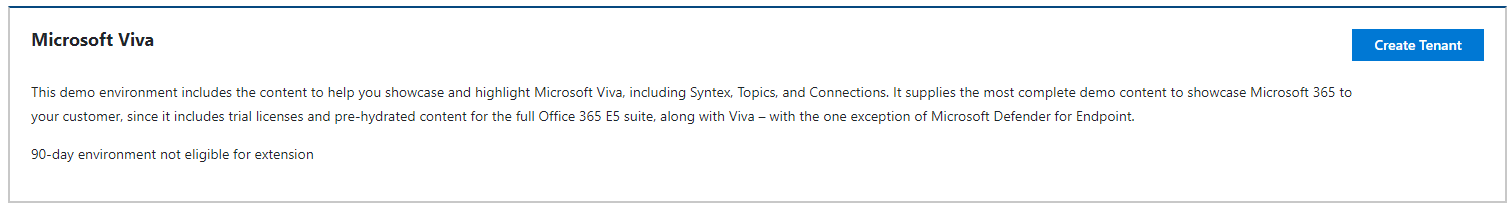
<https://github.com/karlkocar/SyntexDocUnderstanding-Testing>

**Note:** Be sure to download the Training and Testing files onto your laptop, you will need these later.

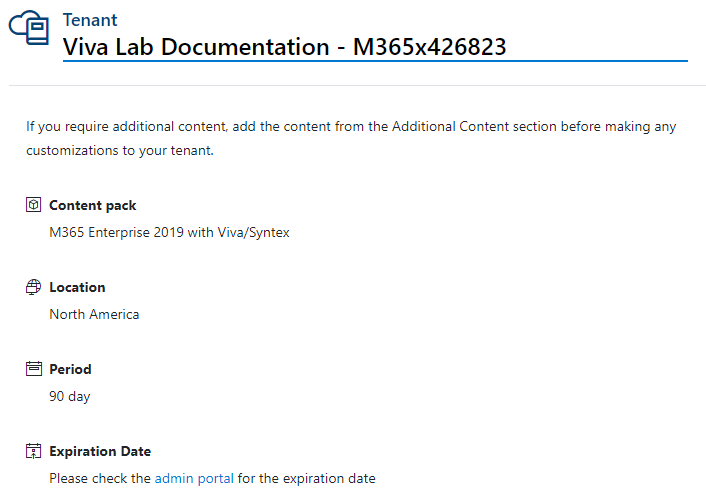
1) Create a 90-day Quick Tenant located in North America.



2) Viva Tenants can be found at the bottom of tenant list.



3) When the lab has been created you should see a screen similar to the one shown below:



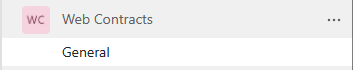
### Model Creation

1) Logon on to your Viva tenant as Megan Bowen (meganb) - <https://portal.office.com>

For example, use - [meganb@M365xXXXXXX.onmicrosoft.com](mailto:meganb@M365xXXXXXX.onmicrosoft.com) (replace the XXXXXX with your tenant number), use the password that was provided in the CDX tenant window. All users have the same password.

Don’t forget to use a private/incognito browser session.

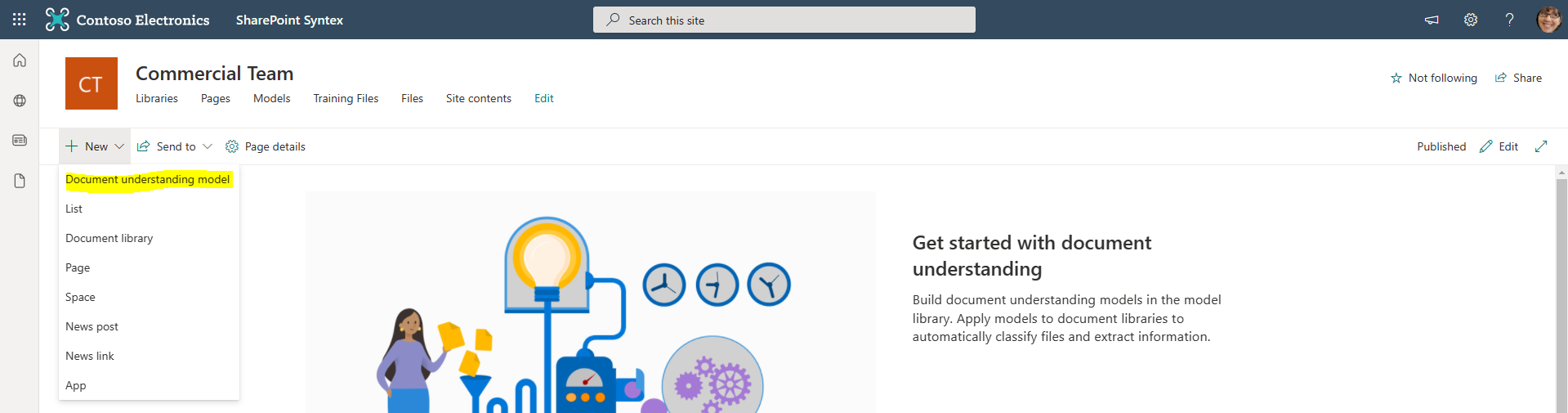
2) Click on the Microsoft Teams icon and create a Microsoft Team called “Web Contracts” we are going to use this to upload our testing files later in the lab.



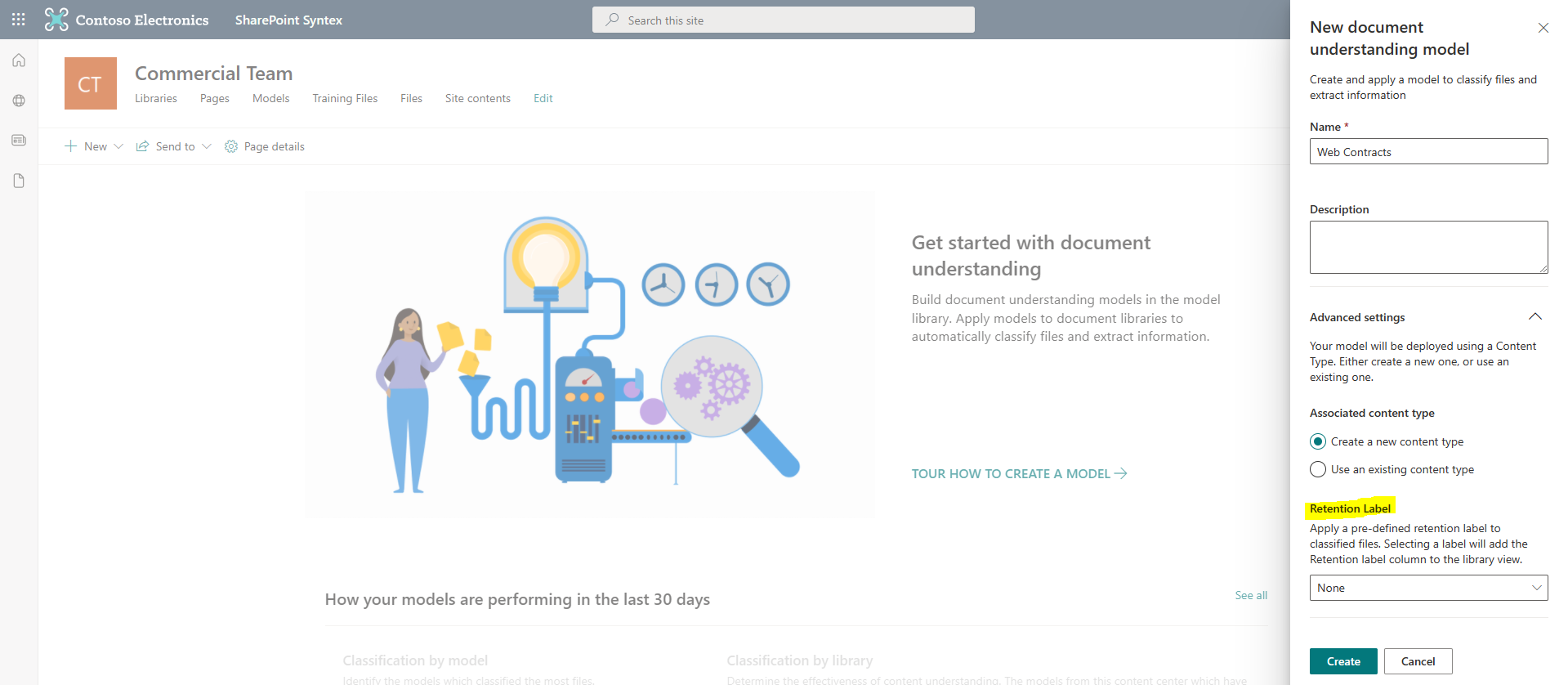
3) Next, navigate to the SharePoint admin portal (accessed through the Office Admin portal) and create a new Content center and name it Commercial Team. Make Megan the primary administrator. Change the Time zone under Advanced settings. Press Finish to create the site.



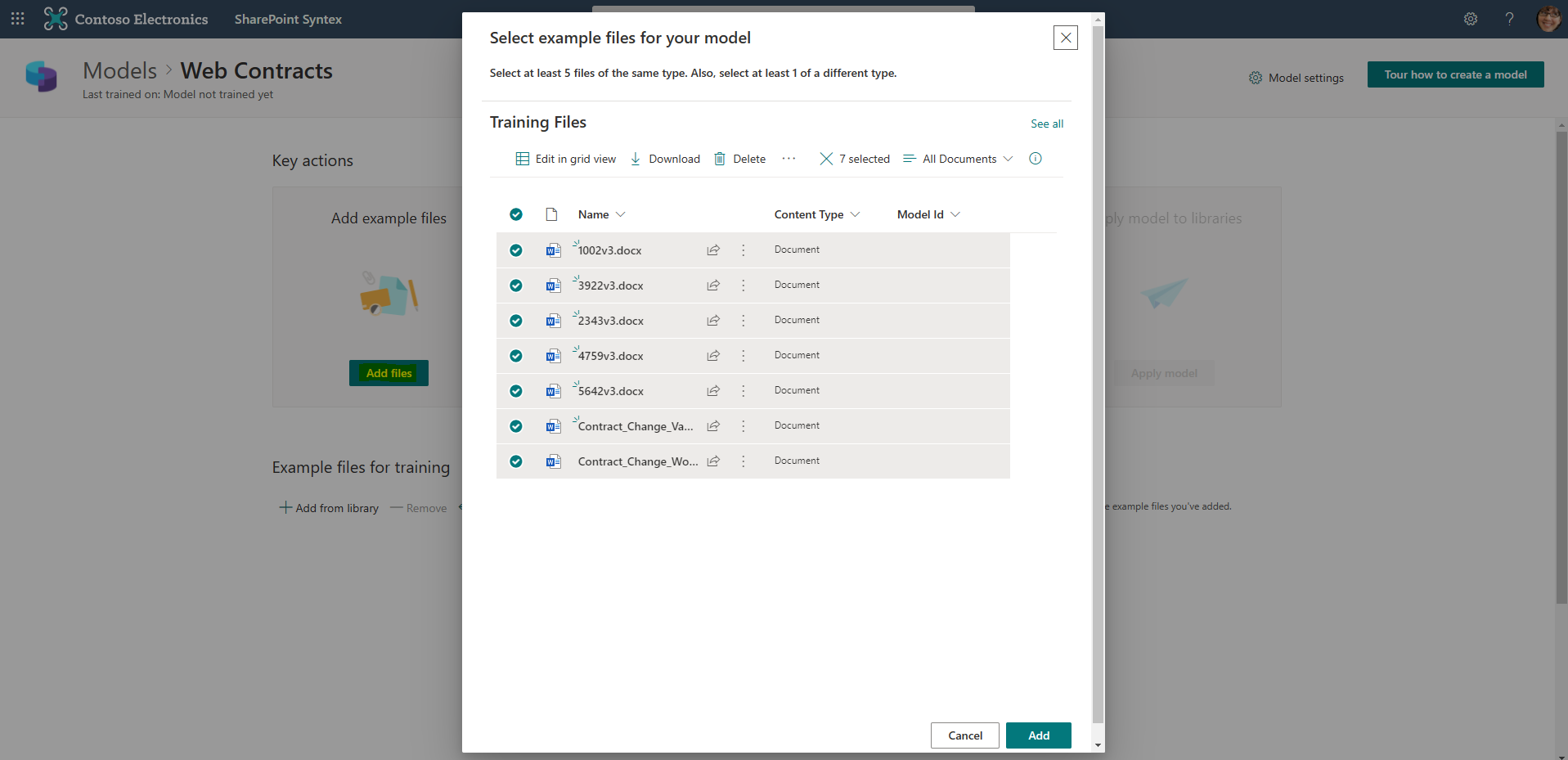
4) Go to the site when it has been built. Feel free to take the tour, which will run through how to create a model. When you are ready use the New button to start building your model.



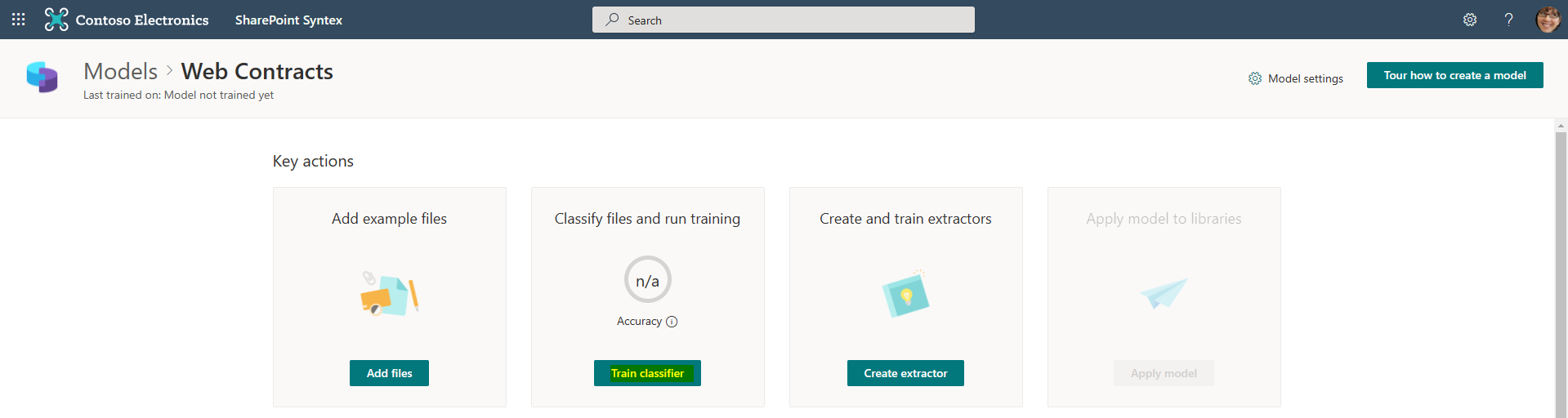
5) Call your model Web Contracts. Also look at the Advanced Settings. Don’t make any changes but note that you could add a Retention policy to classified files. Create the model and move onto the next step.



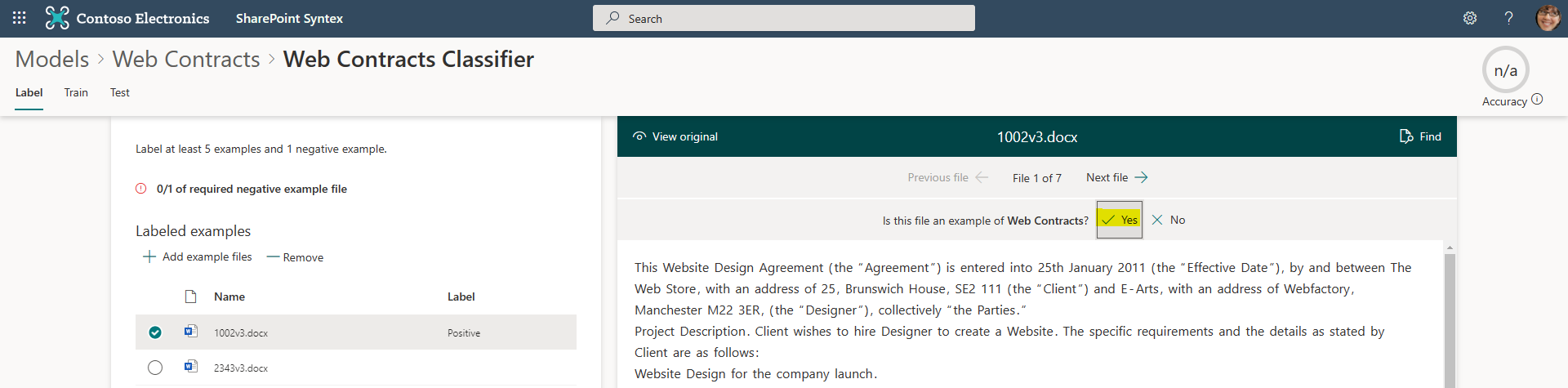
6) You will likely receive a pop up inviting you to read through some guidance about how to “Train a model to classify a type of file”. After you have clicked through the pop up you will be taken to the Web Contracts wizard, where you will be presented with the Key actions you need to perform to create your model. The first step is to upload the Document Understanding Training files you previously downloaded during the prerequisites part of this lab. You should add all the files, five of which are examples of the web contracts you want to classify and two are not. Select the Add button when you have uploaded the required files.

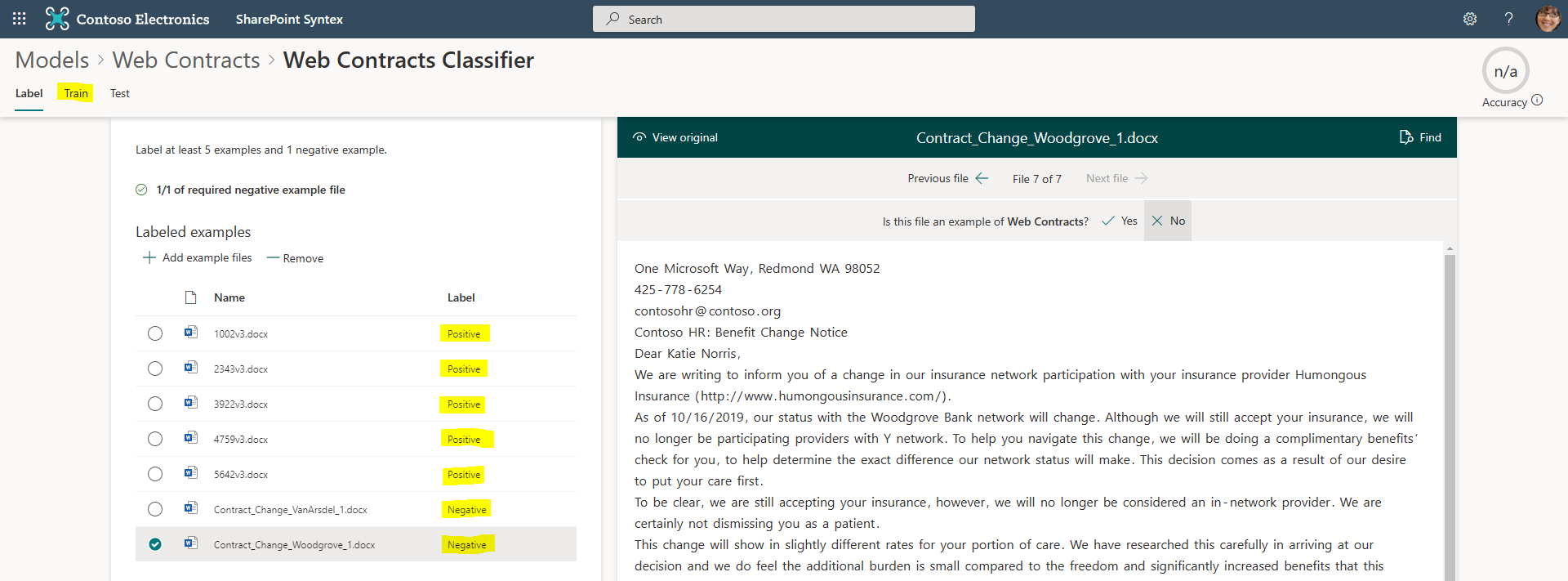


7) The next step is to run through the Train classifier process.

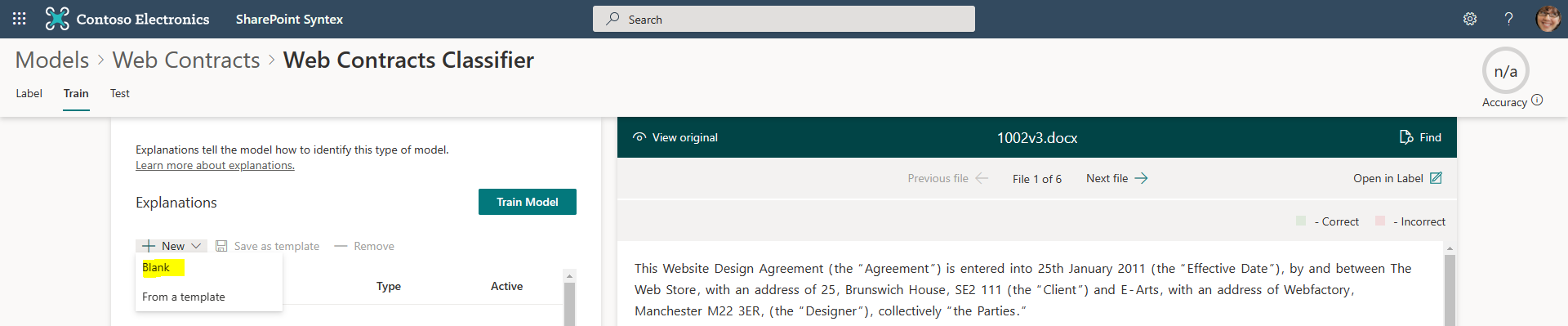


8) For each of our document we need confirm that the document is indeed an example of a Web Contact. Run through the list of documents and click the Yes button for the web contract and the No button when you reach the contract change documents. Use the two screenshots below for guidance. At the end of the procedure, you will have classified all the training document as either Positive or Negative. Then press the Train button at the top left-hand side of the portal.

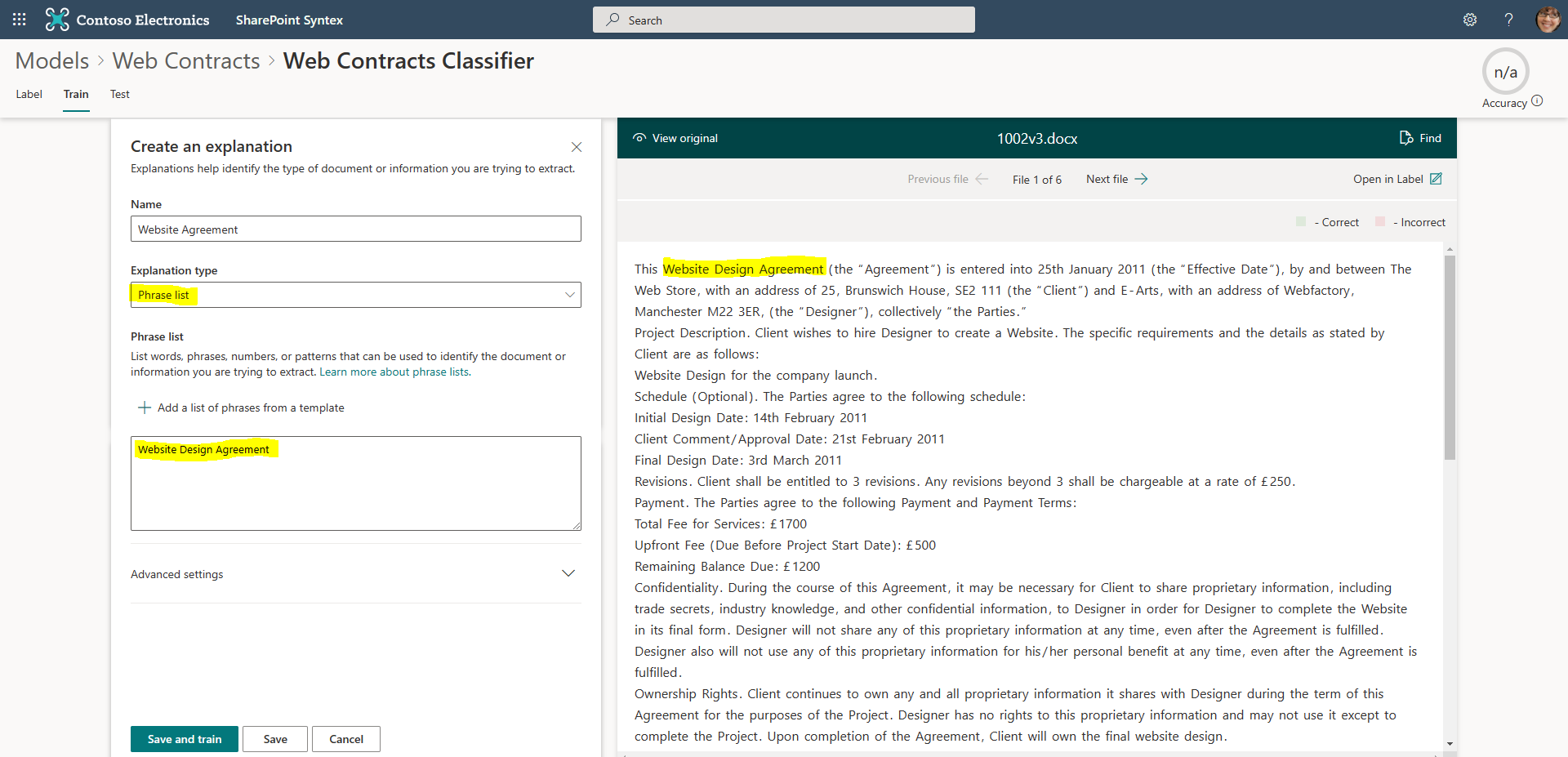




9) We now need to add an explanation that tells the system how to identify a web contract document. Feel free to use the “Learn more about explanations” link to read through the documentation. Add a new Blank explanation.

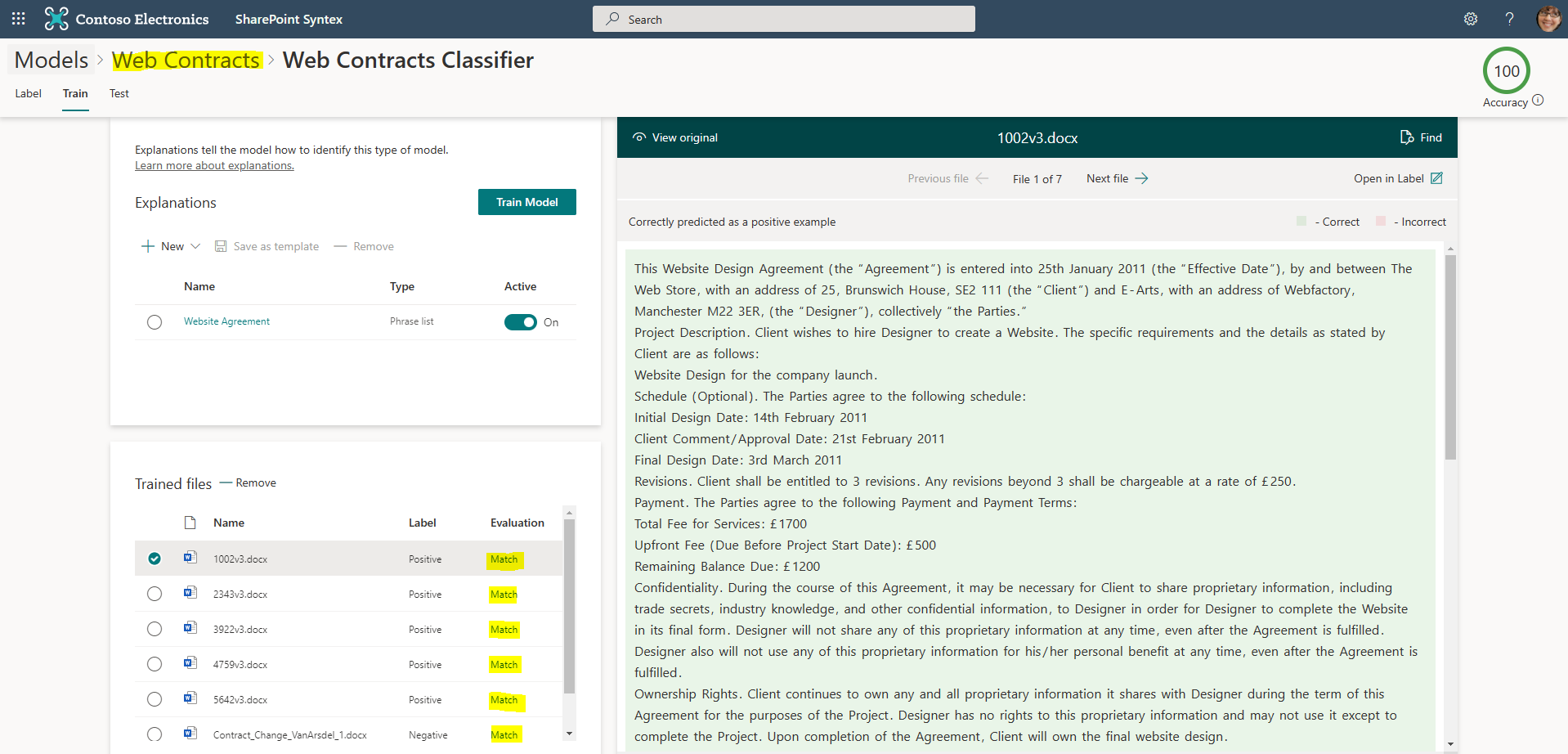


10) Give the explanation a name. Select Phrase List as the Explanation type. Then add the Phrase “Website Design Agreement”, which corresponds to the passage of text at the beginning of the web contract document shown on the right-hand side of the portal. In our lab example this is going to be an adequate explanation, in a production implementation you might need to have additional or more sophisticated explanations. Leave the Advanced setting at default. Save and train the model.

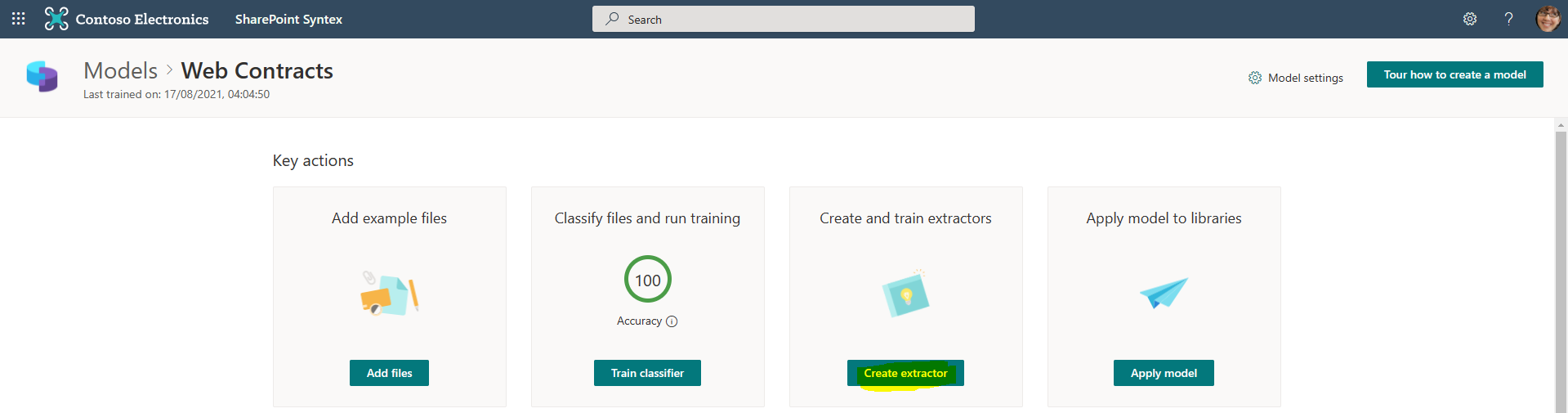


11) When the training has completed you will see the following screen. You should see a Match under the Evaluation column for each of the documents. If you don’t you will need to go back and provide a better explanation. In the lab this should not be necessary.

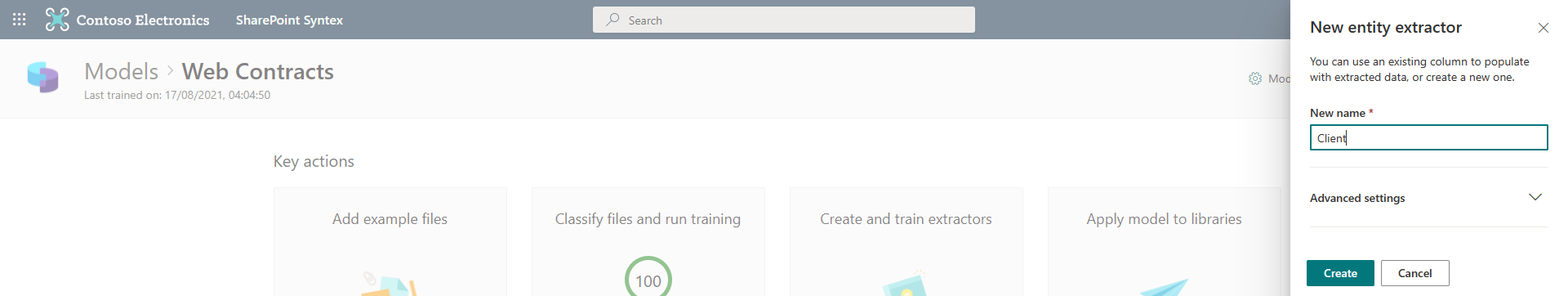
Now go back to the Web Contracts model homepage. The easiest way to do this is to click on Web Contracts at the top of the page: Models > **Web Contracts** > Web Contracts Classifier



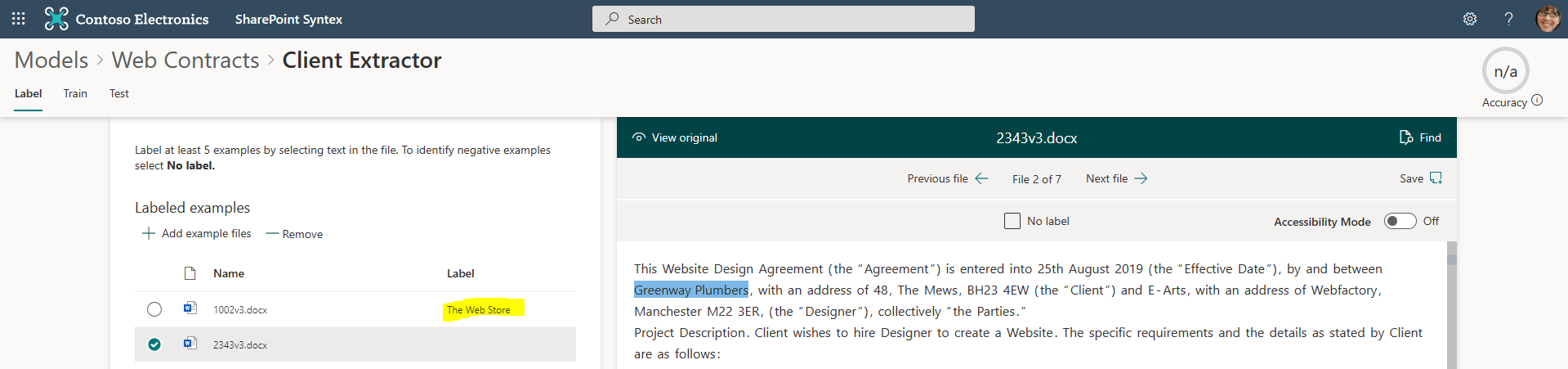
12) Now that we have trained our model to recognize web contracts, we are going to create extractors that will identify and pull-out useful metadata. From the homepage select the Create extractor button.



13) Call the extractor “Client” and create it. Leave the Advanced setting at default.

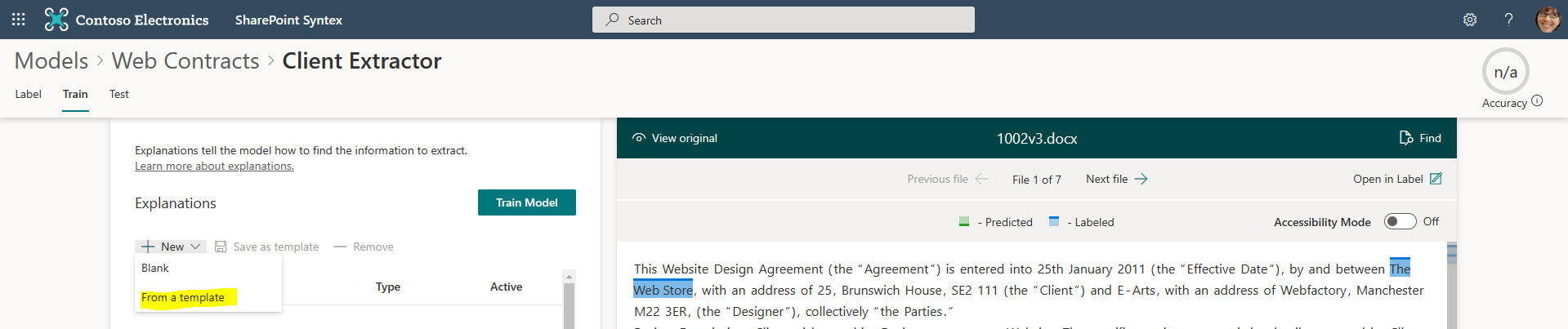


14) For each web contract you now need to highlight (with your mouse) each client name. In my lab example it’s “The Web Store”, for the second document it’s “Greenway Plumbers”.

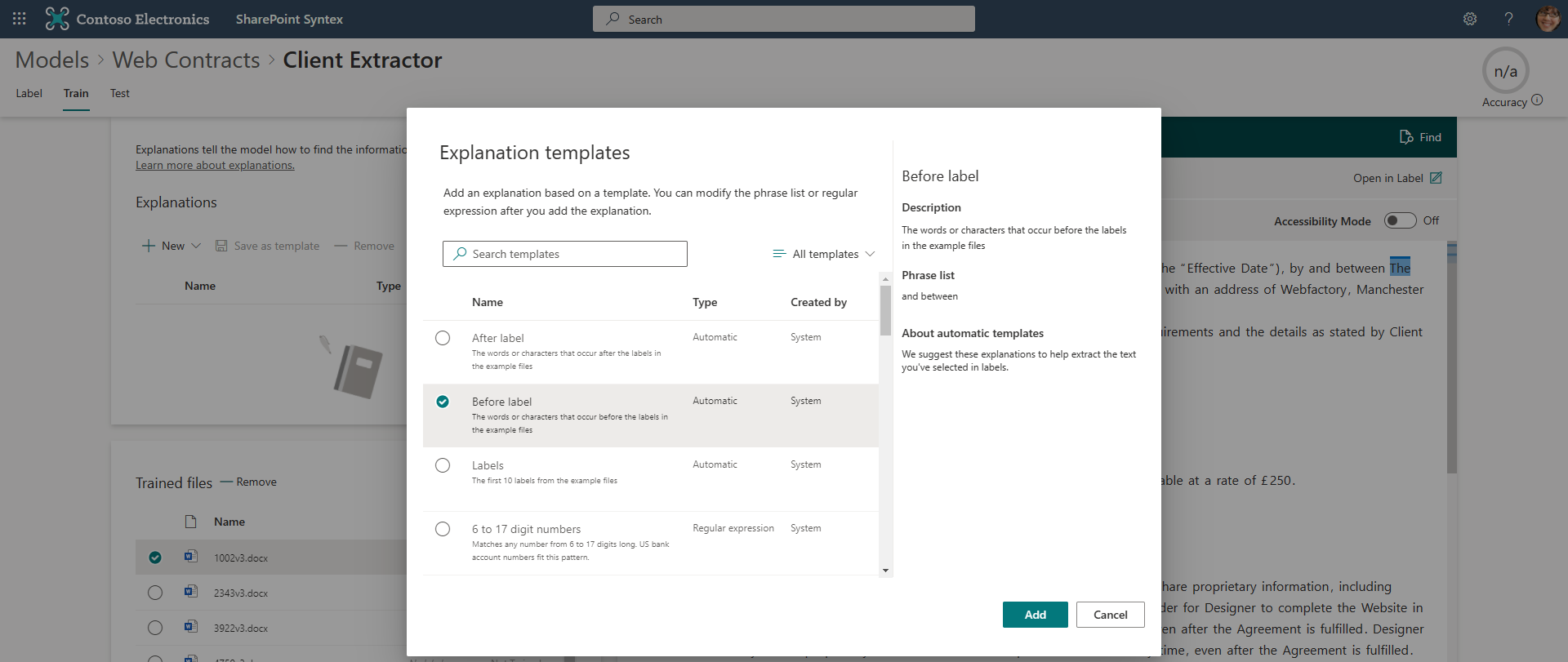


15) You continue to do this for all the web contracts. When you reach the contract change doc you mark them as having no label. After you have clicked the “No label” box for the last document, then press the save button on the top right-hand side of the Contract Change document window.

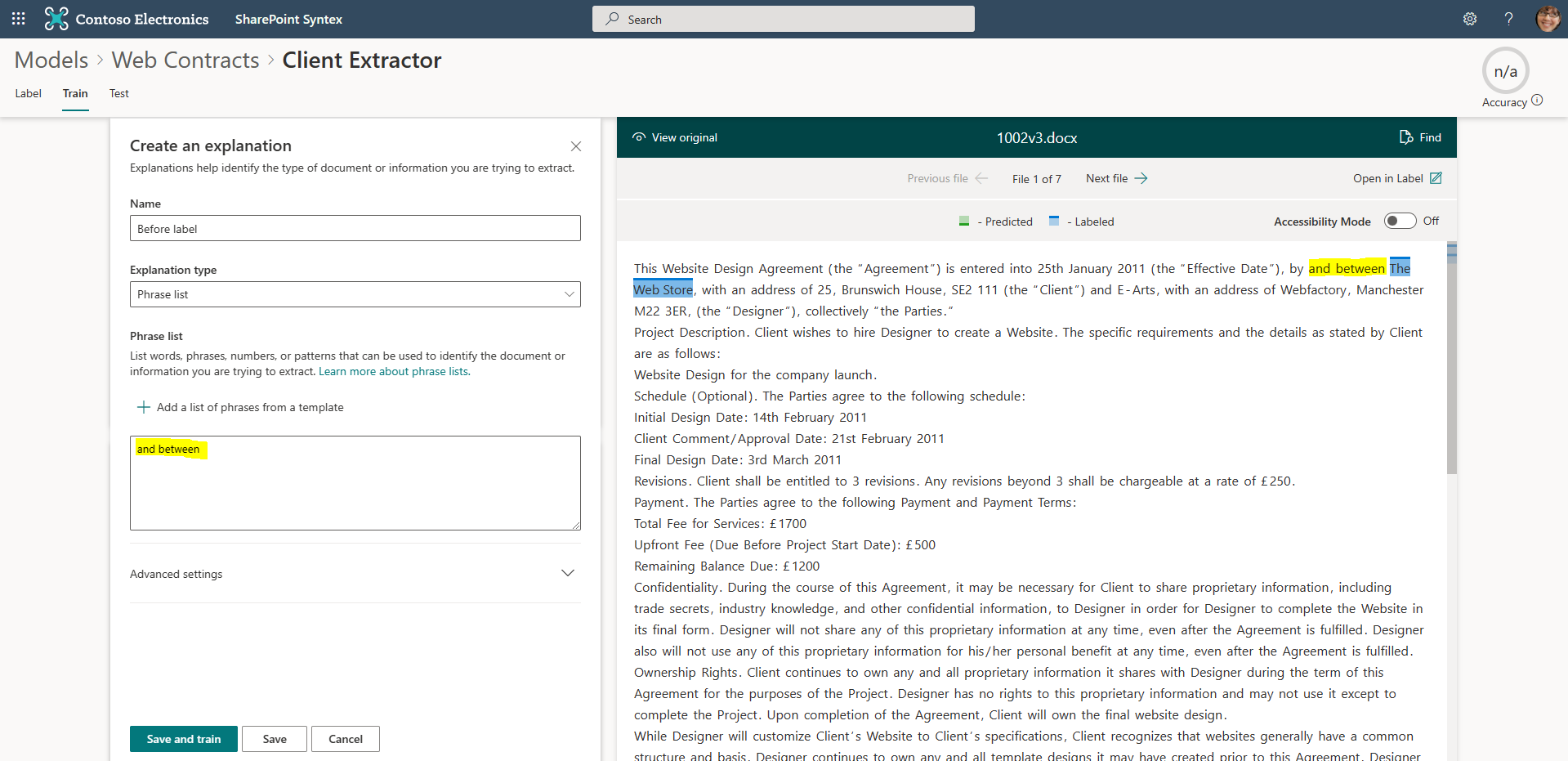
16) Then press the Train button on the top left-hand side of the screen. We are now prompted to add explanations about how to find the Client name in the web contracts. Create a new explanation and choose “From a template”.



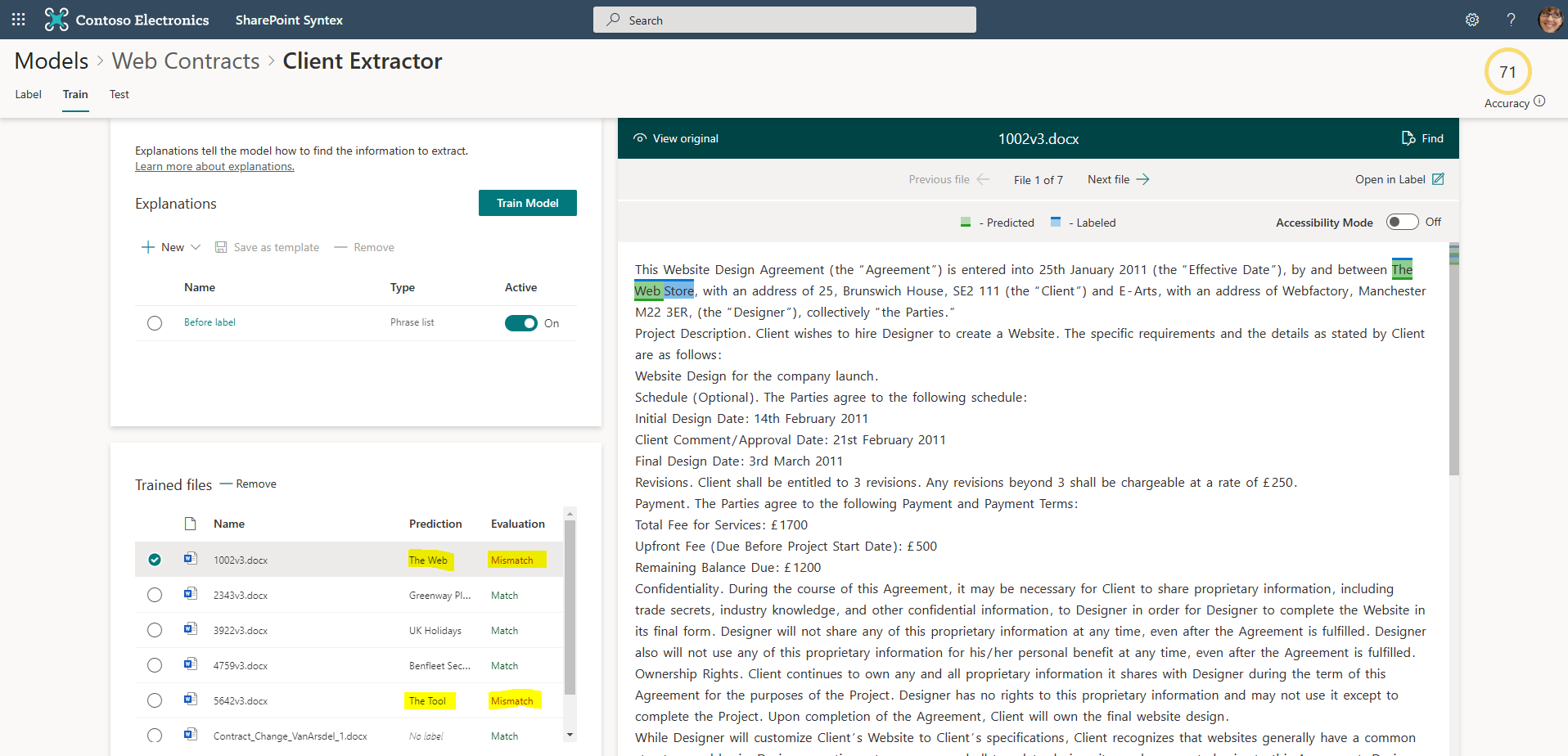
17) Select the Before label and Add it.



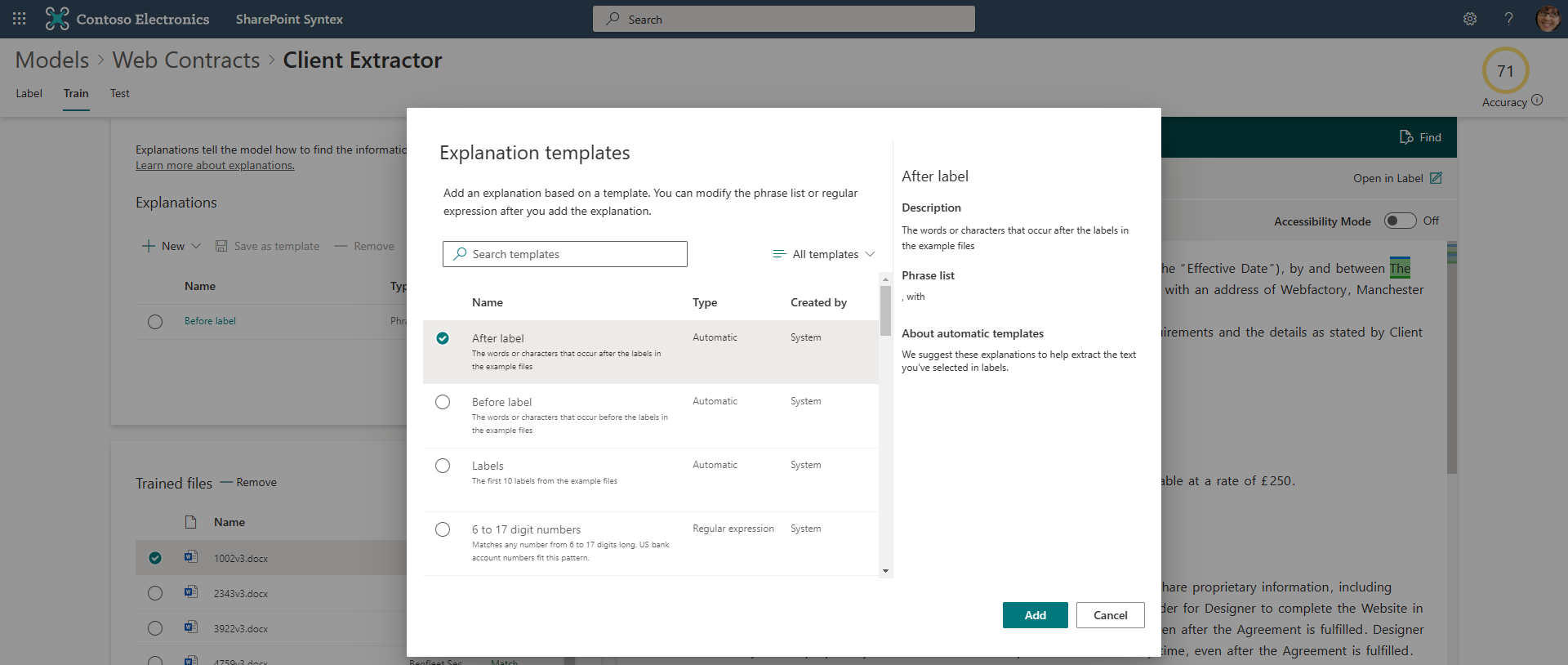
18) You should be able to see that the before template identified the text that appears immediately before the Client name. Note that sometimes you may want to edit this to make it unique as the AI is not infallible. In our example the “and between” text should be fine. Leave the Advanced settings at their defaults and then Save and train.



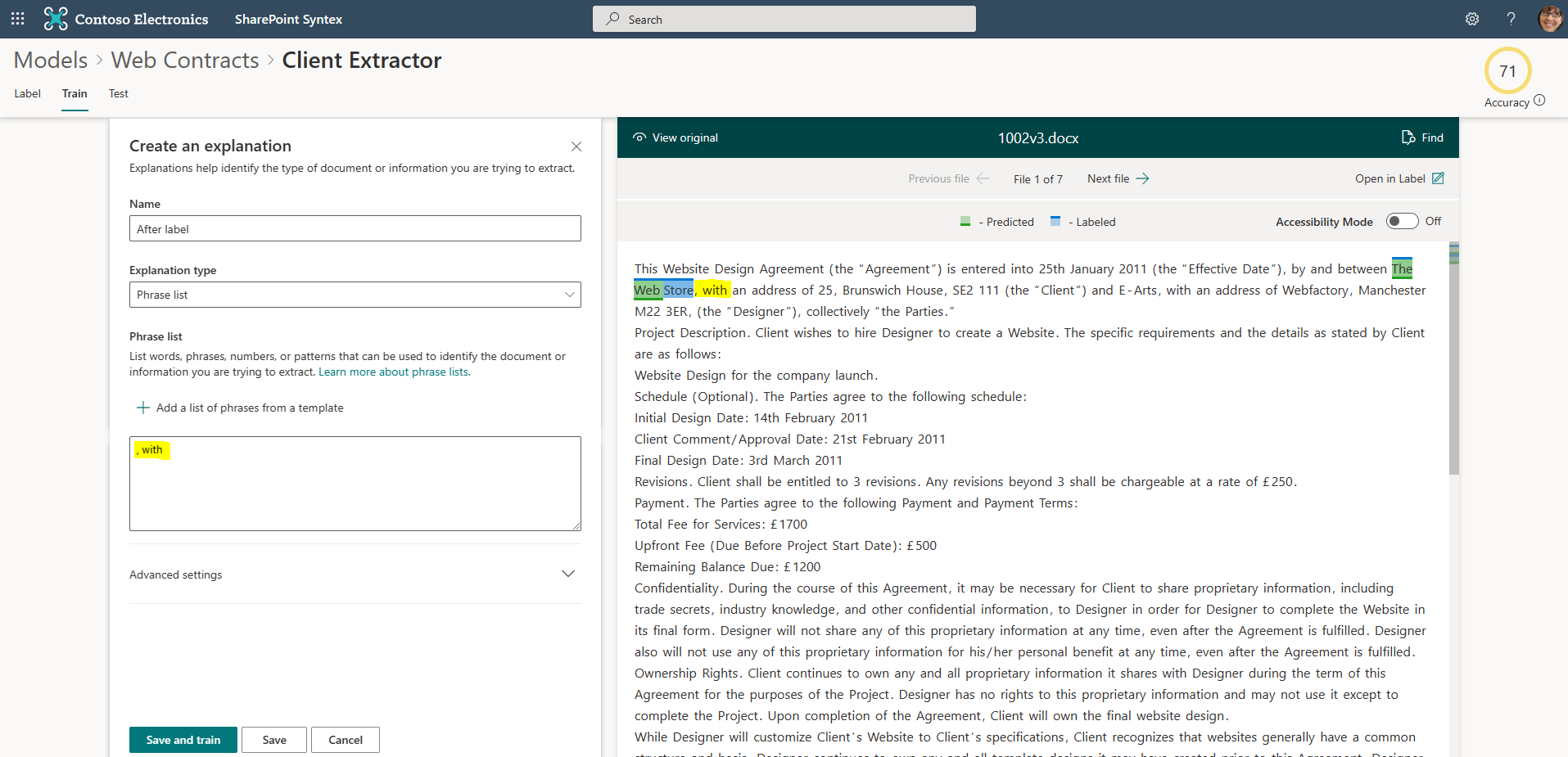
19) You will probably see that the Evaluation column displays one or more mismatches. This means the model needs a better explanation about where to find the client’s name. Let’s add a second Explanation.



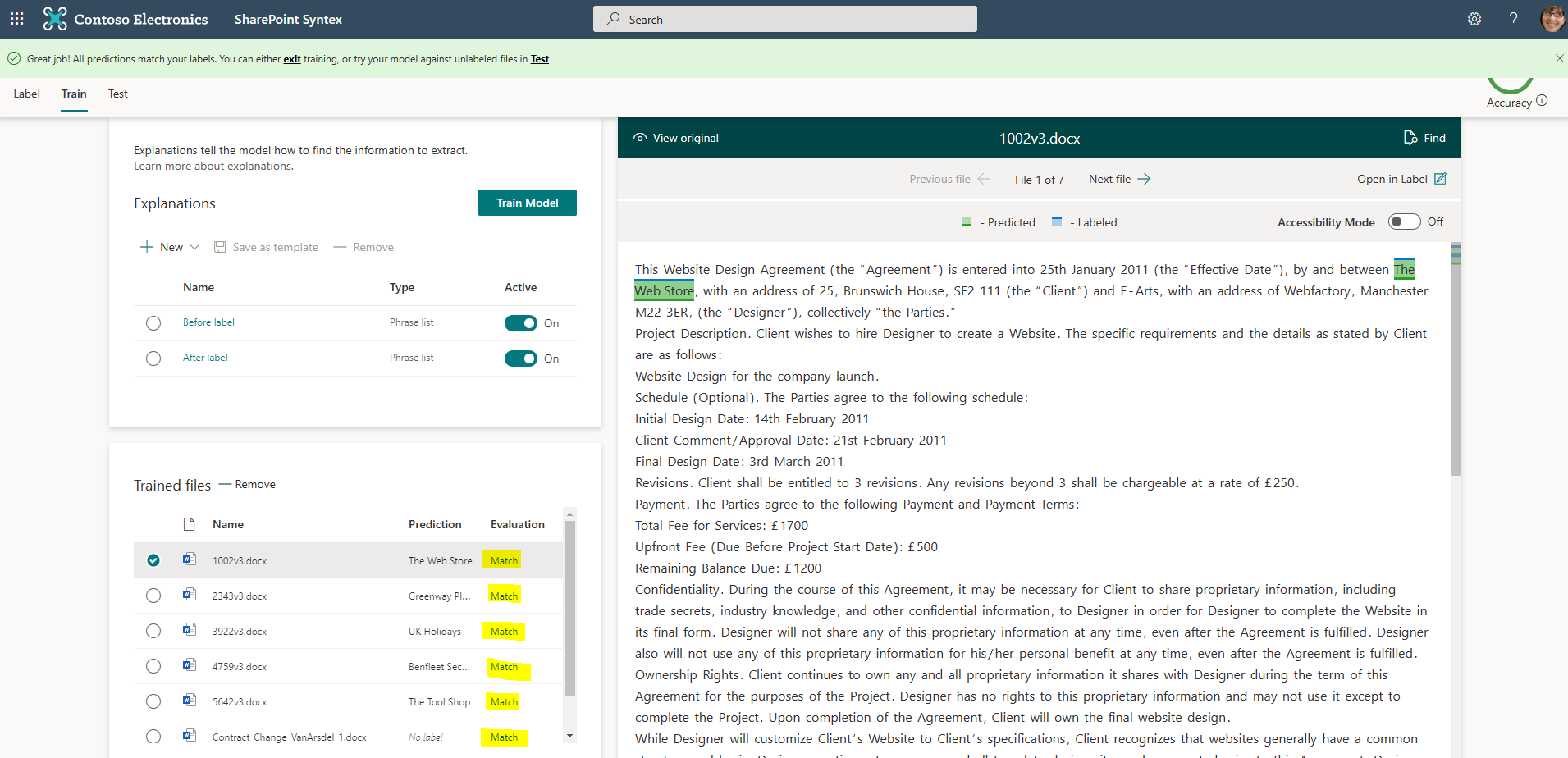
20) Add a new After label template.



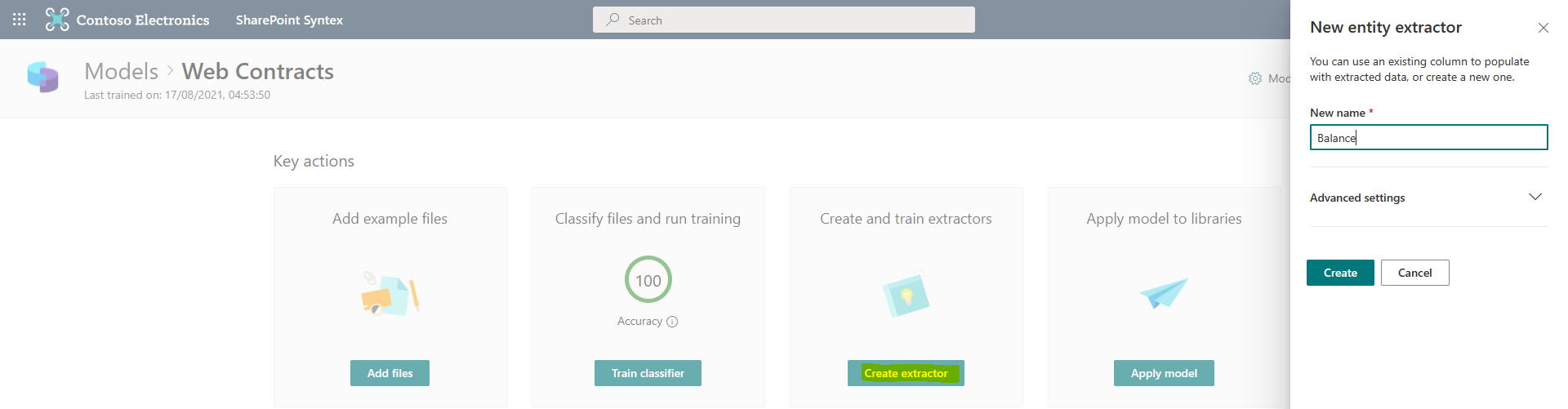
21) You should be able to see that the system has recognized that “,with” proceeds the client’s name. This looks correct so leave the Advanced setting at default and Save and train the model.



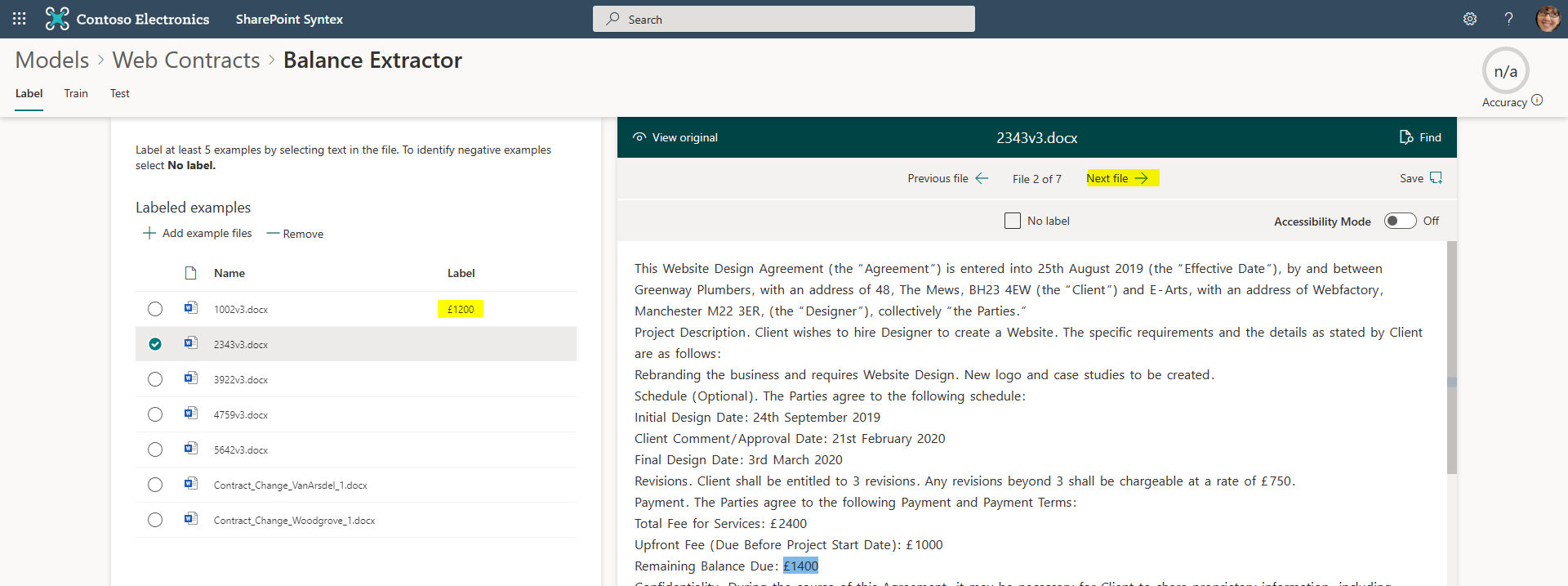
22) This time a full set of matches should be returned, along with the correct client names. This indicates that the model has enough information, and we can move on to create or next Extractor.



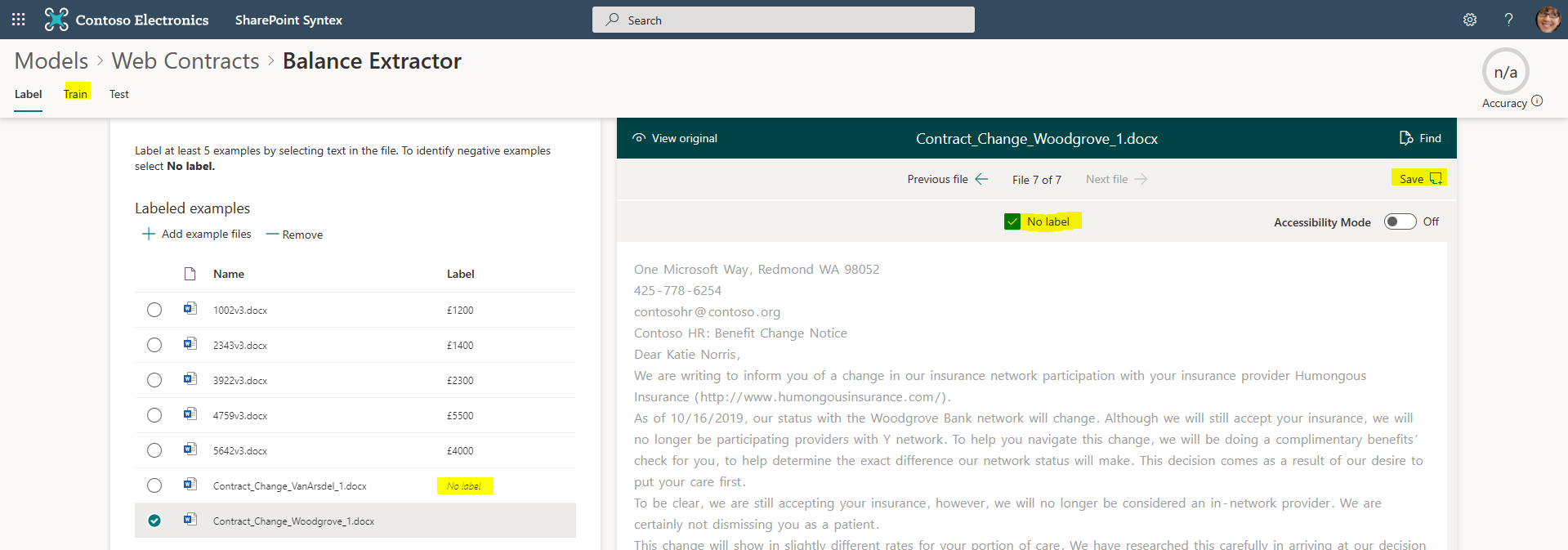
23) Find your way back to the Model’s homepage and create a new Extractor called Balance.



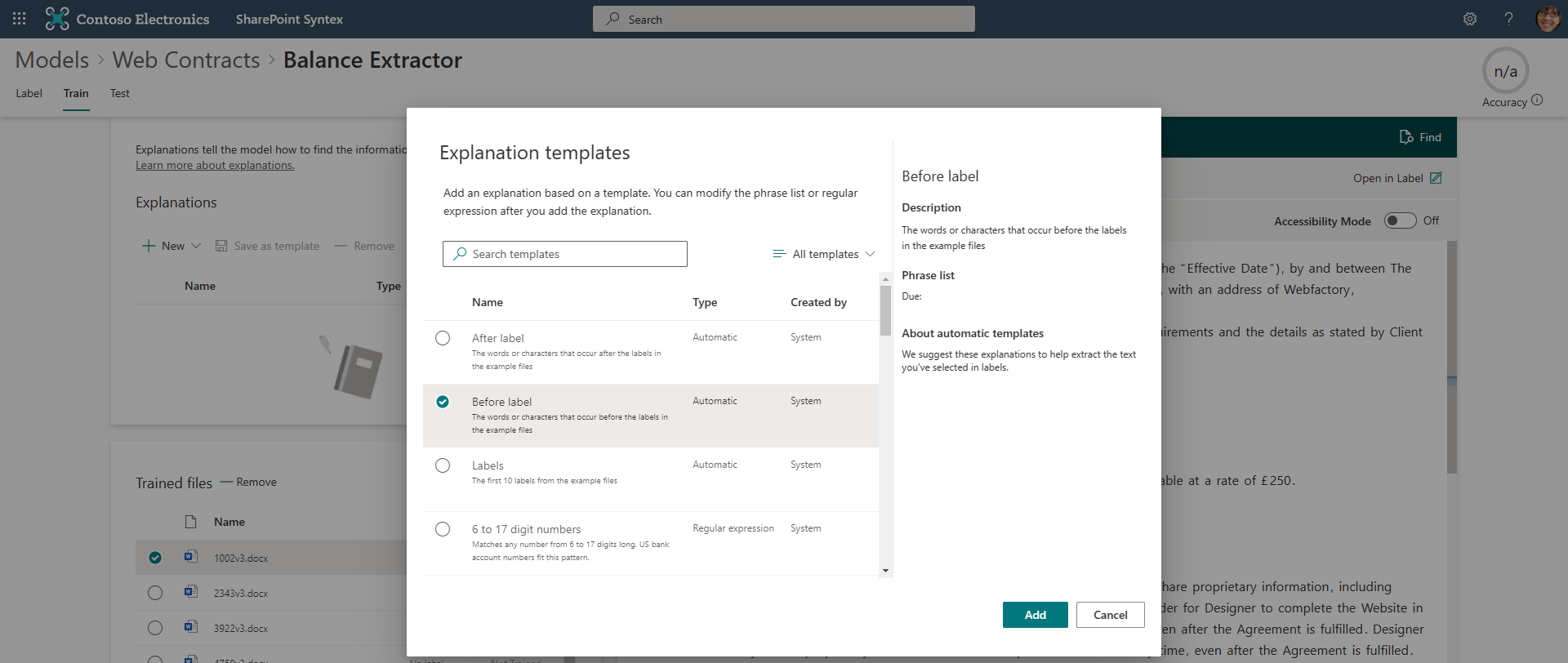
24) In the same way you did for the Client extractor, for each web contract highlight the “Remaining Balance Due” amount. In the lab example shown the first value is £1200 and the second one is £1400. After you have highlighted the value move onto the Next file.



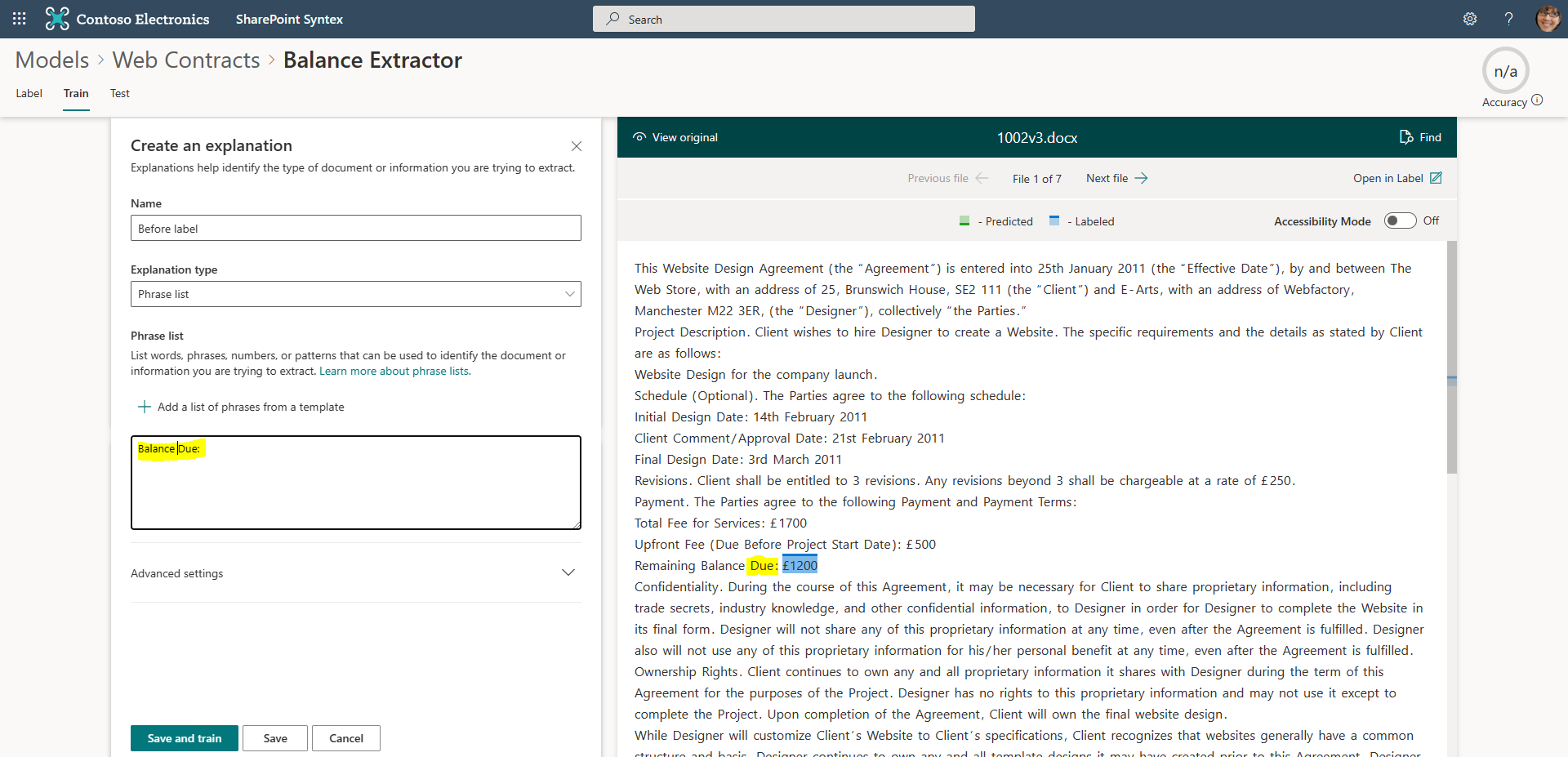
25) When you reach the Contract Change documents select the “No label” field and press the Save button when you have labeled the last document. Then press the Train button (top left-hand side).



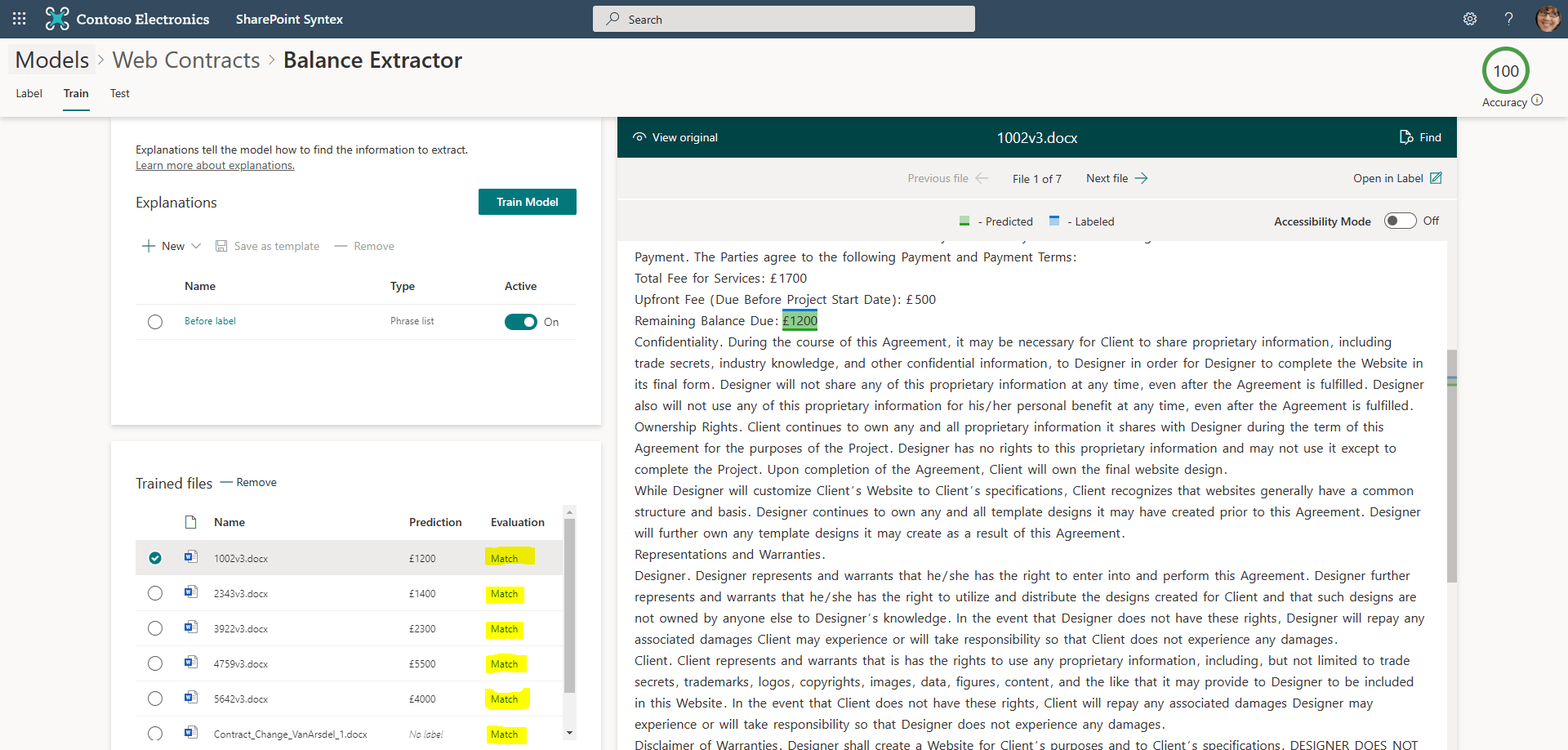
26) As this is a simple model, we are going to use the Before and After Label templates like we did before. Create a new Explanation and select the Before label.



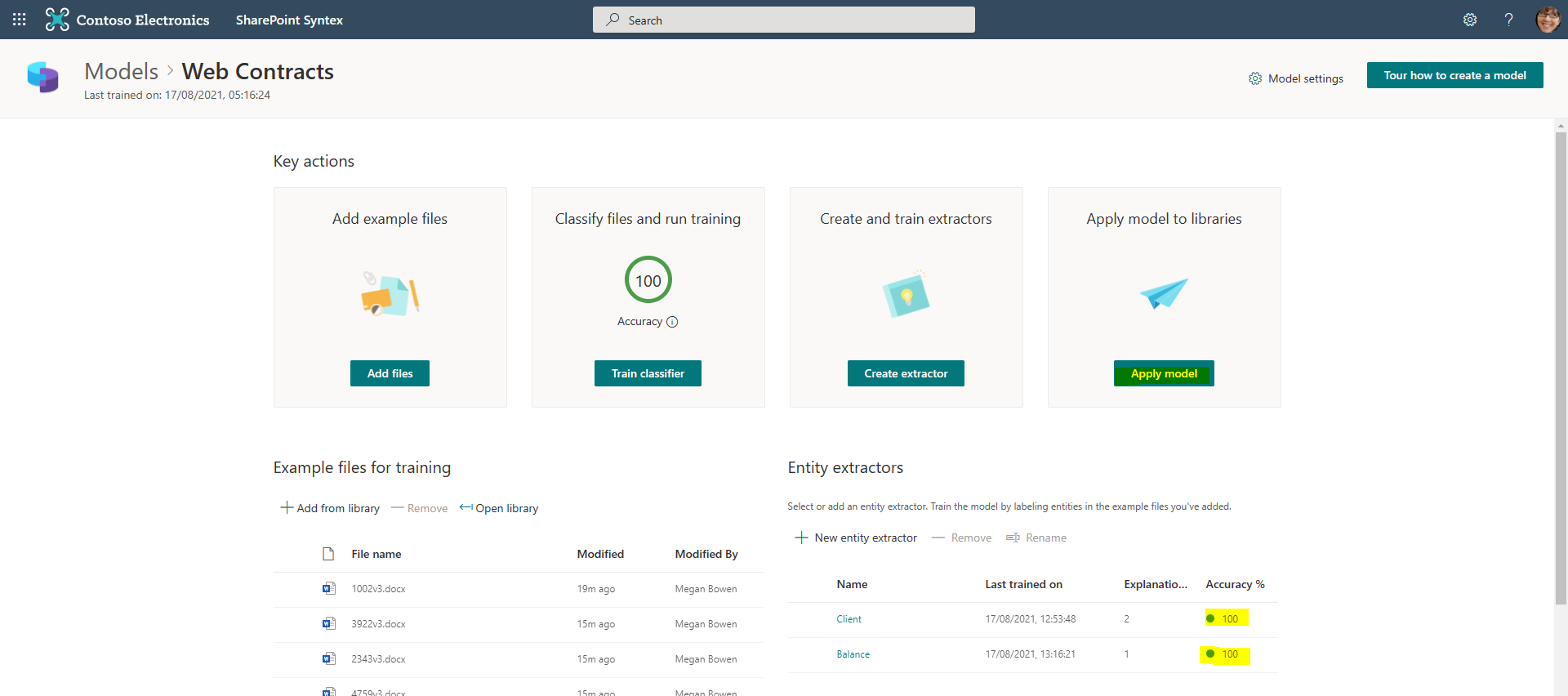
27) you should see that the Before label adds “Due:” in the phrase box list. We are going to add the word “Balance” in front of Due: to demonstrate we can intervene if we believe we can improve on the default explanation. See below. Then Save and Train the model.



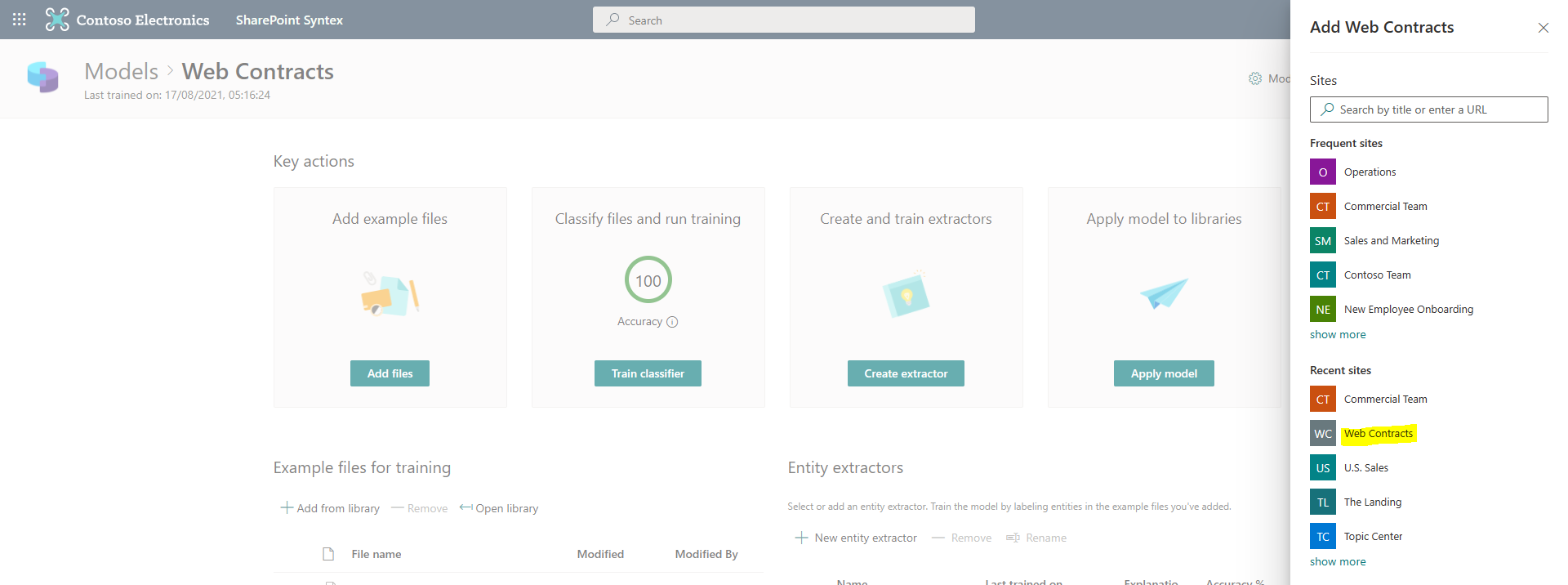
28) If this returns a full suite of matches, we can move on. If we have any mismatches, look to add an After label.



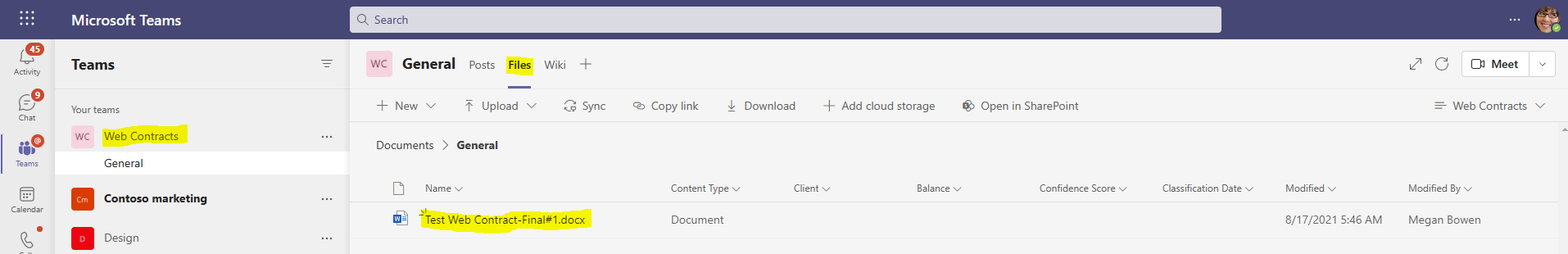
29) If all is well, we can apply our model to a SharePoint library.



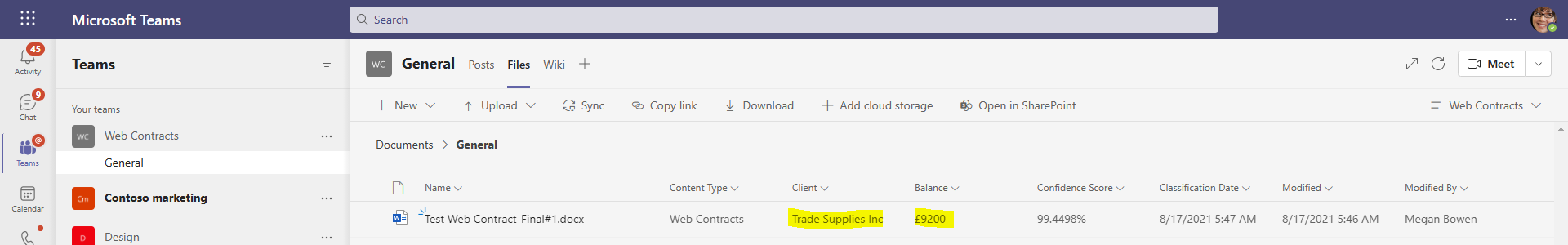
30) In our lab example we are going to add the Microsoft Teams Web Contracts site. Leave the Advanced setting at default.



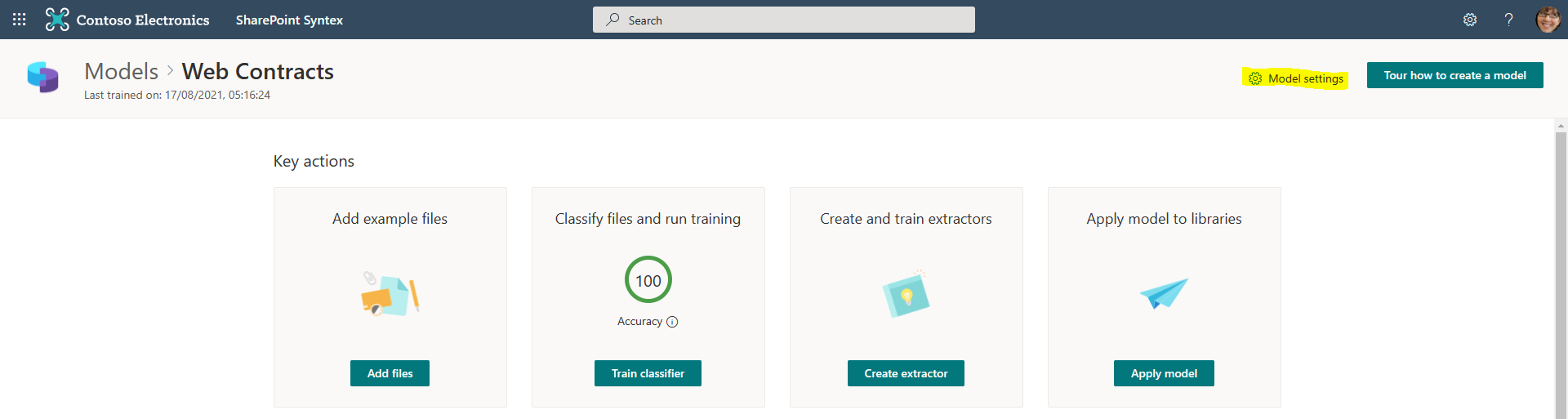
31) Go the Web Contracts Microsoft Team we created previously and go to the Files tab. Then upload one of the test contracts you downloaded in the prerequisites section of this guide.



32) After approximately 2-3 minutes refresh the web page and you will hopefully see the extracted metadata. If the model has completed process, give it some more time.

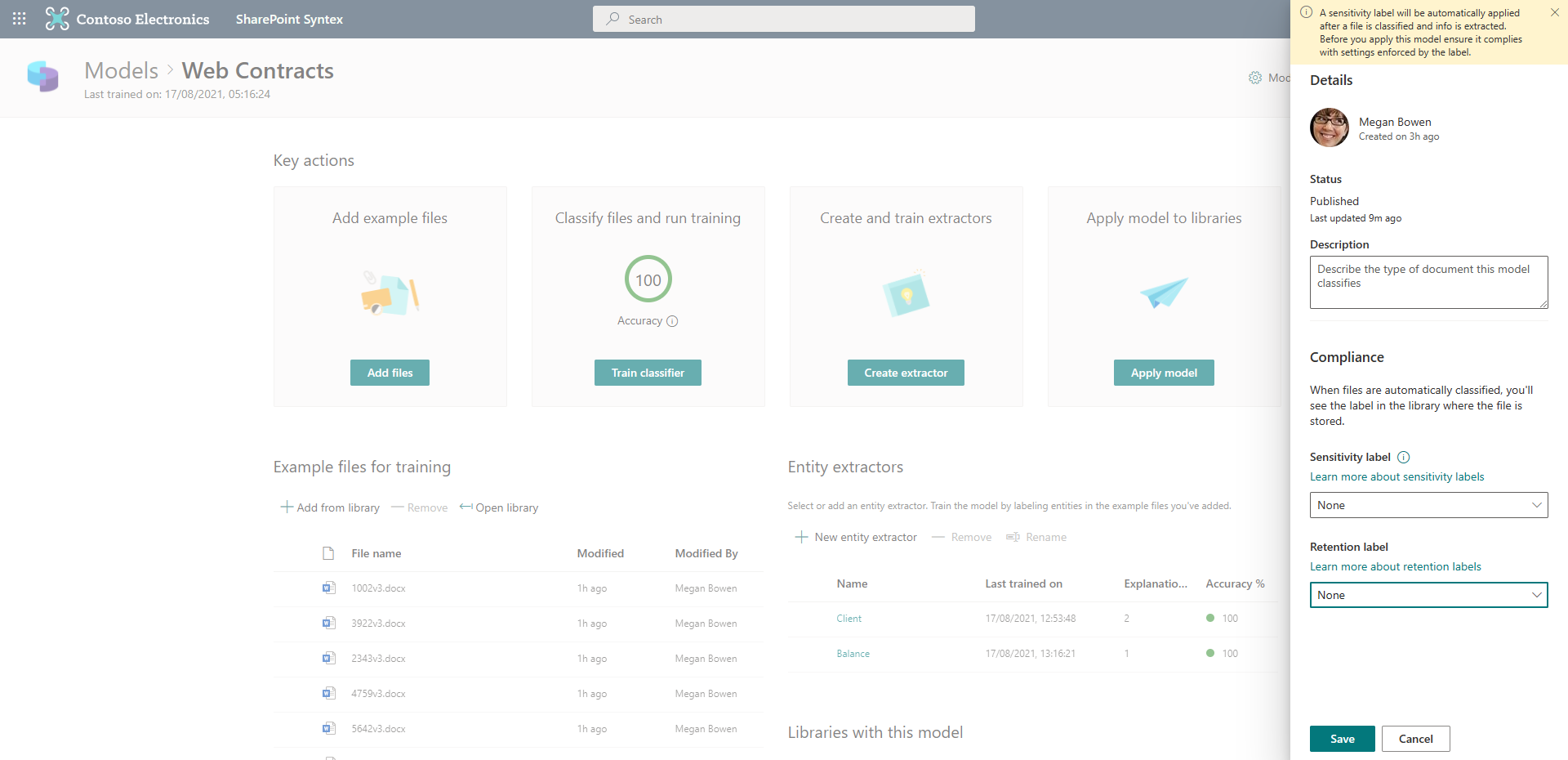


33) Now return to the Web Contracts model homepage and select the model settings.



34) Under Compliance you are provided with an option to add both a Retention and Sensitivity Label.

**Note:** You are welcome to test applying of these labels to your model. The native demo lab Retention policies are assigned correctly. However, if you want to add a Sensitivity label then you will need to create the label from scratch from the Security and Compliance Center. The sensitivity label(s) that are pre-built in the demo environment do not get applied. Feel free to troubleshoot if this tweaks your interest.



Congratulations – you have completed the lab and built a working document understanding model!