Create an AI/ML Model with Power BI Premium

Hackathon Documentation

Contents

[Overview 2](#_Toc105500411)

[Architecture 2](#_Toc105500412)

[Module 1: Modern enterprise AI 3](#_Toc105500413)

[Task 1: Create a Power BI Dataflow 3](#_Toc105500414)

[Task 2: Create an ML model directly in dataflows 7](#_Toc105500415)

[Task 3: Review the Power BI Model Validation Report 9](#_Toc105500416)

[Task 4: Apply the model 11](#_Toc105500417)

# Overview

A few comments on this hackathon

* In this tutorial, you learn how to create a powerful Classification Model without writing a single line of code using automated machine learning functionality provided within Power BI Premium.
* We will use an Accounts Receivable dataset and leverage Power BI premium to build a Classification Model that predicts if a customer will pay his/her account on time (i.e., before or on the invoice due date).
* Automated machine learning functionality within Power BI Premium can be used to automate the time intensive tasks of experimentation and testing ML models. Automated machine learning in Power BI Premium rapidly iterates over many combinations of algorithms and hyperparameters to help you find the best model based on a success metric of your choosing.
* Optionally, we could run a Regression Model and predict when the customer will pay (i.e. the actual number of days it will take a customer will pay).
* Optionally, we could use the Key Influencers visualization in Power BI to …..
* Please share any issues, suggestions, and questions with [alias@microsoft.com](mailto:alias@microsoft.com).

# Architecture

In this scenario, we have an Accounts Receivable dataset that resides in Azure Data Lake and can be queried with Azure Synapse Serverless. At a high level, the architecture of the lab is as follows:

* **Azure Data Lake** - Infinitely scalable azure storage
* **Synapse** - Infinitely scalable compute that can be leveraged in a serverless or dedicated capacity
* **Power BI Premium** - Analytics Platform providing users the ability to create AI/ML models

A picture containing diagram

Description automatically generated

# Module 1: Modern enterprise AI

**Projected time:** 90 minutes

**Product:** Power BI service

**Key message**: Power BI is a powerful platform for processing your data.  
**Persona:** Business analysts and data scientists

### Task 1: Create a [Power BI Dataflow](https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-create)

1. Navigate to <https://github.com/iamjenetzler/pbifinanceml> and download the repo to your computer by choosing **Download ZIP** from the green Code menu. After the folder is downloaded, extract the contents to your Desktop.

Graphical user interface, text, application, email

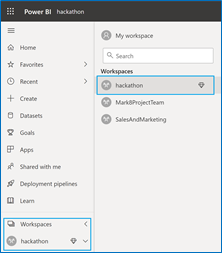
Description automatically generated

1. Login to [OneDrive for Business](https://onedrive.live.com/about/en-us/signin/) using your provided credentials (username/password), chose to **Upload > Files**, and navigate to the file called *WA\_Fn-UseC\_-Accounts-Receivable.csv* in the folder you extracted files to in step 1.

Graphical user interface, text, application, email

Description automatically generated

1. Login to the **Power BI Web Application** at <https://app.powerbi.com> using your provided credentials.
2. Go to the **hackathon workspace.**

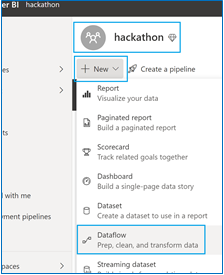


1. Sign up for a Premium per User license by clicking on your profile icon in the top right corner and clicking on “Start Trial”. If the “Start Trial” button is not available, you already have the Premium per User trial and can skip this step.

Graphical user interface, text, application

Description automatically generated

1. Still in your hackathon workspace, create a new dataflow by clicking on the **+ New** button and selecting Dataflow. Click on **Add new tables**.

 Graphical user interface, application

Description automatically generated

1. In the **Choose data source screen**, select Text/CSV.

Graphical user interface, application, Teams

Description automatically generated

1. Click **Browse** and choose **From One Drive Business**.

Graphical user interface, text, application

Description automatically generated

1. **Select** the file called WA\_Fn-UseC\_-Accounts-Receivable.csv. Click **Next**

Graphical user interface, text, application, email

Description automatically generated

1. A preview of the data from the file we just connected to is shown. Click on **Transform data**.

Graphical user interface, table

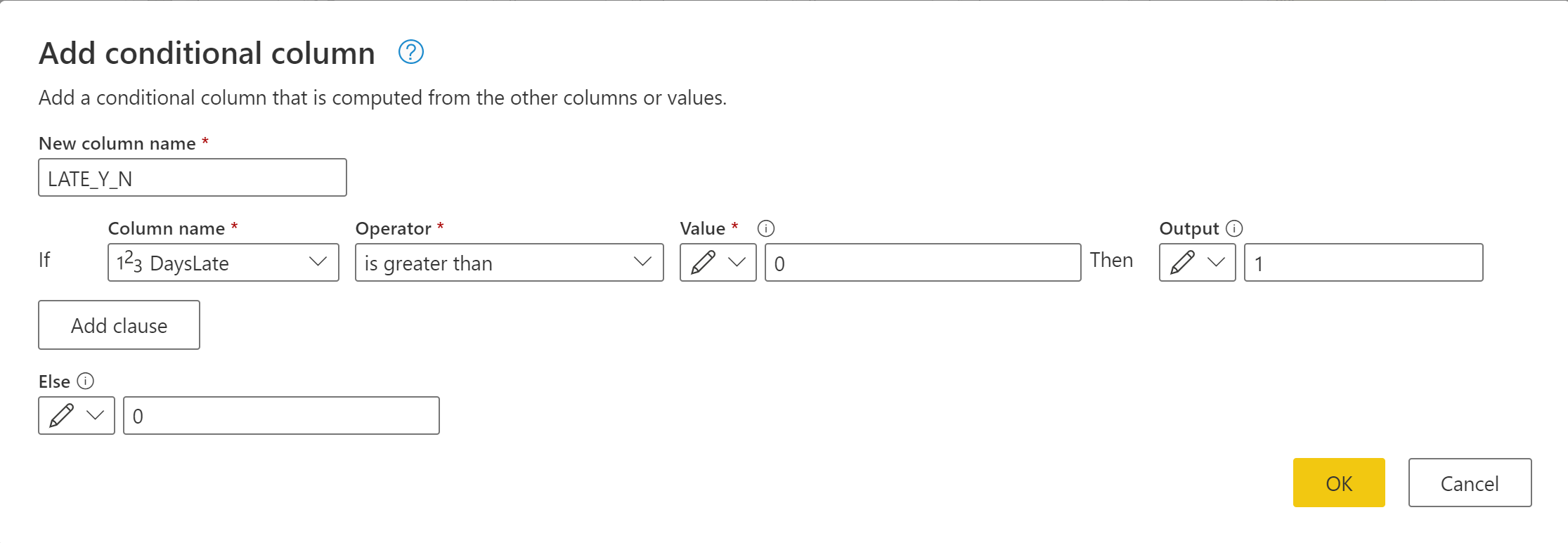
Description automatically generated

1. Select **Add column** from the Power Query menu then select **Conditional Column.**

Graphical user interface

Description automatically generated

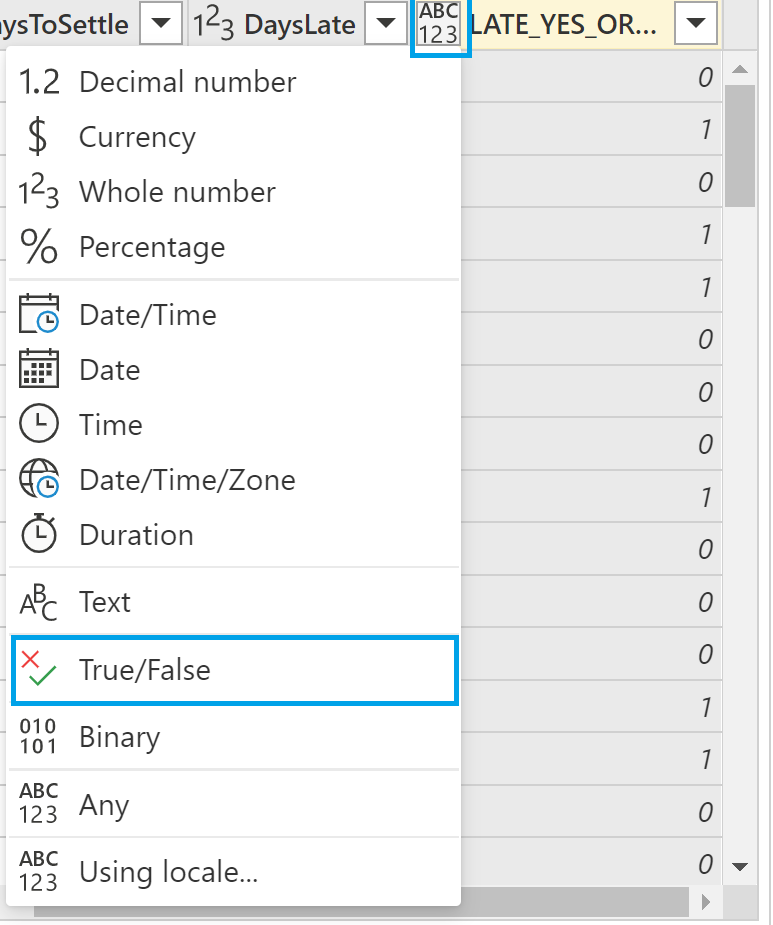
1. In the **Add conditional column** screen, add a new column (called *LATE\_Y\_N* in example) with the logic if column *DaysLate* is greater than 0, then the invoice is late (LATE\_Y\_N equals 1), otherwise it is not late (LATE\_Y\_N equals 0).



This is the PowerQuery formula created for the new column we named *LATE\_Y\_N*:

|  |
| --- |
| Table.AddColumn(#"Changed column type", "LATE\_Y\_N", each if [DaysLate] > 0 then 1 else 0) |

1. Click on the new conditional column we named *LATE\_Y\_N* to change its datatype to True/False then **Save & Close**.



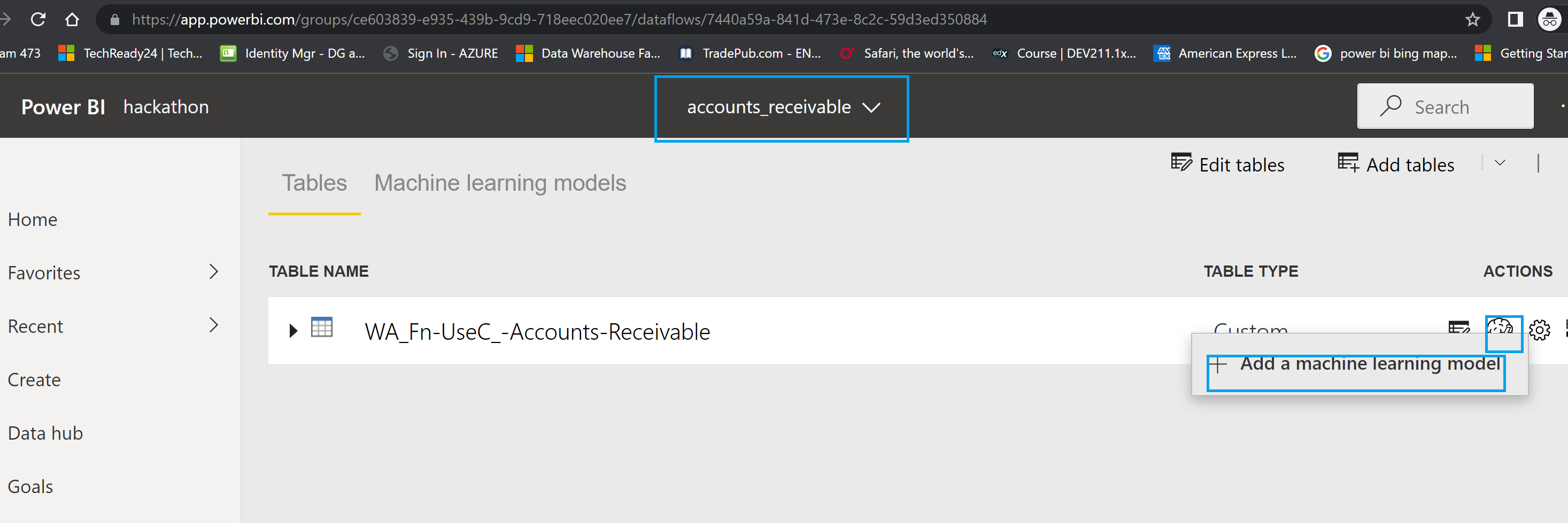
1. Name your dataflow *accounts\_receivable* and click **Save**.

A picture containing graphical user interface

Description automatically generated

### Task 2: Create an ML model directly in dataflows

1. Select the **accounts\_receivable** dataflow.
2. Click the **Brain** icon.
3. Select **+ Add a machine learning model**.

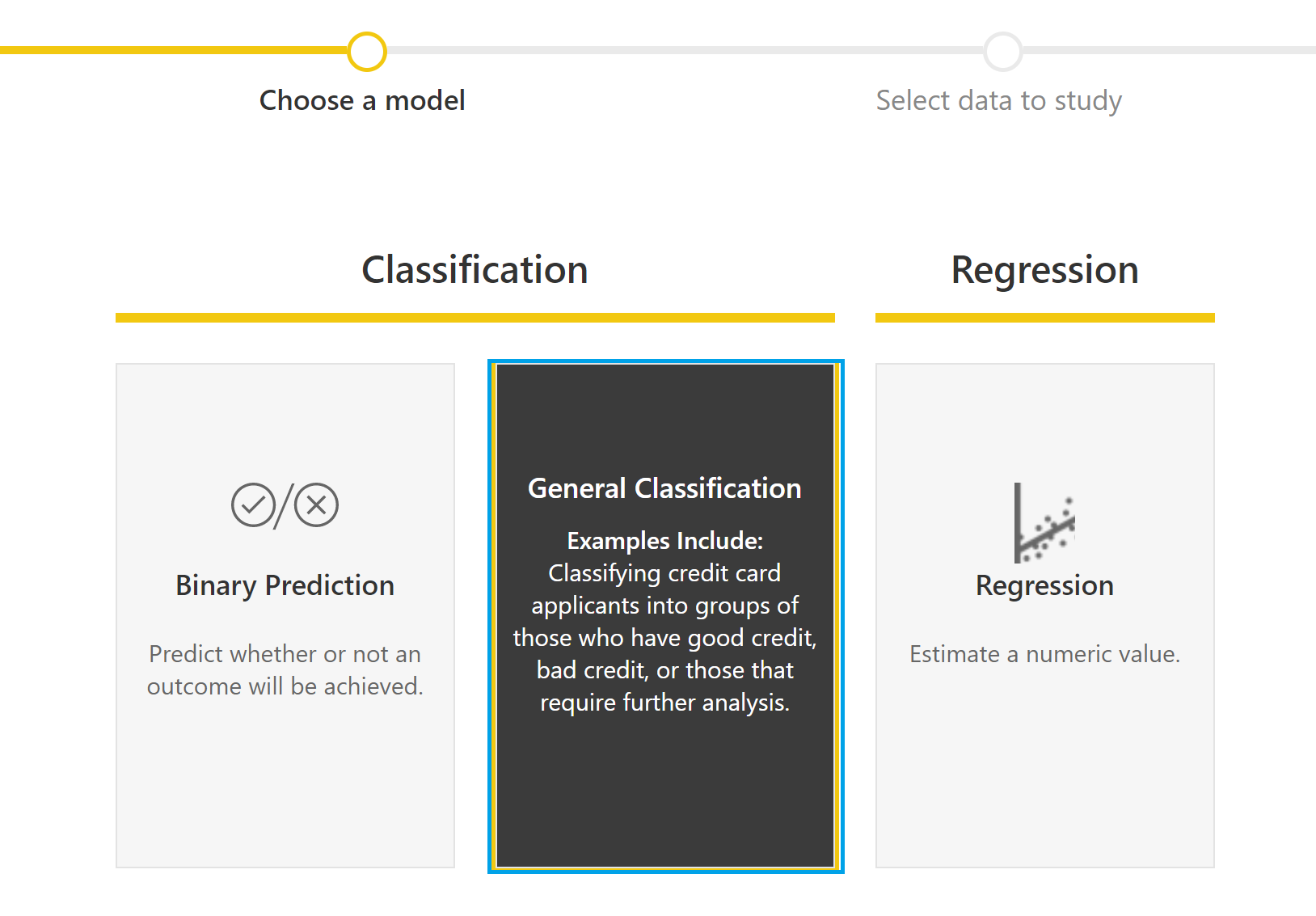
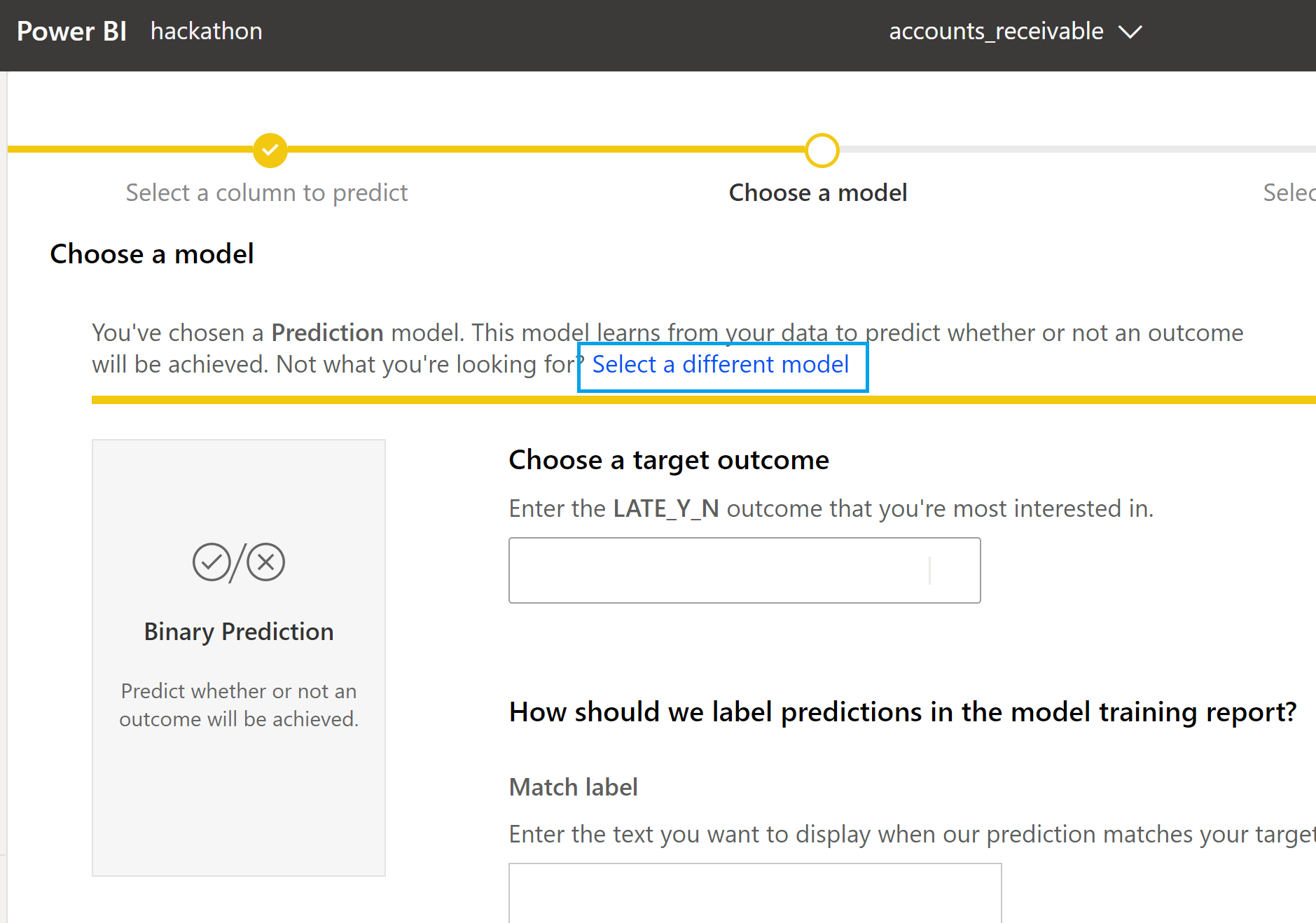


1. Select the table you created earlier, and for the **Outcome column**, select the *LATE* conditional column you created, then Click **Next**.

Graphical user interface, text, application, email

Description automatically generated

1. Choose **General Classification** and click Next. If the General Classification Model isn’t shown, you’ll need to first click on Select a different model.

1. Select all columns except for *CustomerID*, *InvoiceNumber*, *DaysToSettle*, and *DaysLate*.

Graphical user interface, application

Description automatically generated

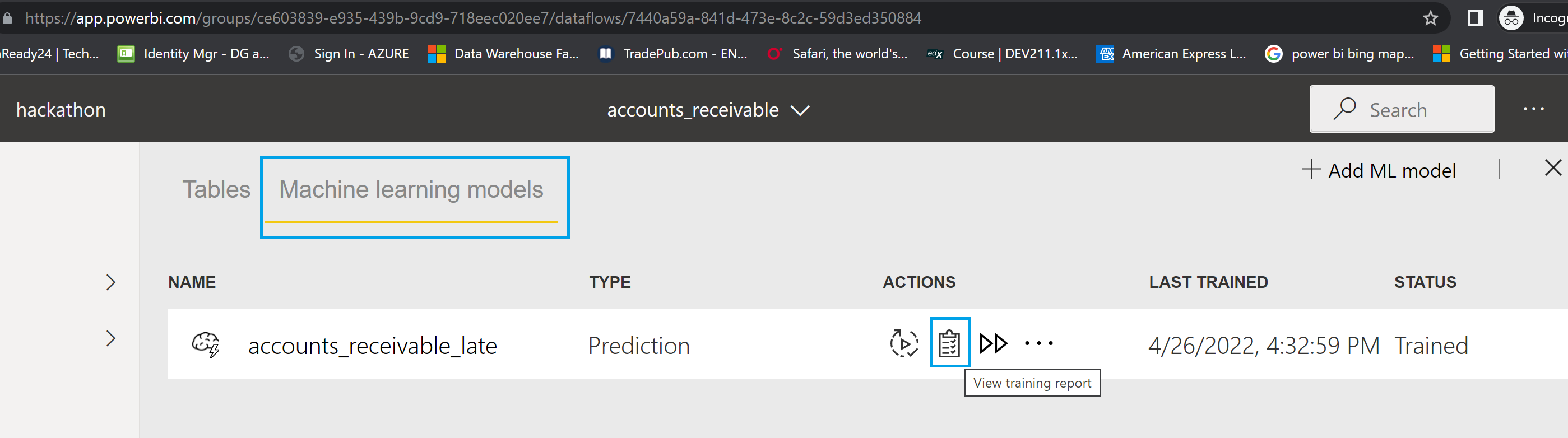
1. Click **Save & Train** keeping the training time set to 5 minutes.

Text, email

Description automatically generated

### Task 3: Review the Power BI Model Validation Report

1. After a few minutes, the model should complete. Go into **Machine learning models** then select the **View training report** icon.

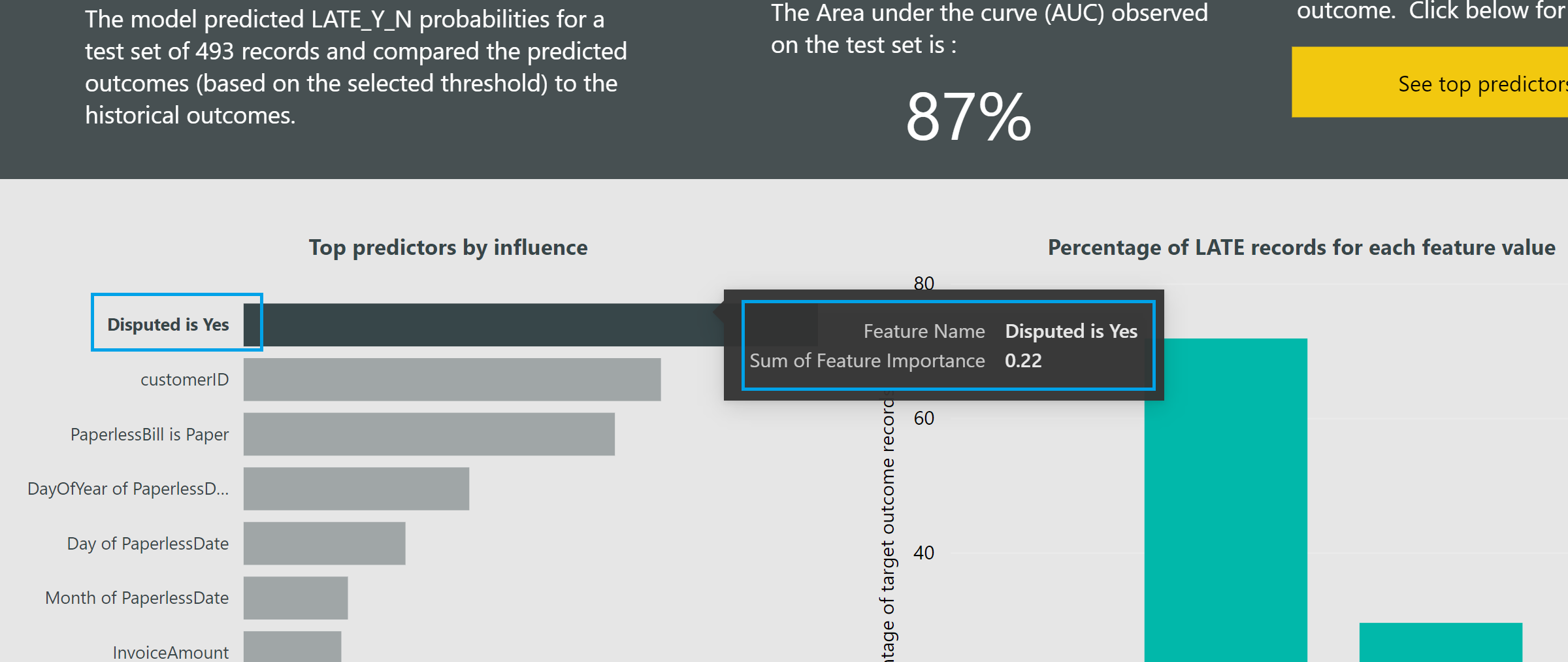


1. In the Model Performance report, click on **See top predictors**.

Graphical user interface

Description automatically generated with medium confidence

1. Observe that the top predictor is *Disputed is Yes*, which you can see explains why 22% of invoices are paid late. It makes sense that disputed invoices might not be paid on time. It’s also interesting that *PaperlessBill is Paper* explains why 14% of invoices are paid late. Perhaps paper bills get lost in the mail more often than digital bills.



1. Notice how there is an inverse relationship between precision and threshold settings. As you increase the precision the recall decreases and vice versa. So, depending on what is more important – i.e. predicting late and actually being late or predicting late but wasn’t late at all you can modify your objective.Table

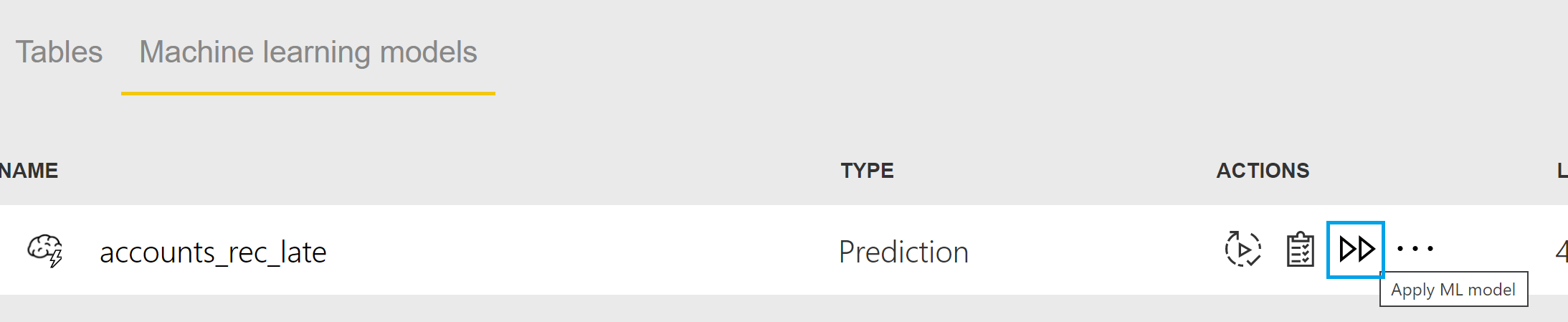
   Description automatically generated with low confidence
2. Click on the Accuracy Tab to see the area under the curve. The area under the curve represents the gain of not just randomly guessing late or not for each customer invoice. Hence, the more area under the curve, the better the model is at predicting lateness or not.

Chart, line chart

Description automatically generated

### Task 4: Apply the model

1. Click on Apply ML model. NOTE: Preparation takes 10-15 minutes to prepare the experiment run. Once running, it takes 2-3 minutes more for each iteration. In production, you'd likely walk away for a bit. But for this tutorial, we suggest you start exploring the tested algorithms on the Models tab as they complete while the others are still running.



1. Call the New output column accounts\_rec\_late and use 0.5 as the Threshold value. Click Save & Apply.

Graphical user interface, text, application, email

Description automatically generated

1. Click on enriched late to see predictions and explanations on a record by record basisGraphical user interface, text, application

   Description automatically generated

Example:

Graphical user interface, application

Description automatically generated