Exercise 3 – AutoML – Banking

**Tutorial: Create your first classification model with automated machine learning**

In this tutorial, you learn how to create your first automated machine learning experiment through Azure Machine Learning studio without writing a single line of code. This example creates a classification model to predict if a client will subscribe to a fixed term deposit with a financial institution.

With automated machine learning, you can automate away time intensive tasks. Automated machine learning rapidly iterates over many combinations of algorithms and hyperparameters to help you find the best model based on a success metric of your choosing.

In this tutorial, you learn how to do the following tasks:

* Create an Azure Machine Learning workspace.
* Run an automated machine learning experiment.
* View experiment details.
* Deploy the model.

**Prerequisites**

* An Azure subscription. If you don’t have an Azure subscription, create a [free account](https://aka.ms/AMLFree).
* Download the [**bankmarketing\_train.csv**](https://automlsamplenotebookdata.blob.core.windows.net/automl-sample-notebook-data/bankmarketing_train.csv) data file.(click link) The **y** column indicates if a customer subscribed to a fixed term deposit, which is later identified as the target column for predictions in this tutorial.

**Create a workspace**

An Azure Machine Learning workspace is a foundational resource in the cloud that you use to experiment, train, and deploy machine learning models. It ties your Azure subscription and resource group to an easily consumed object in the service.

You create a workspace via the Azure Machine Learning studio, a web-based console for managing your Azure resources.

1. Sign in to the [Azure portal](https://portal.azure.com/) by using the credentials for your Azure subscription.
2. In the upper-left corner of the Azure portal, select **+ Create a resource**.
3. Use the search bar to find **Machine Learning**.
4. Select **Machine Learning**.
5. In the **Machine Learning** pane, select **Create** to begin.
6. Provide the following information to configure your new workspace:

| **Field** | **Description** |
| --- | --- |
| Workspace name | Enter a unique name that identifies your workspace. In this example, we use **docs-ws**. Names must be unique across the resource group. Use a name that's easy to recall and to differentiate from workspaces created by others. |
| Subscription | Select the Azure subscription that you want to use. |
| Resource group | Use an existing resource group in your subscription, or enter a name to create a new resource group. A resource group holds related resources for an Azure solution. In this example, we use **docs-aml**. |
| Location | Select the location closest to your users and the data resources to create your workspace. |
| Workspace edition | Select **Enterprise**. This tutorial requires the use of the Enterprise edition. The Enterprise edition is in preview and doesn't currently add any extra costs. |

1. After you're finished configuring the workspace, select **Create**.

**Warning**

It can take several minutes to create your workspace in the cloud.

When the process is finished, a deployment success message appears.

1. To view the new workspace, select **Go to resource**.

**Important**

Take note of your **workspace** and **subscription**. You'll need these to ensure you create your experiment in the right place.

**Create and run the experiment**

You complete the following experiment set-up and run steps in Azure Machine Learning studio, a consolidated interface that includes machine learning tools to perform data science scenarios for data science practitioners of all skill levels. The studio is not supported on Internet Explorer browsers.

1. Sign in to [Azure Machine Learning studio](https://ml.azure.com/).
2. Select your subscription and the workspace you created.
3. Select **Get started**.
4. In the left pane, select **Automated ML** under the **Author** section.

Since this is your first automated ML experiment, you'll see an empty list and links to documentation.

A screenshot of a cell phone

Description automatically generated

1. Select **New automated ML run**.
2. Create a new dataset by selecting **From local files** from the **+Create dataset** drop-down.
   1. Select **Browse**.
   2. Choose the **bankmarketing\_train.csv** file on your local computer. This is the file you downloaded as a [prerequisite](https://automlsamplenotebookdata.blob.core.windows.net/automl-sample-notebook-data/bankmarketing_train.csv).
   3. Select **Tabular** as your dataset type.
   4. Give your dataset a unique name and provide an optional description.
   5. Select **Next** on the bottom left, to upload it to the default container blob store(datastore) that was automatically set up during your workspace creation.

When the upload is complete, the Settings and preview form is pre-populated based on the file type.

* 1. Verify that the **Settings and preview** form is populated as follows and select **Next**.

| **Field** | **Description** | **Value for tutorial** |
| --- | --- | --- |
| File format | Defines the layout and type of data stored in a file. | Delimited |
| Delimiter | One or more characters for specifying the boundary between  separate, independent regions in plain text or other data streams. | Comma |
| Encoding | Identifies what bit to character schema table to use to read your dataset. | UTF-8 |
| Column headers | Indicates how the headers of the dataset, if any, will be treated. | All files have same headers |
| Skip rows | Indicates how many, if any, rows are skipped in the dataset. | None |

* 1. The **Schema** form allows for further configuration of your data for this experiment. For this example, select the toggle switch for the **day\_of\_week** feature, so as to not include it for this experiment. Select **Next**.

A screenshot of a cell phone

Description automatically generated

* 1. On the **Confirm details** form, verify the information matches what was previously populated on the **Basic info** and **Settings and preview** forms.
  2. Select **Create** to complete the creation of your dataset.
  3. Select your dataset once it appears in the list.
  4. Review the **Data preview** to ensure you didn't include **day\_of\_week** then, select **OK**.
  5. Select **Next**.

1. Populate the **Configure Run** form as follows:
   1. Enter this experiment name: my-1st-automl-experiment
   2. Select **y** as the target column, what you want to predict. This column indicates whether the client subscribed to a term deposit or not.
   3. Select **Create a new compute** and configure your compute target. A compute target is a local or cloud-based resource environment used to run your training script or host your service deployment. For this experiment, we use a cloud-based compute.

| **Field** | **Description** | **Value for tutorial** |
| --- | --- | --- |
| Compute name | A unique name that identifies your compute context. | automl-compute |
| Virtual machine size | Select the virtual machine size for your compute. | Standard\_DS12\_V2 |
| Min / Max nodes (in Advanced Settings) | To profile data, you must specify 1 or more nodes. | Min nodes: 1 Max nodes: 6 |

* + 1. Select **Create** to get the compute target.

**This takes a couple minutes to complete.**

* + 1. After creation, select your new compute target from the drop-down list.
  1. Select **Next**.

1. On the **Task type and settings** form, select **Classification** as the machine learning task type.
   1. Select **View additional configuration settings** and populate the fields as follows. These settings are to better control the training job. Otherwise, defaults are applied based on experiment selection and data.

**Note**

In this tutorial, you won't set a metric score or max cores per iterations threshold. Nor will you block algorithms from being tested.

| **Additional configurations** | **Description** | **Value for tutorial** |
| --- | --- | --- |
| Primary metric | Evaluation metric that the machine learning algorithm will be measured by. | AUC\_weighted |
| Automatic featurization | Enables preprocessing. This includes automatic data cleansing, preparing, and transformation to generate synthetic features. | Enable |
| Blocked algorithms | Algorithms you want to exclude from the training job | None |
| Exit criterion | If a criteria is met, the training job is stopped. | Training job time (hours): 1 Metric score threshold: None |
| Validation | Choose a cross-validation type and number of tests. | Validation type:  k-fold cross-validation  Number of validations: 2 |
| Concurrency | The maximum number of parallel iterations executed and cores used per iteration | Max concurrent iterations: 5 Max cores per iteration: None |

Select **Save**.

1. Select **Finish** to run the experiment. The **Run Detail** screen opens with the **Run status** as the experiment preparation begins.

**Important**

Preparation takes **10-15 minutes** to prepare the experiment run. Once running, it takes **2-3 minutes more for each iteration**.  
Select **Refresh** periodically to see the status of the run as the experiment progresses.

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.

**Explore models**

Navigate to the **Models** tab to see the algorithms (models) tested. By default, the models are ordered by metric score as they complete. For this tutorial, the model that scores the highest based on the chosen **AUC\_weighted** metric is at the top of the list.

While you wait for all of the experiment models to finish, select the **Algorithm name** of a completed model to explore its performance details.

The following navigates through the **Model details** and the **Visualizations** tabs to view the selected model's properties, metrics, and performance charts.

A screenshot of a social media post

Description automatically generated

**Deploy the model**

Automated machine learning in Azure Machine Learning studio allows you to deploy the best model as a web service in a few steps. Deployment is the integration of the model so it can predict on new data and identify potential areas of opportunity.

For this experiment, deployment to a web service means that the financial institution now has an iterative and scalable web solution for identifying potential fixed term deposit customers.

Once the run is complete, navigate back to the **Run Detail** page and select the **Models** tab. Select **Refresh**.

In this experiment context, **VotingEnsemble** is considered the best model, based on the **AUC\_weighted** metric. We deploy this model, but be advised**, deployment takes about 20 minutes to complete. T**he deployment process **entails several steps including registering the model, generating resources, and configuring them for the web service**.

1. Select the **Deploy Best Model** button in the bottom-left corner.
2. Populate the **Deploy a model** pane as follows:

| **Field** | **Value** |
| --- | --- |
| Deployment name | my-automl-deploy |
| Deployment description | My first automated machine learning experiment deployment |
| Compute type | Select Azure Compute Instance (ACI) |
| Enable authentication | Disable. |
| Use custom deployments | Disable. Allows for the default driver file (scoring script) and environment file to be autogenerated. |

1. For this example, we use the defaults provided in the *Advanced* menu.
2. Select **Deploy**.

A green success message appears at the top of the **Run** screen, and in the **Recommended model** pane, a status message appears under **Deploy status**. Select **Refresh** periodically to check the deployment status.

Now you have an operational web service to generate predictions.

Proceed to the [**Next Steps**](https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-first-experiment-automated-ml#next-steps) to learn more about how to consume your new web service, and test your predictions using Power BI's built in Azure Machine Learning support.

**Clean up resources**

Deployment files are larger than data and experiment files, so they cost more to store. Delete only the deployment files to minimize costs to your account, or if you want to keep your workspace and experiment files. Otherwise, delete the entire resource group, if you don't plan to use any of the files.

**Delete the deployment instance**

Delete just the deployment instance from the Azure Machine Learning studio, if you want to keep the resource group and workspace for other tutorials and exploration.

1. Go to the [Azure Machine Learning studio](https://ml.azure.com/). Navigate to your workspace and on the left under the **Assets** pane, select **Endpoints**.
2. Select the deployment you want to delete and select **Delete**.
3. Select **Proceed**.

**Delete the resource group**

**Important**

The resources you created can be used as prerequisites to other Azure Machine Learning tutorials and how-to articles.

If you don't plan to use the resources you created, delete them, so you don't incur any charges:

1. In the Azure portal, select **Resource groups** on the far left.

A screenshot of a social media post

Description automatically generated

1. From the list, select the resource group you created.
2. Select **Delete resource group**.
3. Enter the resource group name. Then select **Delete**.

**Next steps**

In this automated machine learning tutorial, you used Azure Machine Learning studio to create and deploy a classification model. See these articles for more information and next steps:

[Consume a web service](https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service#consume-the-service-from-power-bi)

* Learn more about [preprocessing](https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-portal-experiments#preprocess).
* Learn more about [data profiling](https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-portal-experiments#profile).
* Learn more about [automated machine learning](https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml).
* For more information on classification metrics and charts, see the [Understand automated machine learning results](https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#classification) article.

**Note**

This Bank Marketing dataset is made available under the [**Creative Commons (CCO: Public Domain) License**](https://creativecommons.org/publicdomain/zero/1.0/). Any rights in individual contents of the database are licensed under the [**Database Contents License**](https://creativecommons.org/publicdomain/zero/1.0/) and available on [**Kaggle**](https://www.kaggle.com/janiobachmann/bank-marketing-dataset). This dataset was originally available within the [**UCI Machine Learning Database**](https://archive.ics.uci.edu/ml/datasets/bank+marketing).  
  
[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014.