Contents

1 Retired

This repo is now retired. The current content for course DP-100 is at https://github.com/MicrosoftLearning/DP100/blob/mas

2 Lab 1 - Train a Model in Azure Notebook

2.1 Lab 1.0: Objectives

In this lab, you will:

- Use Azure Notebook with Python, a cloud based Jupyter Notebook service.
- Perform exploratory data analysis
- Create machine learning model features
- Train an open source based classification predictive model.

The main goal of this lab is to get students up to speed using Azure Notebooks while getting them started on the Adventure Works use case.

2.2 Introduction

You're going to train a classification model in an Azure Notebook using Python and the open source package scikit-learn. As part of that you will need to perform some basic exploratory data analysis to understand the

data. Then you will need to create features to be used in model training. Finally, you will train and evaluate your model. Note: At this point, you will not be using Azure services.

Your team's data engineer has provided you with a data extract file in CSV format that has all the data you need. The starter lab notebook will guide you through the tasks you need to perform for this lab.

2.3 Lab 1: Resources

You will use the following files from the lab folder:

Name	Description
Starter_Lab1_Notebook.ipynb	The lab notebook you should use. Import this into your Notebook project and open it.
AWData.csv	Adventure Works data extract file. Upload this file to your Azure Notebook project.

For help on create Azure Notebook projects and importing notebooks see: https://docs.microsoft.com/en-us/azure/notebooks/quickstart-migrate-local-jupyter-notebook

For help on uploading data into Azure Notebook see: https://docs.microsoft.com/en-us/azure/notebooks/work-with-project-data-files

2.4 Prerequisites

Before you can do this lab, you need to have:

- Signed Up for the free Azure Notebook Service.

3 Lab 2 - Register and deploy an ML model

3.1 Lab 2.0: Objectives

In this lab, you will:

- Train your model using the Azure Machine Learning service.
- Register the model you created to the registry in the workspace
- Create a model scoring script
- Create a YAML file configuring Python module dependencies
- Create a container image
- Deploy a model as a web service
- Score new data with the deployed model

3.2 Introduction

In this lab, you will train the model you developed in the last lab on Azure using the Azure Machine Learning service and its Python SDK. After it has been trained, you will register the model to the registry and perform the steps necessary to deploy your model to Azure Machine Learning service where it can be used.

3.3 Lab 2: Resources

You will use the following files from the lab folder:

Name	Description
Starter_Lab2_Notebook.ipynb	The lab notebook you should use. Import this into your Notebook project and open it.

For help on creating Azure Notebook projects and importing notebooks see: https://docs.microsoft.com/en-us/azure/notebooks/quickstart-migrate-local-jupyter-notebook

For help on uploading data into Azure Notebook see: https://docs.microsoft.com/en-us/azure/notebooks/work-with-project-data-files

3.4 Prerequisites

Before you can do this lab, you need to have:

- Signed Up for the free Azure Notebook Service.
- An Azure subscription that can provision an Azure Machine Learning service workspace.

4 Lab 3 - Use AutoML and HyperDrive

4.1 Lab 3.0: Objectives

In this lab, you will:

- Understand the machine learning pipeline
- Understand Azure Machine Learning service AutoML and Hyperdrive
- Create a Python script that uses the Azure Machine Learning service's AutoML to recommend a model
- Test the recommended model from your Python script

4.2 Introduction

In this lab, you want to see if there are models that perform better than the one you manually created earlier. You decide to use Azure Machine Learning service's AutoML and HyperDrive to simultaneously execute a number of different types of classification models, compare the results, and recommend the best performing model. This will save you a lot of time picking the best model so you can get the solution delivered sooner.

4.3 Lab 3: Resources

You will use the following files from the lab folder:

Name	Description
Starter_Lab3_Notebook.ipynb	The lab notebook you should use. Import this into your Notebook project and open it.

For help on creating Azure Notebook projects and importing notebooks see: https://docs.microsoft.com/en-us/azure/notebooks/quickstart-migrate-local-jupyter-notebook

For help on uploading data into Azure Notebook see: https://docs.microsoft.com/en-us/azure/notebooks/work-with-project-data-files

4.4 Prerequisites

Before you can do this lab, you need to have:

- Signed Up for the free Azure Notebook Service.
- An Azure subscription that can provision an Azure Machine Learning service workspace.

5 Lab Introduction

5.0.1 The Azure Machine Learning service: Putting it all together.

In this lab, we introduce our workshop case study which involves training and deploying a predictive model using the Azure Machine Learning service. This lab requires the student have access to the Azure Notebook service and an Azure subscription.

6 Workshop Case Study

6.1 Scenario

You've been assigned a new customer, Adventure Works LLC, which sells bicycles and bicycle equipment to its customers.

Adventure Works Cycles is a large multinational manufacturing company. It manufactures and sells bicycles and bicycle components to the North American, European and Asian commercial markets through both an

internet channel and a reseller distribution network. Its base operations are in Kirkland, Washington with 290 employees, and there are several regional sales teams located throughout their market base.

Coming off a successful fiscal year, Adventure Works is looking to increase its revenue by targeting additional sales to their existing customers. Although Adventure Works sells bike accessories and components, the majority of business is in bike sales. A marketing study shows that many customers who bought component and accessories, have never purchased a bike from Adventure Works. They want to target these customers with bike sale promotions. Rather than send blanket emails or display generic online ads, they have asked the data science team to create a predictive machine learning model that identifies the type of bike that a customer is most likely to purchase. The assumption is that based on customer demographics and sales data, it should be feasible to determine the type of bike a customer will buy: Road, Mountain, or, Touring. Note: Phase two of this project will build on this model to predict the specific bike model a customer will buy but we want to start slow.

Using customer and sales history, marketing believes you have what you need to create a predictive model. It is critical that the model be able to scale as needed to support increased usage from multiple applications, i.e. emails, online ads, direct mail, etc. You need to provide a standard way to score data with your model, i.e. an easy to use an API. Marketing is eager to get this developed and deployed as soon as possible so you are encouraged to find ways to speed things up.

Your objective is to do the following:

- Perform Exploratory Data Analysis (EDA) to determine which historical data will help predict the bike purchased.
- Based on the outcome of the EDA, create the model features, aka perform feature engineering.
- Determine the type of machine learning model you need. This will involve trying several models, evaluating their performance, and selecting the best one.
- As part of selecting the best performing model, you need to determine the optimal model hyperparameters.
- Deploy the trained model to an environment where it can be easily called.
- Test using the model.
- Monitor the model's use.

6.1.1 Solution

You will be implementing the solution using Azure Machine Learning service. Azure Notebook provide the interactive development environment (IDE) and Python will be our programming language. Azure Notebook is pre-configured with the Azure Machine Learning software development kits (SDKs) you need and most of the popular Python data science related packages.

- Using a data extract provided by the data engineer, you will do some exploratory data analysis and create several additional features for the model.
- You will create a completely open source based classification model and evaluating its performance.
- You will extend the model training code you created to leverage the Azure Machine Learning service.
- You want to be though in selecting the best model but do not have time to train and compare numerous types. To speed things up, you will use Azure Machine Learning service AutoML to select the best model and determine the optimal hyperparameter values.
- You will register the model and the container image to the Azure Machine Learning service workspace.
- Then you will deploy the model and test it.
- You will monitor your model via Python.

A starting Azure Notebook in each lab provides the structure of the lab. More details about the above steps are provided in the lab technical requirements and the Azure notebook.

7 Labs: Meeting the Technical Requirements

7.1 Labs: What You Need

The labs for the data scientist and focus on using the Azure Machine Learning service to train and deploy a predictive machine learning model.

To do the labs, you need the following:

- Access to the Azure Notebook Service which is a cloud based hosting for Jupyter Notebook. You will use the free version. All the required Python libraries are pre-installed in this environment.
- An Azure subscription that can provision an Azure Machine Learning service workspace.
- The Azure Notebook for this lab.

7.1.1 Account Setup

Your instructor will guide you through the necessary steps to get the environment up and running.

7.1.1.1 Setup your Azure Account:

You may activate an Azure free trial at https://azure.microsoft.com/en-us/free/.

If you have been given an Azure Pass to complete this lab, you may go to http://www.microsoftazurepass.com/ to activate it. Please follow the instructions at https://www.microsoftazurepass.com/howto, which document the activation process. A Microsoft account may have **one free trial** on Azure and one Azure Pass associated with it, so if you have already activated an Azure Pass on your Microsoft account, you will need to use the free trial or use another Microsoft account.

Note: Internet Explorer is not supported. We recommend using Chrome, Edge, or, Firefox.

7.2 Lab Files

A Jupyter Notebook is provided for each lab that provides the structure of the lab. Because Azure Machine Learning service is a very exacting technology, most of the code that interfaces with AMLS is provided ready to run. What is left undone is to add in the Python machine learning code as that is the job of a data scientist. Bear in mind, the goal of the labs is to help the data scientist understand how to use AMLS in their work.

Lab 1 provides a comma separated (CSV) file to jump start the labs. The assumption is that a data engineer provided the extract to the data scientist which is a typical workflow.

7.2.1 Notebooks prefixed with Starting_ are the labs to be worked on.

7.2.2 Notebooks prefixed with Finished_ are completed notebooks, i.e. the solution.

7.3 Resources for future projects/learning

To deepen your understanding of the Azure Machine Learning service, and to involve your broader team in the development of data science solutions, we recommend reviewing the following resources:

- Azure Machine Learning Service Documentation
- Azure Machine Learning service

7.4 Credits