

Applied Machine Learning

Lab 1 – Working with Time Series Data

Overview

In this lab, you will use R to create a forecasting model for time series data. Specifically, you will write R code to predict dairy production levels for the next twelve months based on historical data.

What You'll Need

To complete this lab, you will need the following:

- An Azure ML account
- The files for this lab

Note: To set up the required environment for the lab, follow the instructions in the [Setup Guide](#) for this course.

Exploring and Uploading Historical Data

In this lab, you will create a forecasting model for dairy production. The forecasting model is based on an existing dataset of dairy production history for California.

Explore the Dataset

1. In the folder where you extracted the lab files for this module (for example, C:\DAT203.3x\Lab01), open the **cadairydata.csv** file, using either a spreadsheet application such as Microsoft Excel, or a text editor such as Microsoft Windows Notepad.
2. View the contents of the **cadairydata.csv** file, noting that it contains dairy production data from January 1995 to December 2013, as shown here:

	A	B	C	D	E	F	G	H	I
1		Year.Month	Month.Number	Year	Month	Cotagecheese.Prod	Icecream.Prod	Milk.Prod	N.CA.Fat.Price
2	1	1995.01	1	1995	Jan	4.37	51.595	2.112	0.9803
3	2	1995.02	2	1995	Feb	3.695	56.086	1.932	0.8924
4	3	1995.03	3	1995	Mar	4.538	68.453	2.162	0.8924
5	4	1995.04	4	1995	Apr	4.28	65.722	2.13	0.8967
6	5	1995.05	5	1995	May	4.47	73.73	2.227	0.8967
7	6	1995.06	6	1995	June	4.238	77.994	2.124	0.916
8	7	1995.07	7	1995	July	4.377	81.475	2.184	0.916
9	8	1995.08	8	1995	Aug	4.368	74.981	2.152	0.8934
10	9	1995.09	9	1995	Sept	3.917	61.53	2.062	0.8934
11	10	1995.1	10	1995	Oct	4.078	60.022	2.121	0.9434
12	11	1995.11	11	1995	Nov	3.611	52.772	2.03	0.9434
13	12	1995.12	12	1995	Dec	3.591	50.85	2.091	1.0811
14	13	1996.01	1	1996	Jan	3.698	53.116	2.109	1.0811
15	14	1996.02	2	1996	Feb	3.429	57.463	1.996	1.1022
16	15	1996.03	3	1996	Mar	3.717	68.929	2.173	1.1022
17	16	1996.04	4	1996	Apr	3.688	69.673	2.163	1.048
18	17	1996.05	5	1996	May	3.897	77.387	2.233	1.048
19	18	1996.06	6	1996	June	3.756	86.814	2.142	1.073
20	19	1996.07	7	1996	July	4.099	83.006	2.182	1.073
21	20	1996.08	8	1996	Aug	3.070	70.36	2.120	1.2222

3. Close the data file without saving any changes.

Upload the Dataset to Azure Machine Learning

1. Browse to <https://studio.azureml.net> and sign in using the Microsoft account associated with your free Azure ML account.
2. If the **Welcome** page is displayed, close it by clicking the **OK** icon (which looks like a checkmark). Then, if the **New** page (containing a collection of Microsoft samples) is displayed, close it by clicking the **Close** icon (which looks like an X).
4. At the bottom left, click **NEW**; and in the **NEW** dialog box, in the **DATASET** tab, click **FROM LOCAL FILE**. Then in the **Upload a new dataset** dialog box, browse to select the **cadairydata.csv** file from the folder where you extracted the lab files on your local computer. Enter the following details, and then click the ✓ icon.
 - **This is a new version of an existing dataset:** Unselected
 - **Enter a name for the new dataset:** cadairydata.csv
 - **Select a type for the new dataset:** Generic CSV file with a header (.csv)
 - **Provide an optional description:** Historical dairy data.
5. Wait for the upload of the dataset to complete, then click **OK** on the status bar at the bottom of the Azure ML Studio page.

Working with Time-Series Data in Jupyter

Now you are ready to use R code in a Jupyter notebook to work with the time-series data and create a forecasting model for dairy production.

Upload a Jupyter Notebook

1. In Azure ML Studio, click **NEW**; and in the **NEW** dialog box, in the **NOTEBOOK** tab, click **Upload**. Then in the **Upload a new notebook** dialog box, browse to select the **TimeSeries.ipynb** file from the folder where you extracted the lab files on your local computer. Enter the following details, and then click the ✓ icon.
 - **Enter a name for the new notebook:** TimeSeries
 - **Select a language for the new notebook:** R
2. Wait for the upload of the notebook to complete, then click **OK** on the status bar at the bottom of the Azure ML Studio page.

Use R to Work with the Time Series Data

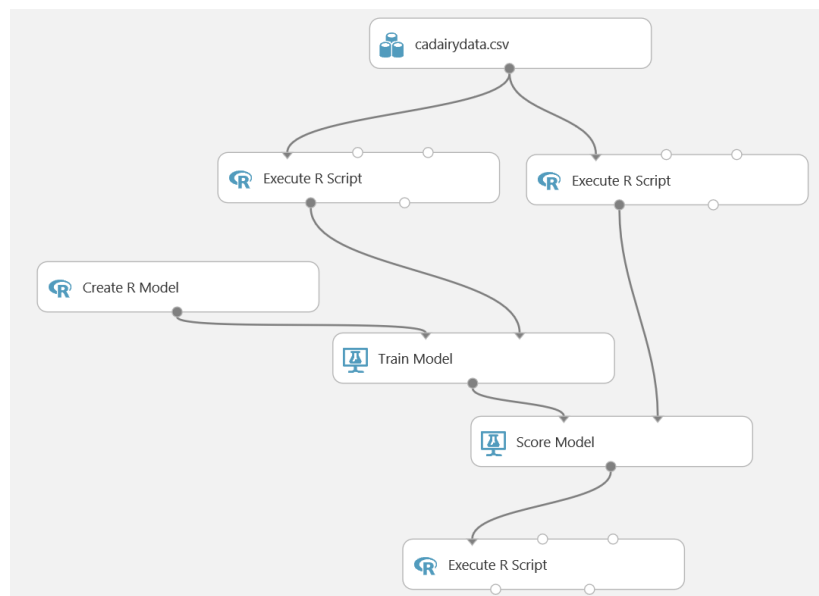
1. In Azure ML Studio, on the Notebooks tab, open the **TimeSeries** notebook you uploaded in the previous procedure.
2. Follow the instructions in the notebook to work with the time series data.
3. When you have completed all of the coding tasks in the notebook, save your changes and then close and halt the notebook.

Forecasting in Azure ML

In the previous exercises, you used a Jupyter notebook to explore time series data and create a forecasting model. Now you will use Azure ML to publish a forecasting model that uses similar code, and use it from a client application.

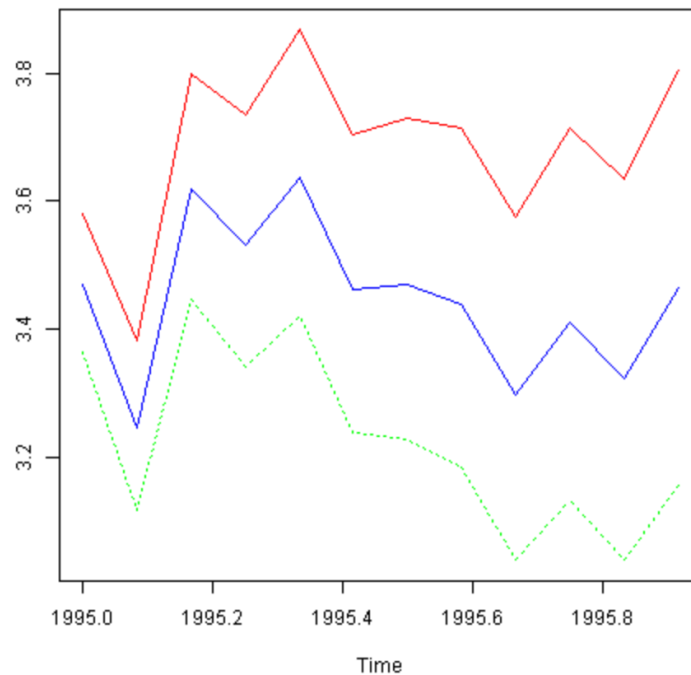
Create an Azure ML Experiment

1. In your Web browser, open the gallery experiment at <https://aka.ms/edx-dat203.3x-forecast>, and then open it in Azure ML Studio, copying it to your workspace. The copied experiment should look like this:



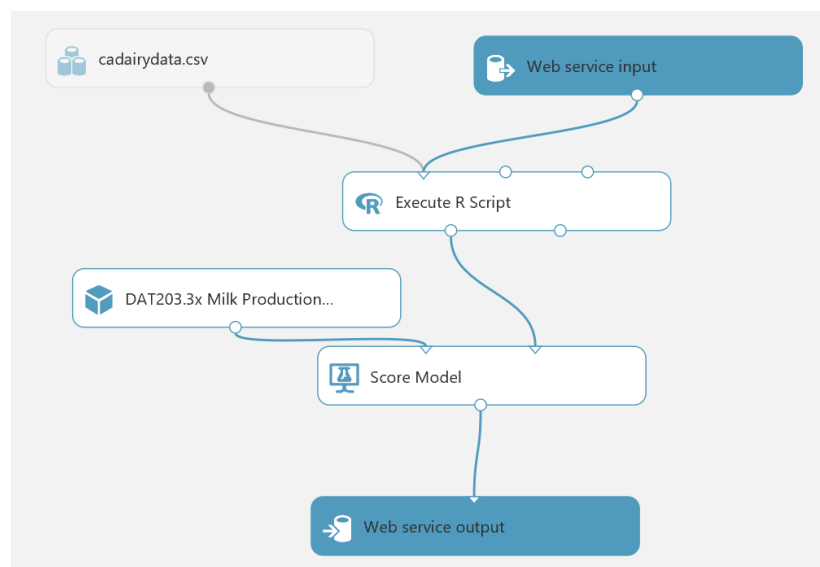
2. Note that the experiment contains two **Execute R Script** modules to prepare the time-series data for training and scoring, a **Create R Model** module that contains the R script to train and score a time series model, and a final **Execute R Script** module to visualize the results. Review the code in these modules.

- Save and run the experiment. Then, when it has finished running, visualize the **R Device** (right-most) output from the final **Execute R Script** module to view the forecast data, which should look like this:



Create a Predictive Web Service

- With the **DAT203.x:Milk Production Forecast** experiment still open, click **Set Up Web Service**, and then click **Predictive Web Service (Recommended)**. When a banner at the bottom of the screen notifies you that the experiment has been created, click **Close** to remove it.
- Verify that the initial predictive experiment looks like this:

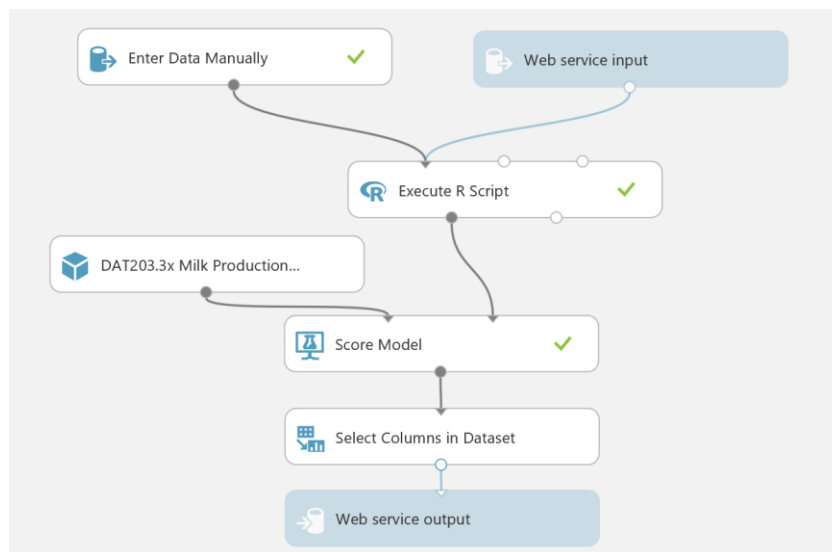


- Delete the **cadairydata.csv** dataset from the predictive experiment. Then add an **Enter Data Manually** module to the predictive experiment, and connect its output to the **Dataset1** (left-most) input of the **Execute R Script** module to replace the original **cadairydata.csv** dataset.

4. Edit the properties of the Enter Data Manually module, setting the **DataFormat** property to CSV, selecting the **HasHeader** checkbox, and entering the following data:

```
Year, Month, MonthNumber
2014, Jan, 1
2014, Feb, 2
2014, Mar, 3
2014, Apr, 4
2014, May, 5
2014, Jun, 6
2014, Jul, 7
2014, Aug, 8
2014, Sep, 9
2014, Oct, 10
2014, Nov, 11
2014, Dec, 12
```

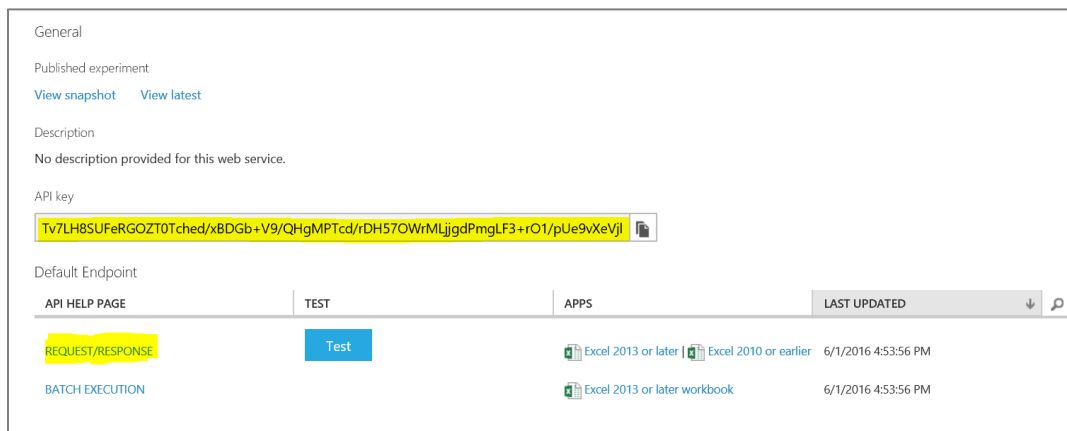
5. Save and run the experiment to read the data and pass it through the workflow. Then visualize the output of the **Score Model** module to verify that the output columns include a **forecast** along with a prediction for the **upper95** and **lower95** percentiles.
6. Add a **Select Columns in Dataset** module to the predictive experiment, and connect the output from the **Score Model** module to its input. Then connect its output to the **Web service output**.
7. In the properties for the **Select Columns in Dataset** module, use the column selector to select only the **forecast**, **upper95**, and **lower95** columns.
8. Verify that your predictive experiment looks like this:



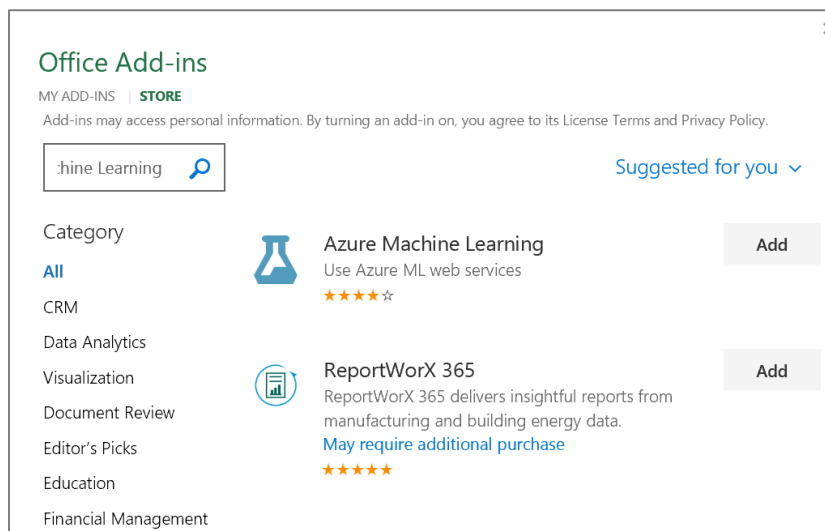
9. Save and run the experiment, and visualize the output of the **Select Columns in Dataset** module to verify that only the **forecast**, **upper95**, and **lower95** columns are returned by the web service.

Deploy and Use the Web Service

1. In the **DAT203.x:Milk Production Forecast [Predictive Exp.]** experiment, click the **Deploy Web Service** icon at the bottom of the Azure ML Studio window.
2. Wait a few seconds for the dashboard page to appear, and note the **API key** and **Request/Response** link. You will use these to connect to the web service from a client application.



3. Leave the dashboard page open in your web browser, and open a new browser tab.
4. In the new browser tab, navigate to <https://office.live.com/start/Excel.aspx>. If prompted, sign in with your Microsoft account (use the same credentials you use to access Azure ML Studio.)
5. In Excel Online, create a new blank workbook.
6. On the **Insert** tab, click **Office Add-ins**. Then in the **Office Add-ins** dialog box, select **Store**, search for *Azure Machine Learning*, and add the **Azure Machine Learning** add-in as shown below:



7. After the add-in is installed, in the **Azure Machine Learning** pane on the right of the Excel workbook, click **Add Web Service**. Boxes for the URL and API key of the web service will appear.
8. On the browser tab containing the dashboard page for your Azure ML web service, right-click the **Request/Response** link you noted earlier and copy the web service URL to the clipboard. Then return to the browser tab containing the Excel Online workbook and paste the URL into the URL box.
9. On the browser tab containing the dashboard page for your Azure ML web service, click the **Copy** button for the **API key** you noted earlier to copy the key to the clipboard. Then return to the browser tab containing the Excel Online workbook and paste it into the **API key** box.
10. Verify that the **Azure Machine Learning** pane in your workbook now resembles this, and click **Add**:

Azure Machine Learning

Web Services

Titanic Survivor Predictor (Excel Add-in Sa...

Text Sentiment Analysis (Excel Add-in Sam...

URL

https://studio.azureml.net/apihelp/workspaces/b2101c3182ae42c58c2466ab40607479/webservices/4d98abd657a449a895374b16fa2722aa/endpoints/66ce37974550469ba9b9c3d90eb25814/score

API key

Tv7LH8SUFERGOZT0Tched/xBDGb+V9/QHgMP.Tcd/rDH57OWrMLjgdPmgLF3+rO1/pUe9vXeVJlU7dHjU8TPyg==

Cancel

Add

☐ Auto-predict

Predict All

11. After the web service has been added, in the **Azure Machine Learning** pane, click **1. View Schema** and note the *inputs* expected by the web service (which consist of the fields toy added in the **Enter Data Manually** module) and the *outputs* returned by the web service (the fields you specified for the **Select Columns in Dataset** module).
12. In the Excel worksheet select cell A1. Then in the **Azure Machine Learning** pane, collapse the **1. View Schema** section and in the **2. Predict** section, click **Use sample data**. this enters some sample input values in the worksheet.
13. Modify the sample data as follows:

Year	Month	MonthNumber
2014	Jan	1
2014	Feb	2
2014	Mar	3
2014	Apr	4
2014	May	5
2014	Jun	6
2014	Jul	7
2014	Aug	8
2014	Sep	9
2014	Oct	10
2014	Nov	11
2014	Dec	12

14. Select the cells containing the input data (cells A1 to C13), and in the **Azure Machine Learning** pane, click the button to select the input range and confirm that it is **'Sheet1'!A1:C13**.
15. Ensure that the **My data has headers** box is checked.

16. In the **Output** box type **D1**, and ensure the **Include headers** box is checked.
17. Click the **Predict** button, and after a few seconds, view the predicted **forecast**, **upper95**, and **lower95** values for each month.

Summary

In this lab, you used R in a Jupyter notebook to work with time-series data. You then published an Azure ML web service that forecasts milk production, and used it from Excel.