



Azure Machine Learning (Hands-On Workshop) [Session B1/B4]

Zürich, 03.07.2017

www.make-bi.ch



INTRODUCTION

➤ GOAL

Having a beneficial experience with the use of Azure Machine Learning

- ❑ Familiarize ourselves with the AML platform
- ❑ Understand its capabilities
- ❑ Understand key-concepts of the field “Machine Learning”
- ❑ Run end-to-end a real case study

➤ Session Logistics

- ❑ Overview from Sotiris Dimopoulos (20’)
 - presentation
 - quick demo
- ❑ Implement the AML solution yourself (25’)
 - the user guide from Microsoft
 - assistants

ABOUT THE USER GUIDE

- Overview (pg. 3-4)

Next 2 slides

- Create a Workspace and a “New Experiment” (pg. 5-9)

Almost Done

- Explore/Visualize MPG-Data (pg. 10 - 13)

Quickly / Skip

- Create a predictive model (pg. 14 - 24)

Primary focus of this session



ABOUT THE USER GUIDE

- Overview (pg. 3-4)

Next 2 slides

- Create a Workspace and a “New Experiment” (pg. 5-9)

Almost Done

- Explore/Visualize MPG-Data (pg. 10 - 13)

Quickly / Skip

- Create a predictive model (pg. 14 - 24)

Primary focus of this session












OVERVIEW

➤ Scenario

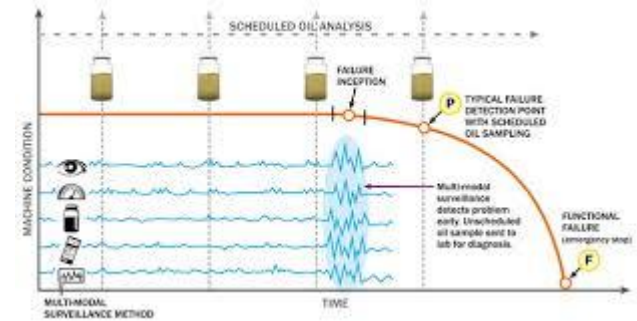
Predict the fuel consumption of cars (in Miles Per Gallons, MPG) from other automobile characteristics, like number of cylinders, horsepower, etc

➤ Dataset (~400 rows, 9 columns)

MPG	Cyl	Displacement	Horsepower	Weight	Acceleration	Year	CountryCode	Model
								
18	8	307	130	3504	12	70	1	chevrolet chevelle malibu
15	8	350	165	3693	11.5	70	1	buick skylark 320
18	8	318	150	3436	11	70	1	plymouth satellite
16	8	304	150	3433	12	70	1	amc rebel sst
17	8	302	140	3449	10.5	70	1	ford torino
15	8	429	198	4341	10	70	1	ford galaxie 500
14	8	454	220	4354	9	70	1	chevrolet impala
14	8	440	215	4312	8.5	70	1	plymouth fury iii
14	8	455	225	4425	10	70	1	pontiac catalina

OVERVIEW

- Regression Analysis: similar case studies



ABOUT THE USER GUIDE

- Overview (pg. 3-4)

Next 2 slides

- Create a Workspace and a “New Experiment” (pg. 5-9)

Almost Done

- Explore/Visualize MPG-Data (pg. 10 - 13)

Quickly / Skip

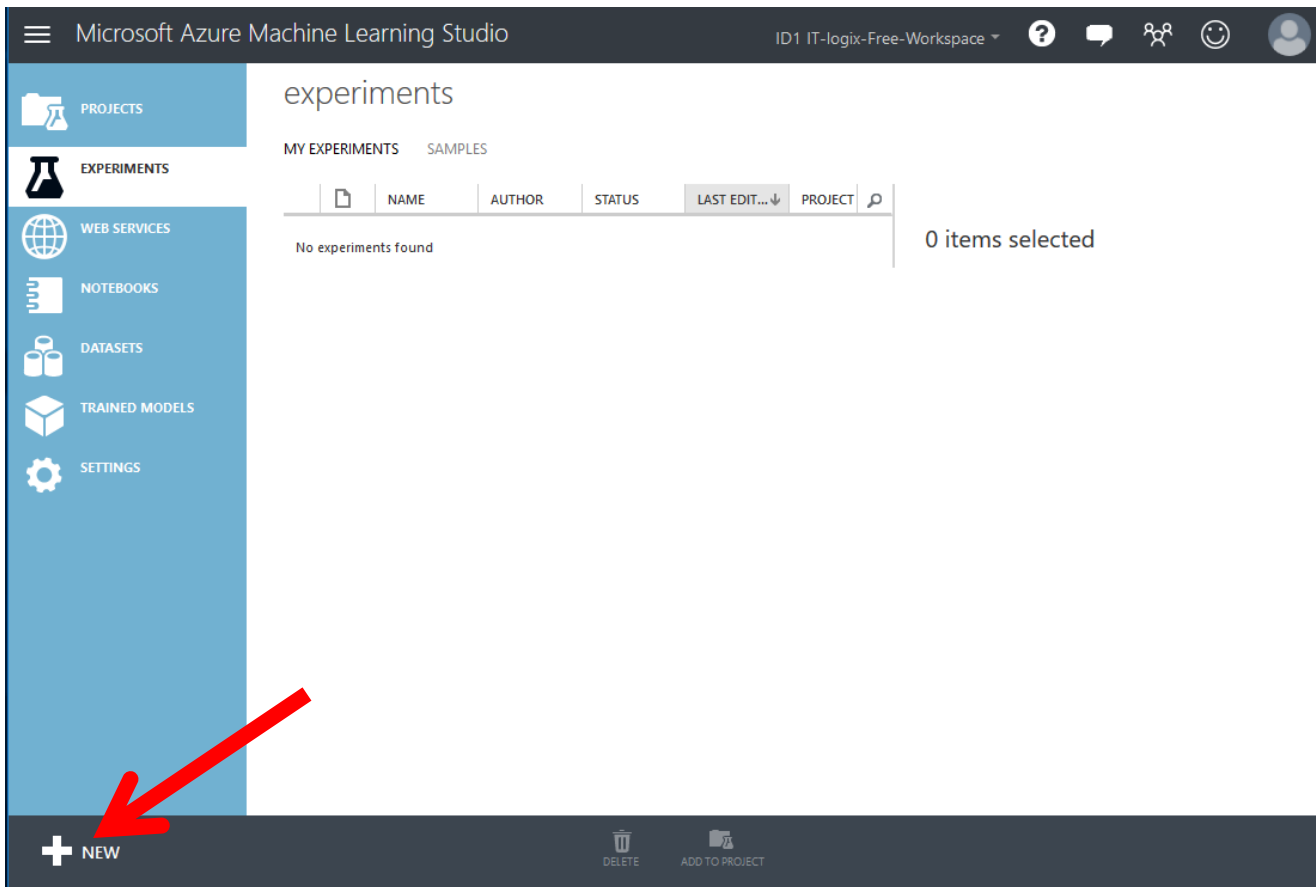
- Create a predictive model (pg. 14 - 24)

Primary focus of this session



CREATE A “NEW EXPERIMENT”

- Create a “New Experiment”



CREATE A “NEW EXPERIMENT”

➤ ML Studio

The screenshot displays the Microsoft Azure Machine Learning Studio interface. At the top, the title bar reads "Microsoft Azure Machine Learning Studio" and "ID0 IT-logix-Free-Workspace". The main workspace is titled "Experiment created on 6/23/2017" and "In draft". A central instruction says "To create your experiment, drag and drop datasets and modules here". A dashed-line flowchart shows a sequence of modules: a "Data" module (represented by a cylinder icon), followed by two "Data Transformation" modules (represented by document icons), and finally a "Machine Learning" module (represented by a flask icon). A callout box labeled "Drag Items Here" points to the first module. On the left, a sidebar lists various experiment items: Saved Datasets, Trained Models, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, Text Analytics, Time Series, Web Service, and Deprecated. On the right, a "Properties" panel is visible, containing sections for "Experiment Properties" (STATUS CODE: InDraft), "Summary" (a text area for a brief description), and "Description" (a text area for a detailed description). At the bottom, a toolbar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

EXPLORE / VISUALIZE MPG-DATA

➤ New experiment: Drag & drop example

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top bar shows "Microsoft Azure Machine Learning Studio" and "ID0 IT-logix-Free-Workspace". The main workspace is titled "Experiment created on 6/23/2017" and is in "In draft" status. On the left, a sidebar contains a search bar with "MPG" and a tree view with "Saved Datasets", "Samples", and "Trained Models". The "Samples" folder is expanded, showing "MPG data for various au...". A red arrow points from this item to the main workspace. In the workspace, a card labeled "MPG data for various autom..." with a circled "1" is visible. Another red arrow points from the search bar to this card. The right sidebar shows "Properties" and "Project" tabs, with "Properties" selected, displaying details for "MPG data for various automobil...": SUBMITTED BY: Microsoft C..., SIZE: 17.4 KB, FORMAT: GenericCSV, CREATED ON: 4/9/2015 1..., and a "View dataset" link. A "Quick Help" section at the bottom right provides context: "This data can be used to predict the fuel economy of automobiles based on various information, such as fuel economy (MPG), number of cylinders, engine displacement, horsepower, total weight, ...". The bottom toolbar includes icons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

ABOUT THE USER GUIDE

- Overview (pg. 3-4)

Next 2 slides

- Create a Workspace and a “New Experiment” (pg. 5-9)

Almost Done

- Explore/Visualize MPG-Data (pg. 10 - 13)

Quickly / Skip

- Create a predictive model (pg. 14 - 24)

Primary focus of this session



EXPLORE / VISUALIZE MPG-DATA

➤ Explore/Visualize MPG-Data (pg. 10 - 13)



MPG	Cyl	Displacement	Horsepower	Weight	Acceleration	Year	CountryCod	
18	8	307	130	3504	12	70	1	chevrolet chevelle malibu
15	8	350	165	3693	11.5	70	1	buick skylark 320
18	8	318	150	3436	11	70	1	plymouth satellite
16	8	304	150	3433	12	70	1	amc rebel sst
17	8	302	140	3449	10.5	70	1	ford torino
15	8	429	198	4341	10	70	1	ford galaxie 500
14	8	454	220	4354	9	70	1	chevrolet impala
14	8	440	215	4312	8.5	70	1	plymouth fury iii
14	8	455	225	4425	10	70	1	pontiac catalina

ABOUT THE USER GUIDE

- Overview (pg. 3-4)

Next 2 slides

- Create a Workspace and a “New Experiment” (pg. 5-9)

Almost Done

- Explore/Visualize MPG-Data (pg. 10 - 13)

Quickly / Skip

- Create a predictive model (pg. 14 - 24)

Primary focus of this session



CREATE A PREDICTIVE MODEL

➤ Split Dataset (pg. 14-16)

The image shows a screenshot of the Azure Machine Learning interface. On the left, a pipeline is visible with a step named 'Split Data' highlighted with a blue border and a circled '1' next to it. A red arrow points to the 'Split Data' step. The pipeline also includes a data source 'MPG data for various autom...' and a 'Split Data' step with a circled '2' next to it. On the right, the 'Properties' pane is open, showing the configuration for the 'Split Data' step. The 'Splitting mode' is set to 'Split Rows'. The 'Fraction of rows in the first split' is set to '0.75', which is highlighted with a red arrow. The 'Randomized split' checkbox is checked. The 'Random seed' is set to '0'. The 'Stratified split' is set to 'False'.

In draft

MPG data for various autom...

Split Data

1

2

Properties Project

Split Data

Splitting mode

Split Rows

Fraction of rows in the first split

0.75

Randomized split

Random seed

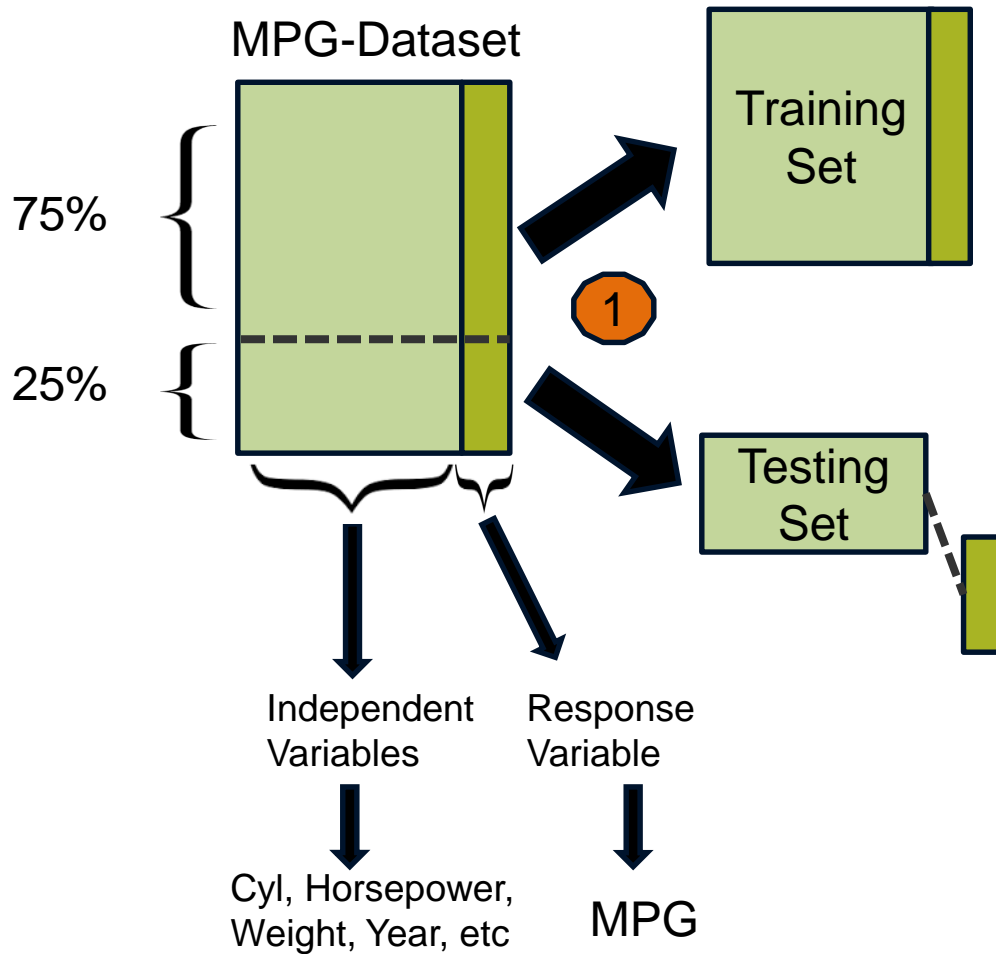
0

Stratified split

False

CREATE A PREDICTIVE MODEL

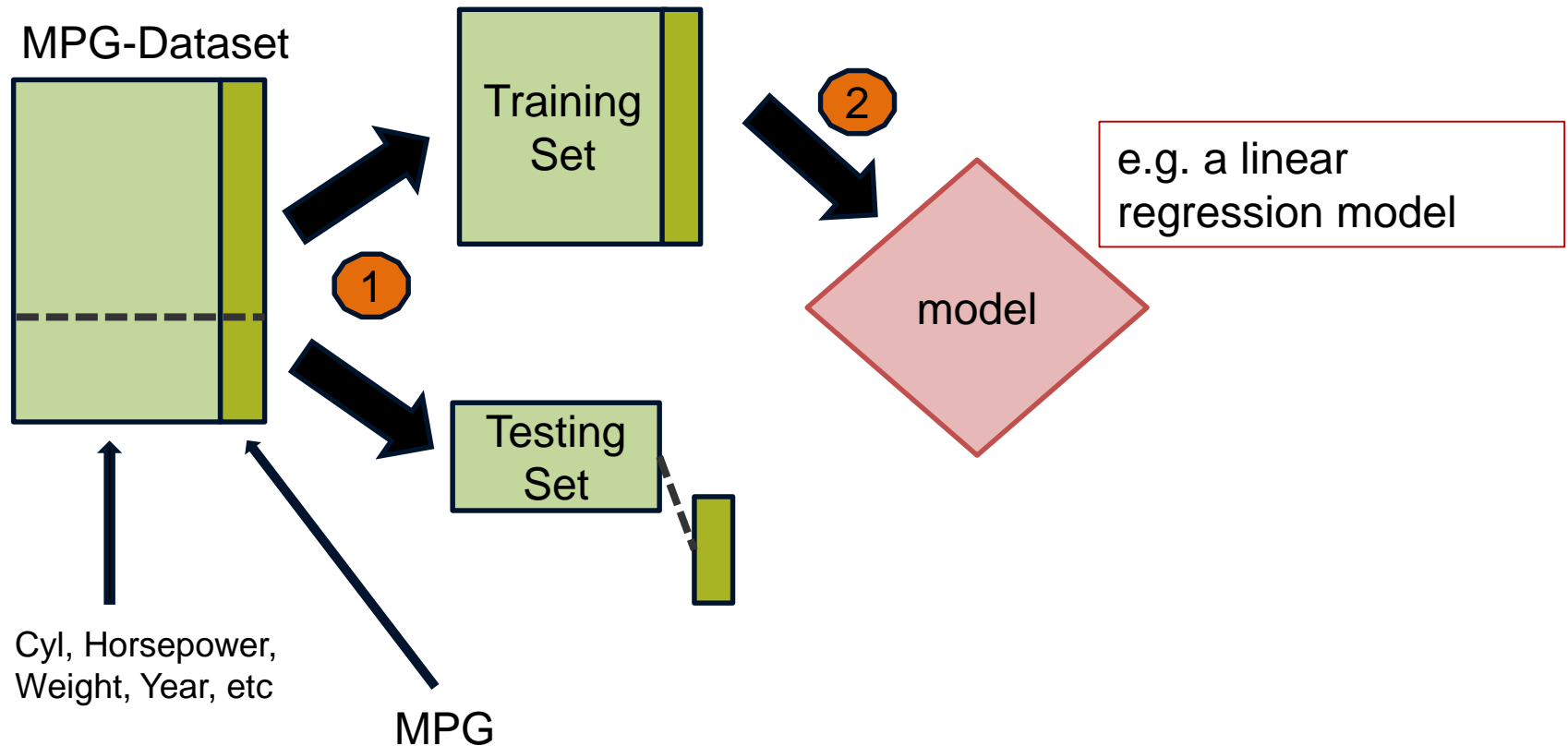
➤ Split Dataset (pg. 14-16)



MPG	Cyl	Displacement	Horsepower	Weight	Acceleration
18	8	307	130	3504	12
15	8	350	165	3693	11.5
18	8	318	150	3436	11
16	8	304	150	3433	12
17	8	302	140	3449	10.5
15	8	429	198	4341	10
14	8	454	220	4354	9
14	8	440	215	4312	8.5
14	8	455	225	4425	10

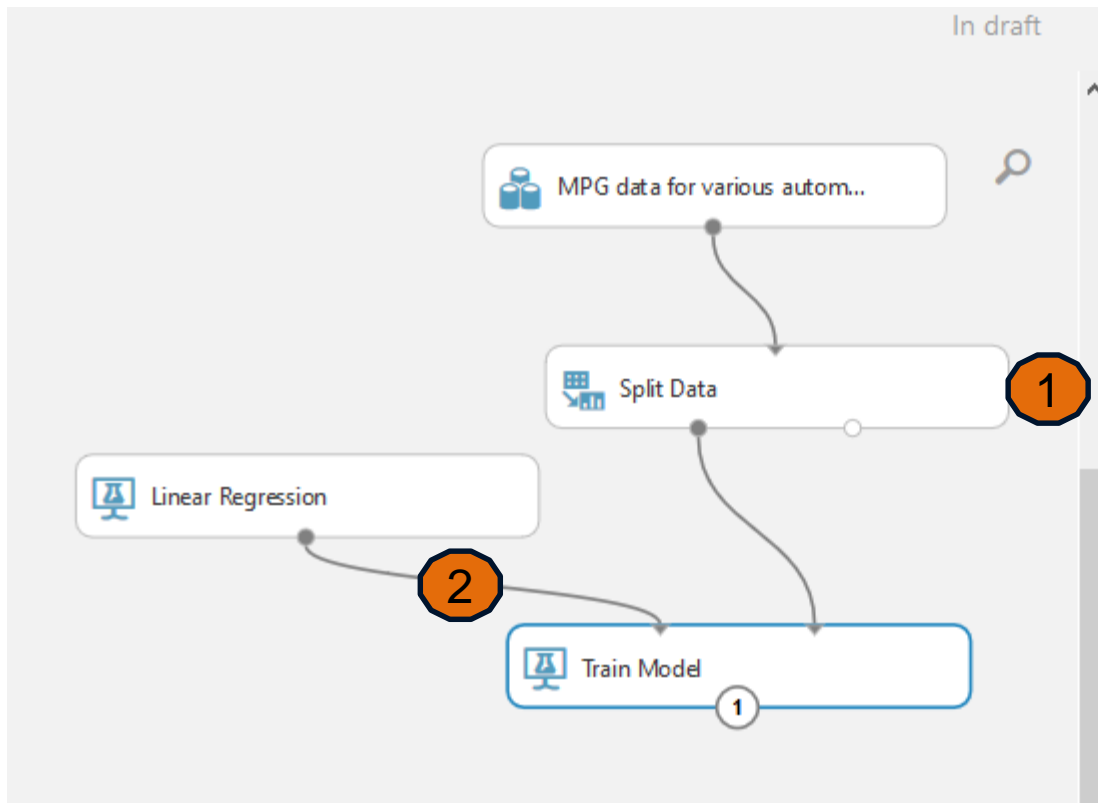
CREATE A PREDICTIVE MODEL

- Train a Model (pg. 16-19)



CREATE A PREDICTIVE MODEL

- Train a Model (pg. 16-19)



Properties Project

Train Model

Label column

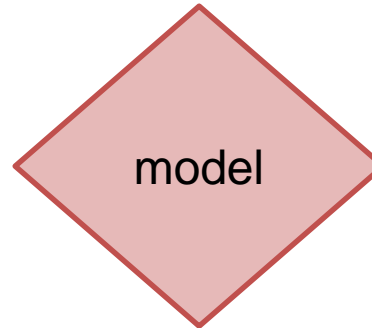
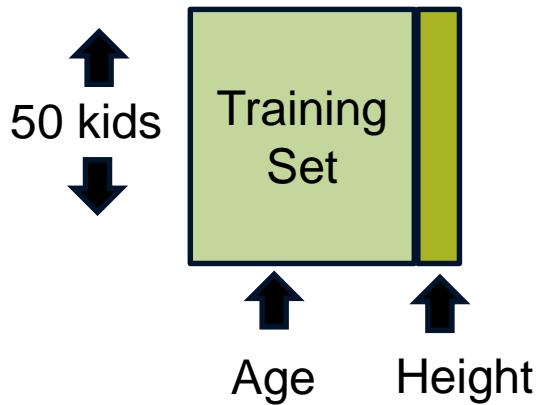
Selected columns:
Column names: MPG

Launch column selector



CREATE A PREDICTIVE MODEL

➤ Train a Model (pg. 16-19)



$$y = a_0 + a_1 * x$$

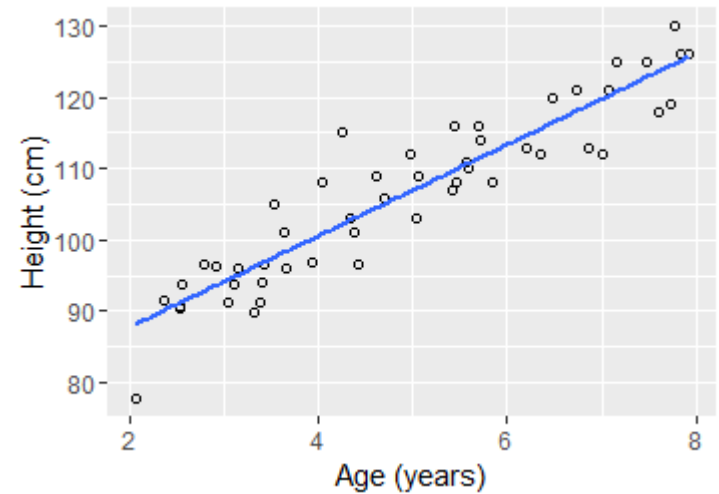
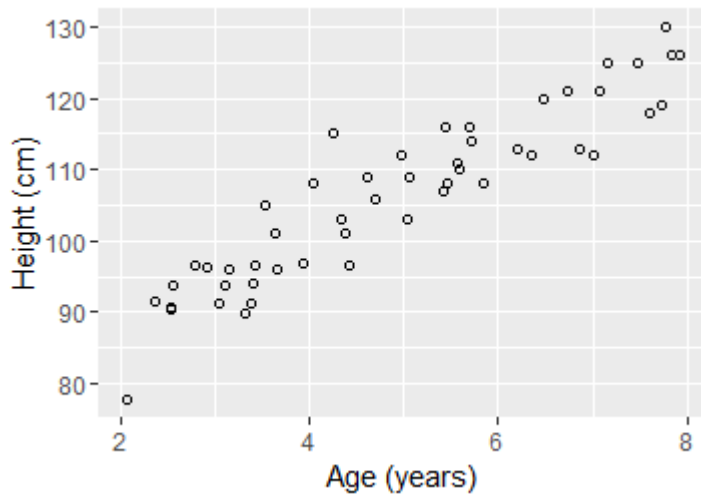
↓

$$\text{Height} = a_0 + a_1 * \text{Age}$$

↓

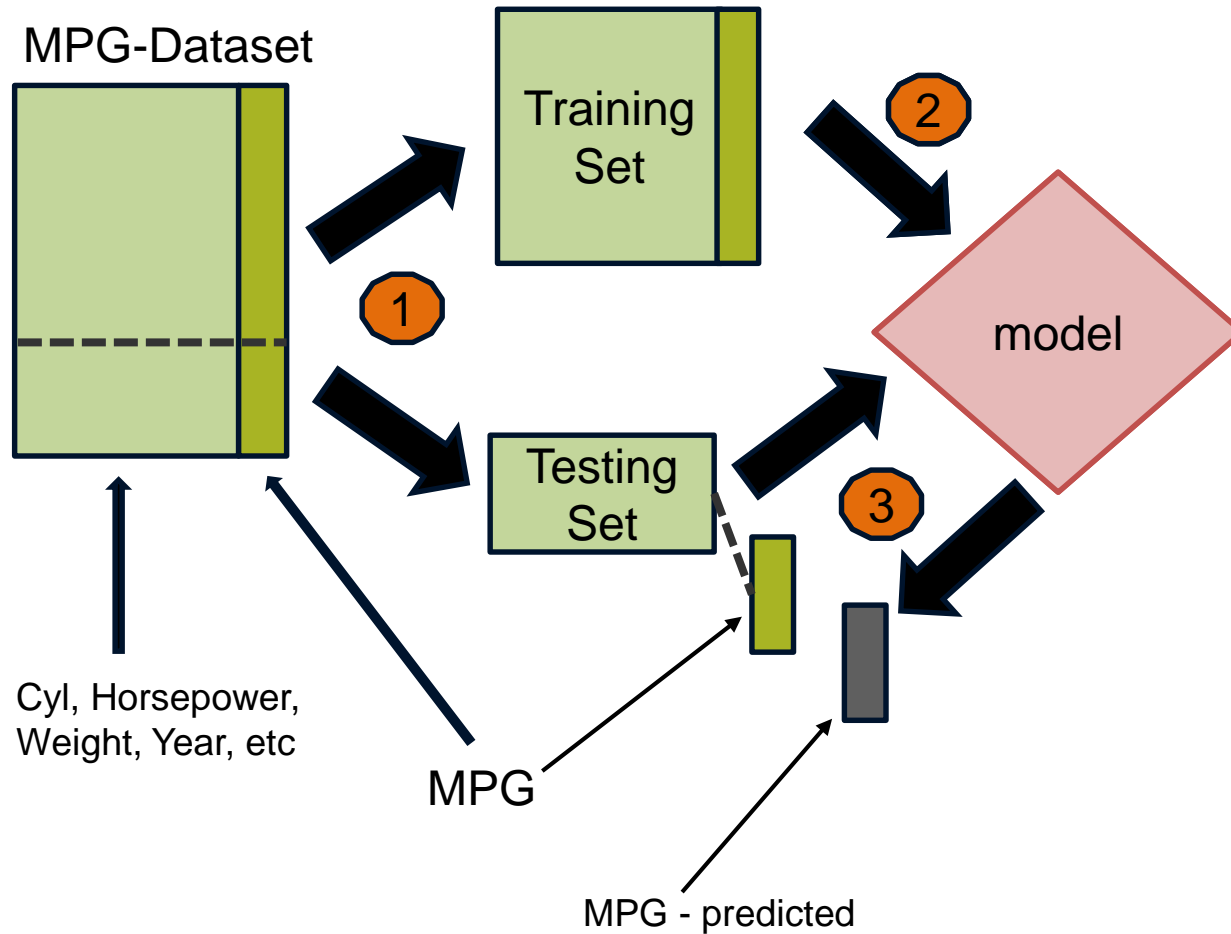
Find a_0, a_1

$$\text{Height} = 75 + 6.4 * \text{Age}$$



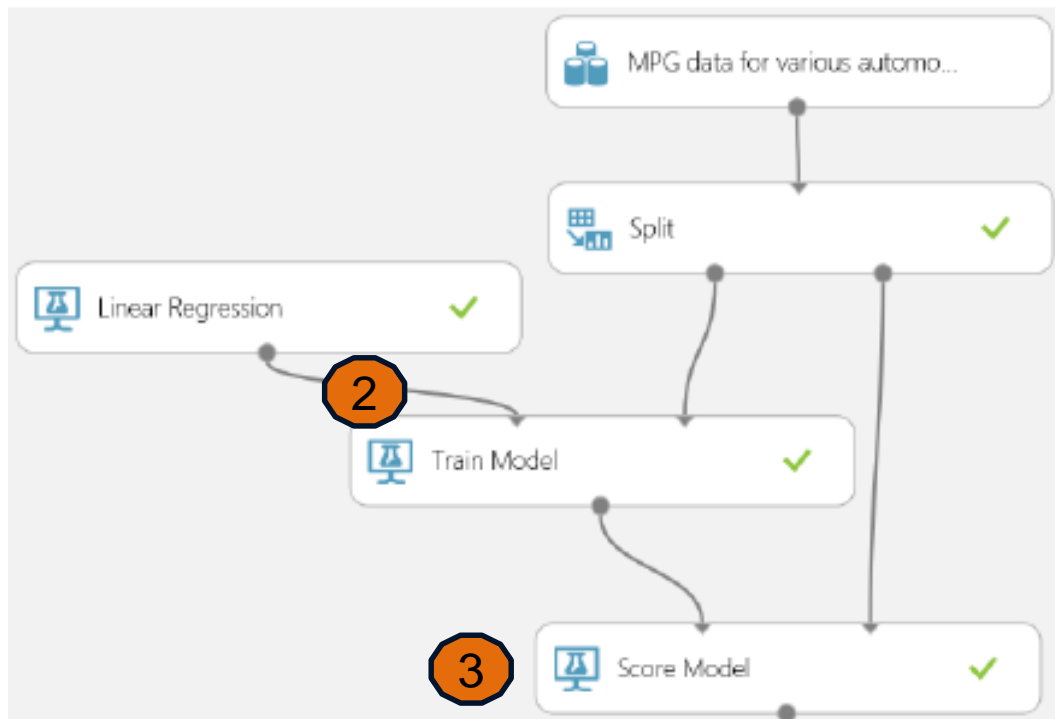
CREATE A PREDICTIVE MODEL

- Score the model (pg. 19-22)



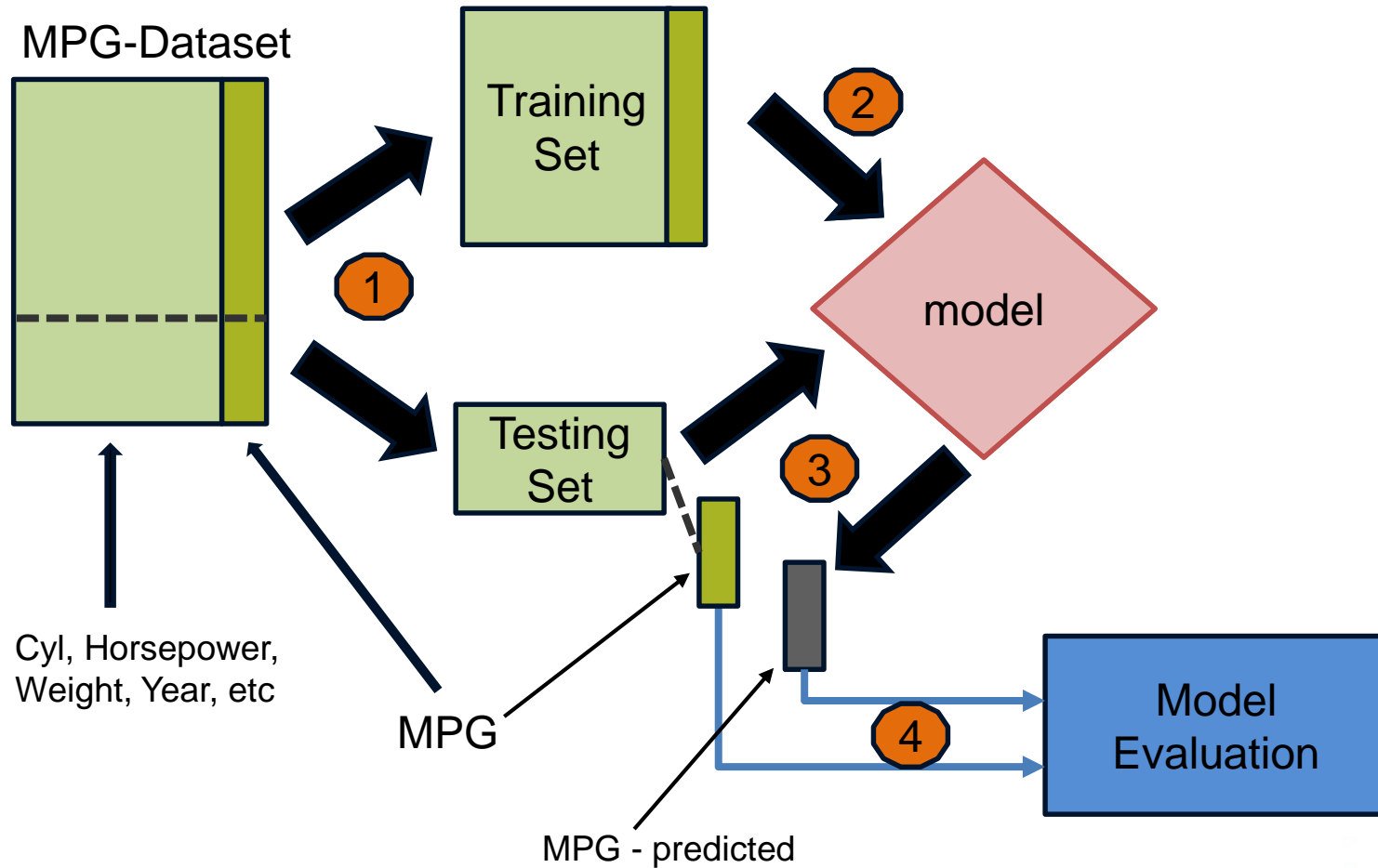
CREATE A PREDICTIVE MODEL

- Score the model (pg. 19-22)



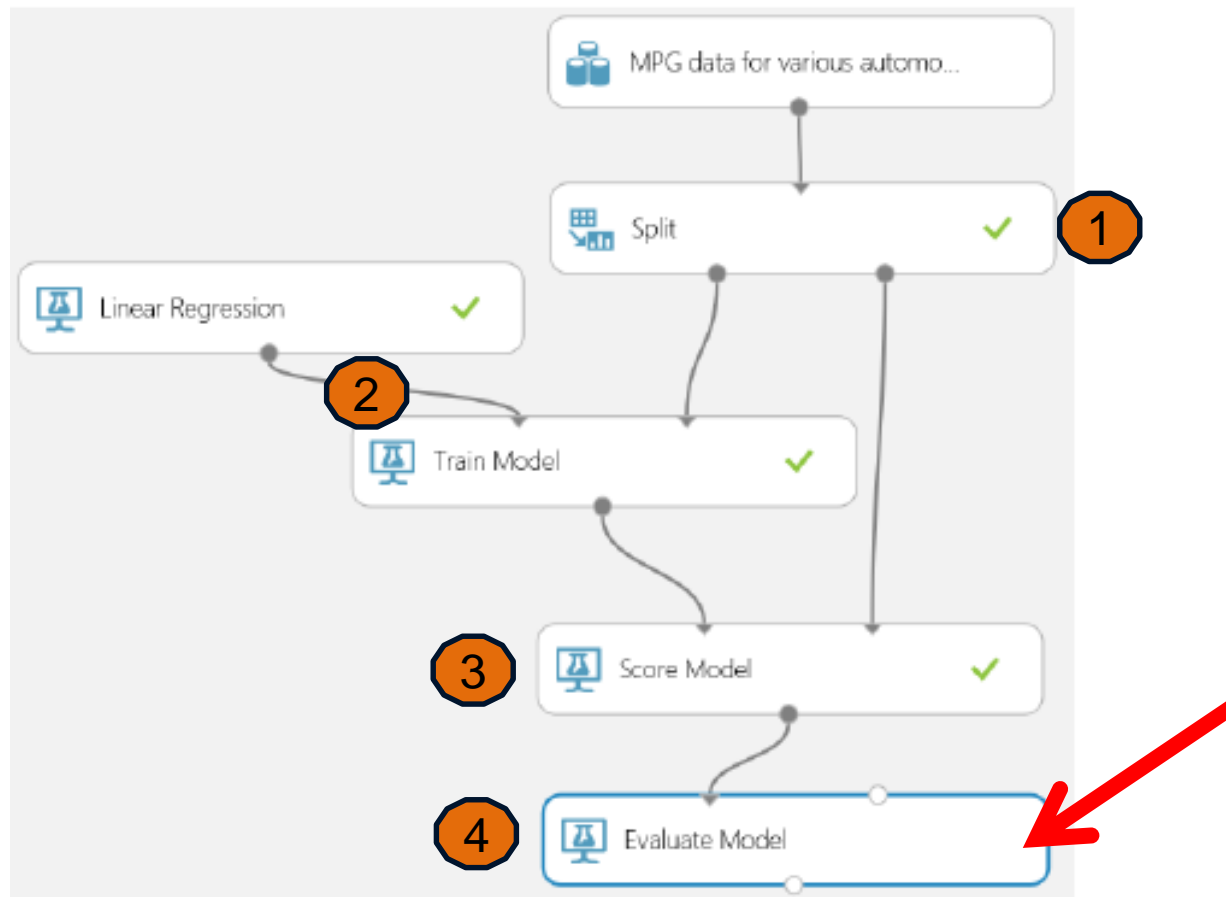
CREATE A PREDICTIVE MODEL

- Evaluate the model (pg. 22-24)



CREATE A PREDICTIVE MODEL

- Evaluate the model (pg. 22-24)



CREATE A PREDICTIVE MODEL

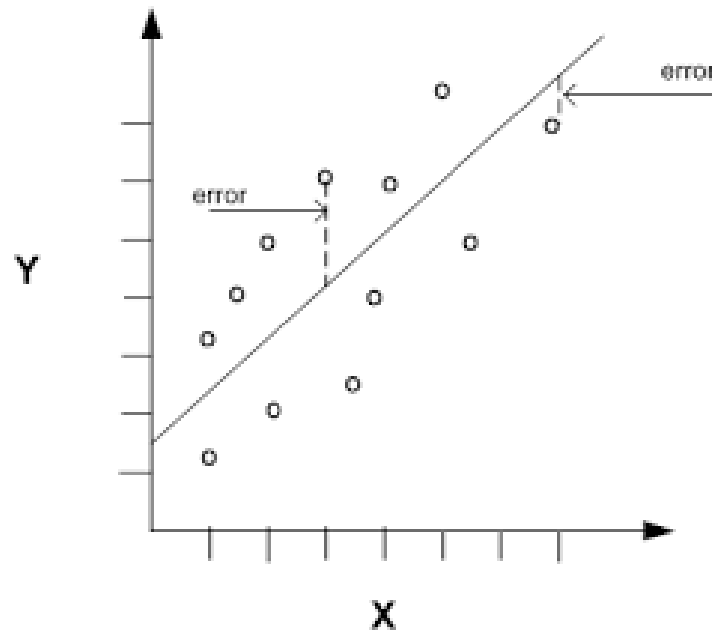
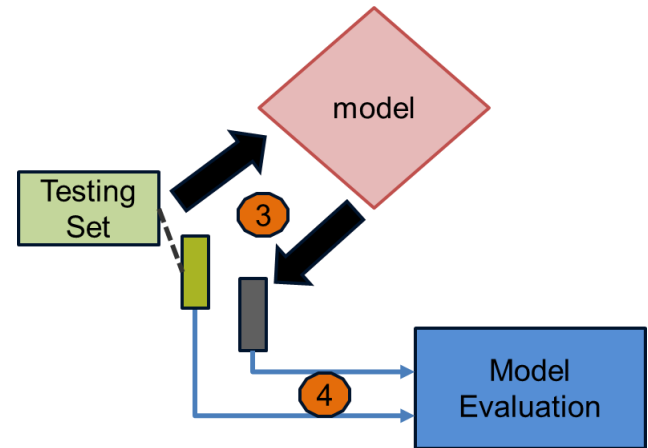
➤ Model Evaluation Metrics

AzureML-Lab1 ➤ Evaluate Model ➤ Evaluation results

▸ Metrics

Mean Absolute Error	2.656148
Root Mean Squared Error	3.439296
Relative Absolute Error	0.383569
Relative Squared Error	0.174408
Coefficient of Determination	0.825592

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$



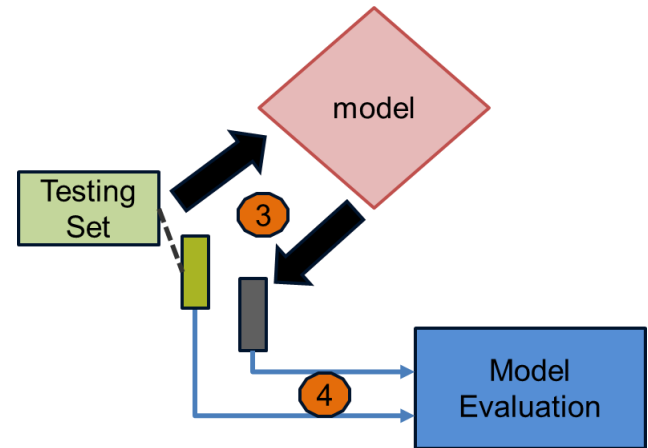
CREATE A PREDICTIVE MODEL

► Model Evaluate Metrics

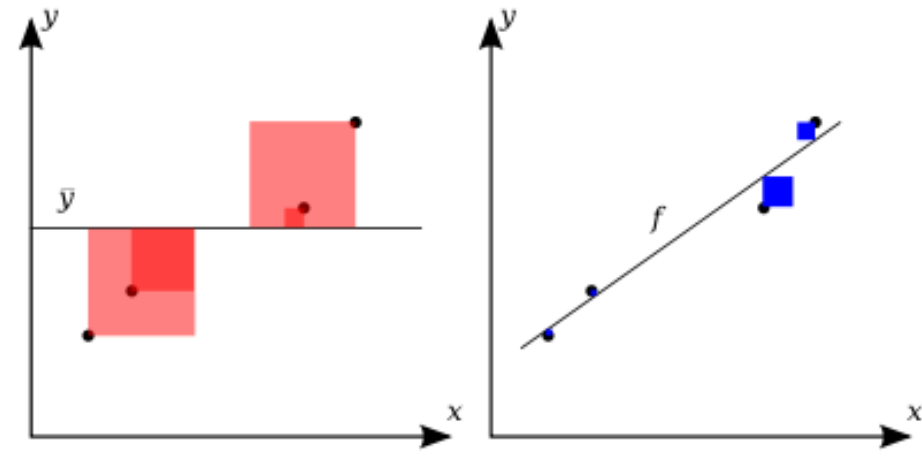
AzureML-Lab1 ► Evaluate Model ► Evaluation results

▸ Metrics

Mean Absolute Error	2.656148
Root Mean Squared Error	3.439296
Relative Absolute Error	0.383569
Relative Squared Error	0.174408
Coefficient of Determination	0.825592

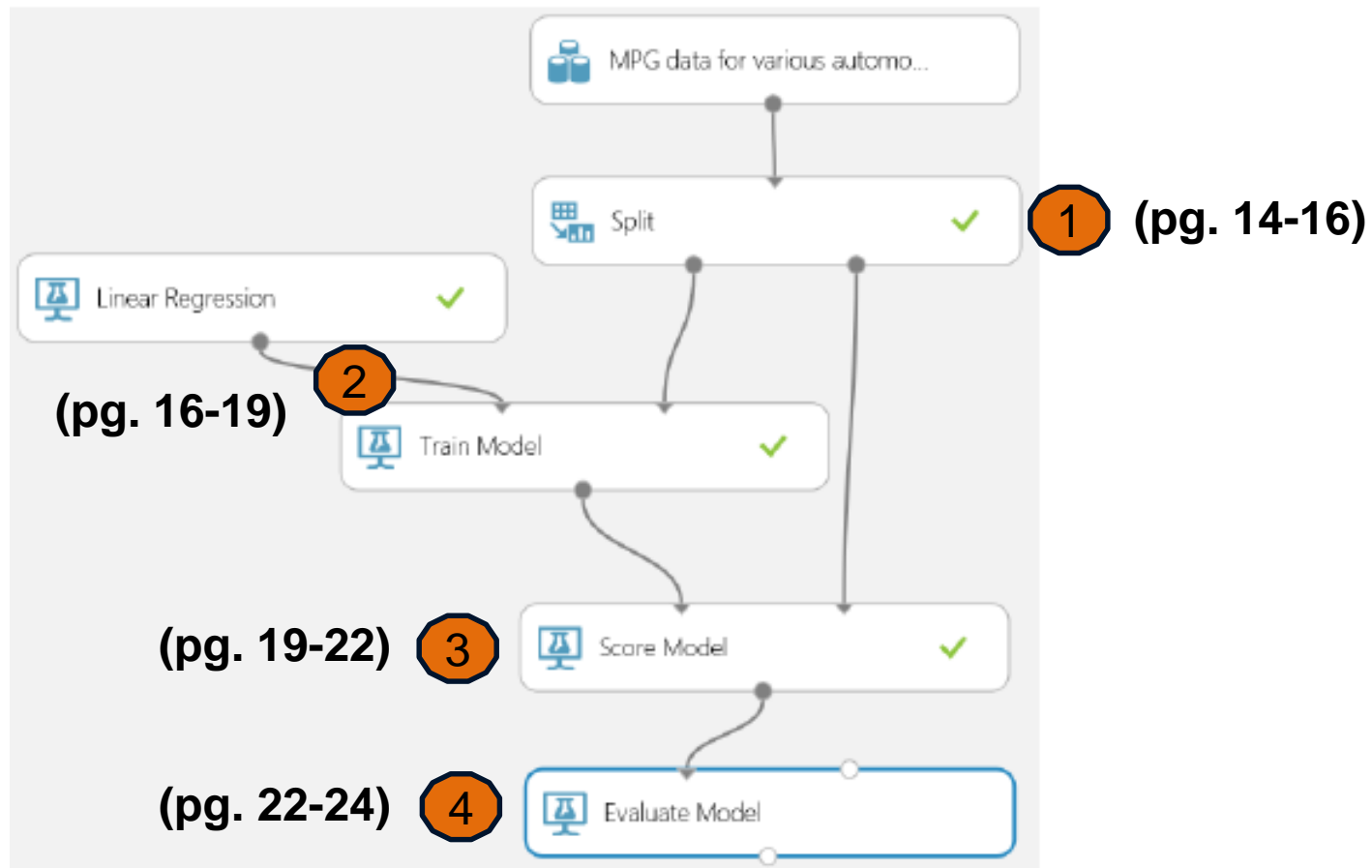


$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$



CREATE A PREDICTIVE MODEL

➤ Overview



Wir freuen uns auf angeregte Gespräche mit Ihnen

- Dr. Sotiris Dimopoulos
Senior Data Science Consultant

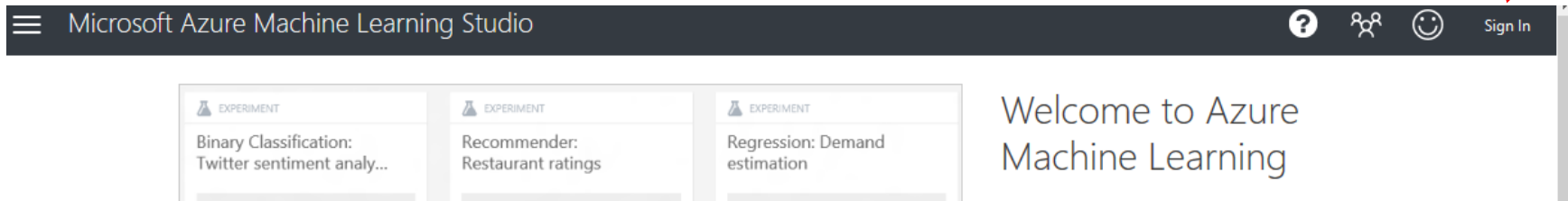
 Follow us: @itlogixag

 <http://blog.it-logix.ch>

AZURE MACHINE LEARNING

➤ Sign in

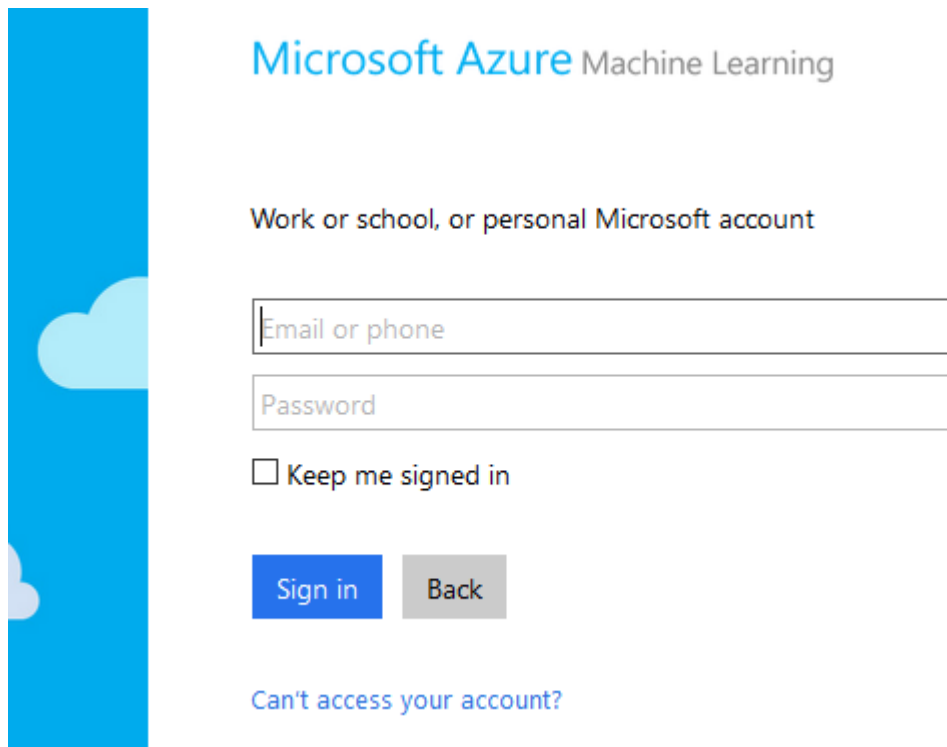
- ❑ Go to the ML Studio website by typing the address:
<http://studio.azureml.net>
- ❑ Click Sign In on the top right corner of the web page



AZURE MACHINE LEARNING

➤ Sign in

- ❑ Enter the email address and password associated with your Microsoft ID, and click the Sign In button.



Microsoft Azure Machine Learning

Work or school, or personal Microsoft account

Email or phone

Password

Keep me signed in

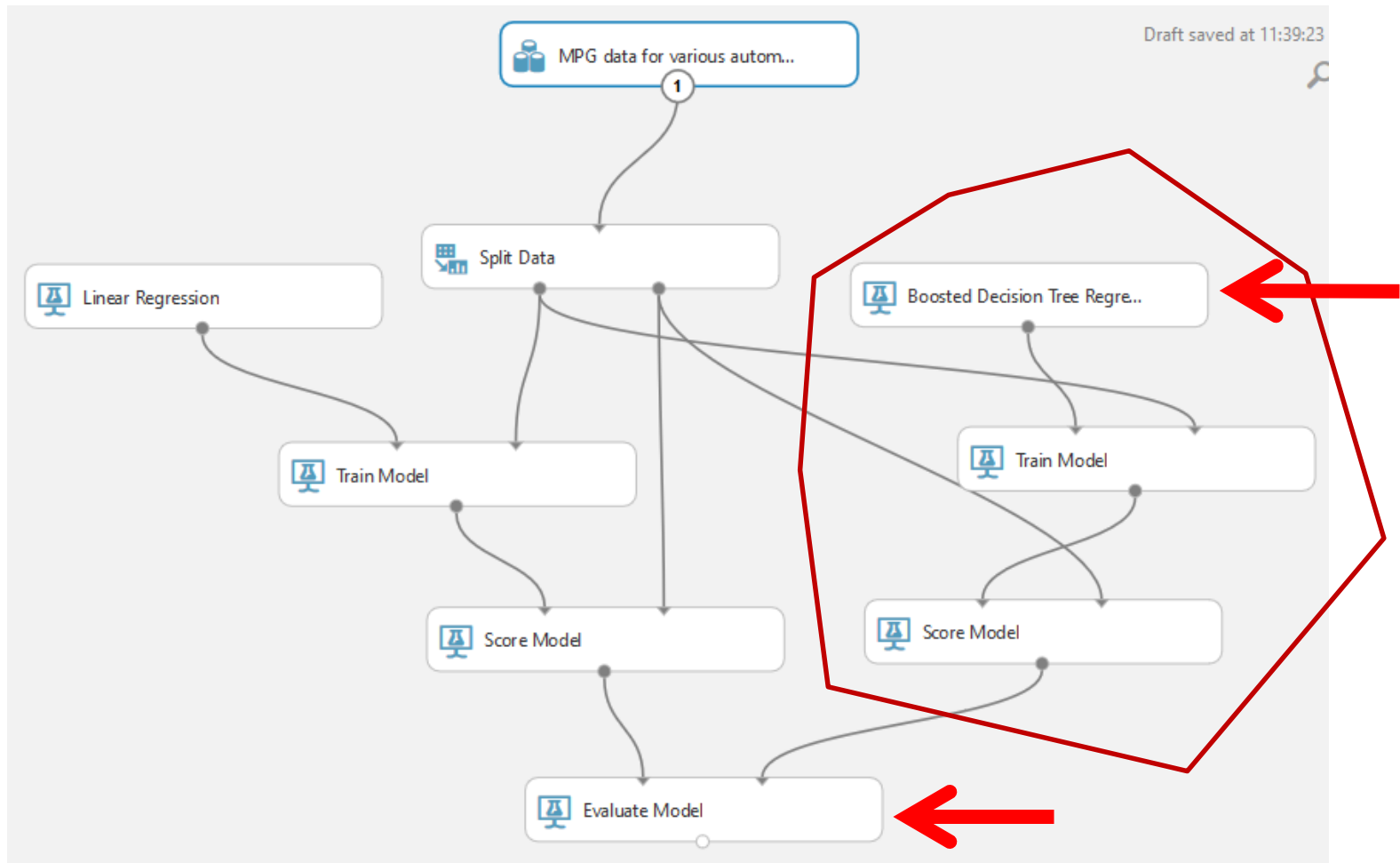
Sign in Back

[Can't access your account?](#)

Usr: MAKEBI-IDX@outlook.com
Pswd: makeBlidX
X: your ID number

EXTRA...

- Repeat with a different type of model & Compare models



EXTRA...

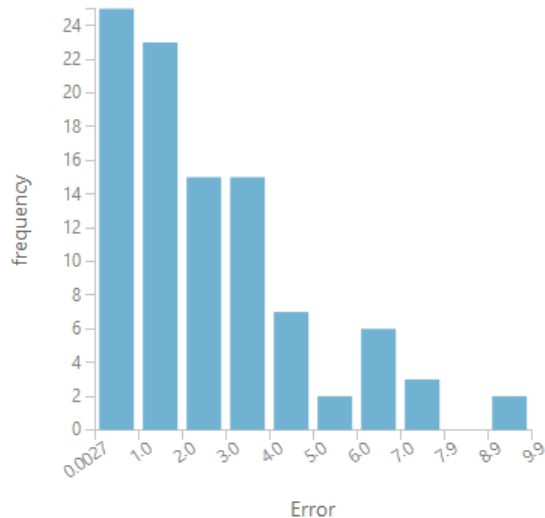
➤ Compare Models

AzureML-Lab1 ➤ Evaluate Model ➤ Evaluation results

▾ Metrics

Mean Absolute Error	2.656148
Root Mean Squared Error	3.439296
Relative Absolute Error	0.383569
Relative Squared Error	0.174408
Coefficient of Determination	0.825592

▾ Error Histogram



▾ Metrics

Mean Absolute Error	2.175115
Root Mean Squared Error	2.884716
Relative Absolute Error	0.314104
Relative Squared Error	0.122697
Coefficient of Determination	0.877303

▾ Error Histogram

