



Using the Cloud for Machine Learning with Azure ML

数据科学 -
使用 **Azure ML**

July 2020
Microsoft Reactor | Ryan Chung

```
led by player to
            .load_image("kg.png")
        (self):
            alize Dog object and create Text o
g, self).__init__(image = Dog.image
                    x = games.mouse.x
                    bottom = games.screen.height)
            re = games.Text(value = 0, size = 24,
                            top = 5, right = game
            screen.add(self.score)
            t = games.Text(value = 0, size = 24,
                           top = 5, left = game
            screen.add(t)
```



Ryan Chung

Instructor / DevelopIntelligence
Founder / MobileDev.TW

@ryanchung403 on WeChat
Ryan@MobileDev.TW





Reactor



developer.microsoft.com/reactor/
@MSFTReactor on Twitter

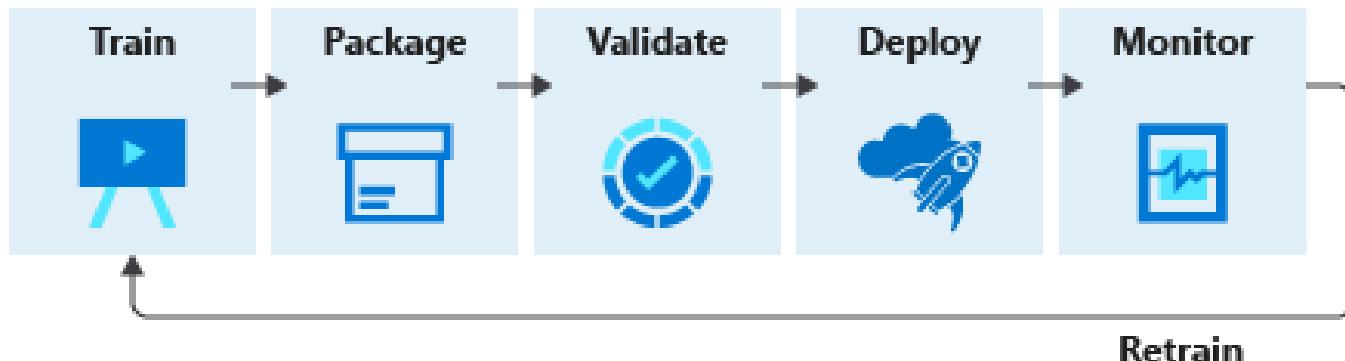
DS On-line Workshop agenda **数据科学在线研讨会议程**

19:30	Welcome 开场
19:35	Overview of Azure Machine Learning 微软新一代机器学习解决方案
20:00	Case : Predict automobile price 实作 : 汽车价格预测
20:20	5 -minute lab break 中场休息 / 实作练习
20:25	Case : Predict income 实作 : 收入预测
20:45	Automated ML in Azure Machine Learning 使用自动化ML建立分类
21:00	Event end 研讨会结束

Azure Machine Learning 微软 Azure 机器学习

- 云端环境
- 可以进行模型的训练/部署/自动化/管理/追踪
- 适用于
 - 传统机器学习 / 深度学习 / 监督式学习 / 非监督式学习
- 使用弹性
 - 可自行撰写Python/R 或 使用Azure ML 图形化界面

Azure Machine Learning Model Workflow



Azure Machine Learning Designer

- 微软 Azure 机器学习设计师
- 最新推出，以拖拉方式建立机器学习流程

Jupyter Notebook

- 可以在云端上使用，上面有许多范例也可以自己从头建立

VS Code Extension 插件

- 可结合本地端运行

CLI Extension

- 以指令列方式使用

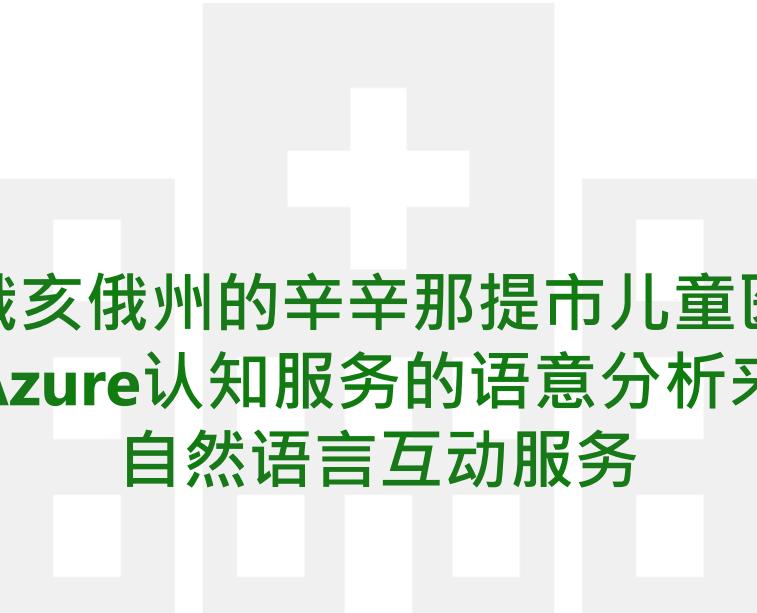
Reinforcement learning

- 实验中，使用Ray RLLib

优点与案例

为什么要使用云端平台上的机器学习?

- 可以快速、低成本地存取成熟的算法
- 可以将完成训练、调校的模型在在线或装置上重复使用



美国俄亥俄州的辛辛那提市儿童医院，
使用**Azure**认知服务的语意分析来开发
自然语言互动服务

Azure Machine Learning Designer 拖拉式界面

The screenshot shows the Azure Machine Learning Designer interface. On the left, there's a sidebar with navigation links: New, Home, Designer (which is selected), Notebooks, Automated ML, Datasets, Experiments, Pipelines, Models, and Endpoints. Below these are sections for Author and Assets, each with a 'Search' bar and a list of items like Datasets, Data Input and Output, Data Transformation, etc. The main workspace is titled 'Flight Delays' and contains a data flow diagram. A 'Flight Delays Data' dataset is connected to a 'Normalize Data' step. A mouse cursor is hovering over the connection line between the two components.

```
graph TD; FlightDelaysData[Flight Delays Data] --> NormalizeData[Normalize Data]
```

Azure 机器学习

歡迎使用工作室!

X

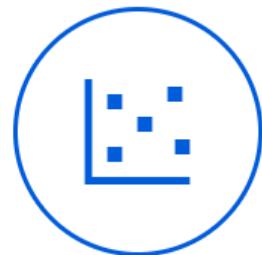
Azure Machine Learning 可協助您建置、訓練、部署及管理雲端等級的模型。



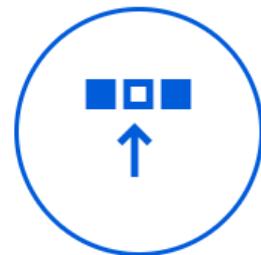
登錄資料



訓練模型



評估模型



部署模型

從儲存體參考資料，以便在模型訓練期間輕鬆存取並使用摘要統計資料來探索。

使用具備訓練資料的機器學習演算法，建立模型。

使用測試資料尋找最佳模型。

將模型部署在 Azure 雲端中成為 Web 服務，或是部署到 IoT Edge 裝置。

建立工作区

- 建立资源 +

The screenshot shows the Microsoft Azure portal's home page. At the top, there is a navigation bar with back, forward, and search icons, followed by the URL 'portal.azure.com/#home'. Below the navigation bar is a blue header bar with the Microsoft Azure logo and a search bar labeled '搜尋資源、服務及文件 (G+ /)'. The main content area is titled 'Azure 服務' and features several service icons: '建立資源' (Create Resource), '機器學習' (Machine Learning), '成本管理 + 計費' (Cost Management + Billing), '認知服務' (Cognitive Services), '資源群組' (Resource Groups), '所有資源' (All Resources), and 'Bot 服務' (Bot Services). A large blue arrow points from the '建立資源' button to the '更多服務' (More Services) link at the bottom left.

<https://portal.azure.com/>

Azure Marketplace

- AI+机器学习服务 -> 机器学习

Microsoft Azure 搜尋資源、服務及文件 (G+/-) 首頁 >

新增

搜尋 Marketplace

Azure Marketplace [查看全部](#) 精選 [查看全部](#)

開始使用

最近建立

AI + 機器學習服務

分析

 [機器學習](#)
[深入了解](#)

 [Web App Bot](#)
[快速入門 + 教學課程](#)

10

建立机器学习工作区

- 订用帳戶、工作区名称、区域、工作区版本(企业)
- 标签可略过
- 按下[检阅+建立]

首頁 > 新增 >

機器學習

建立機器學習工作區

選取用以管理部署資源及成本的訂用帳戶。使用像資料夾這樣的資源群組來安排及管理您的所有資源。

訂用帳戶 * ⓘ

Azure for Students

資源群組 * ⓘ

cloud-shell-storage-southeastasia

新建

工作區詳細資料

指定工作區的名稱、區域和版本。

工作區名稱 * ⓘ

Hello_ML_Designer

區域 * ⓘ

東南亞

工作區版本 * ⓘ

企業

檢閱 + 建立

< 上一步

下一步: 標籤

汽车价格预测

- 利用马力、引擎种类、车门数目、款式、驱动、轴距、车子长度、车子宽度等，来预测价格



Source : www.inovex.de

Microsoft Azure Machine Learning

- 设计工具 -> 立即启动

Microsoft Azure Machine Learning

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left, there's a sidebar with navigation options like '新增' (New), '首頁' (Home), '作者' (Author), '筆記本' (Notebook), '自動化 ML (預覽)' (Automated ML (Preview)), '設計工具 (預覽)' (Design Tools (Preview)), '資產' (Assets), '資料集' (Dataset), and '實驗' (Experiment). The '設計工具 (預覽)' option is highlighted with a red box. The main area has a title '歡迎使用工作室!' (Welcome to the Studio!). It features four cards: '建立新的' (Create new) with a plus sign, '筆記本' (Notebook) with a code icon, '自動化 ML (預覽)' (Automated ML (Preview)) with a gear icon, and '設計工具 (預覽)' (Design Tools (Preview)) with a cube icon. Each card has a '立即啟動' (Start now) button, with the one for '設計工具 (預覽)' also highlighted with a red box.

<https://ml.azure.com/>

Designer

- 新增易于使用的预建模块

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽)

設計工具 (預覽)

新增管線

易於使用的預建模組 ②

- HelloML
- 設計工具 (預覽)
- 新增
- 首頁
- 作者
- 筆記本
- Automated ML (preview)
- 設計工具 (預覽) (selected)
- 資產
- 資料集
- 實驗

修改项目名称

- Car Price Prediction

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽) > 正在撰寫

Search by name, tags and description

Car Price Prediction

Modules

Datasets Models

自動儲存開啟

100%

提交

設定

預設計算目標

選取計算目標以執行管線。

未選取任何計算目標

選取計算目標

管線參數

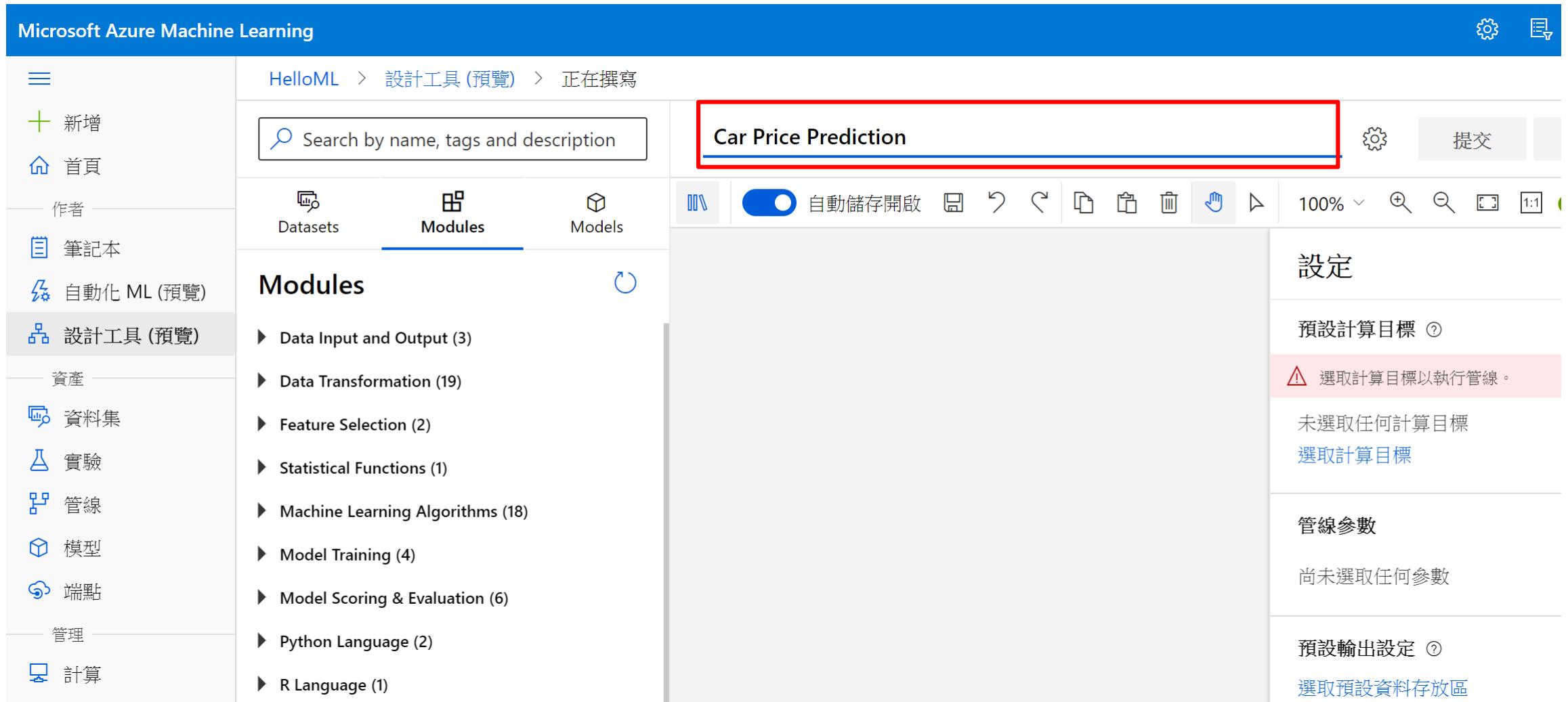
尚未選取任何參數

預設輸出設定

選取預設資料存放區

Modules

- ▶ Data Input and Output (3)
- ▶ Data Transformation (19)
- ▶ Feature Selection (2)
- ▶ Statistical Functions (1)
- ▶ Machine Learning Algorithms (18)
- ▶ Model Training (4)
- ▶ Model Scoring & Evaluation (6)
- ▶ Python Language (2)
- ▶ R Language (1)



运算资源设定

- 点击专案名称旁的「齿轮」
- 右下方点击「选取计算目标」

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽) > 正在撰寫

Car Price Prediction 

提交 發行 ...

新增 首頁 作者 筆記本 自動化 ML (預覽) **設計工具 (預覽)** 資產 資料集 實驗 管線 模型 端點

自動儲存開啟 100% 1:1

設定

預設計算目標 ⑦

⚠ 選取計算目標以執行管線。

未選取任何計算目標 **選取計算目標**

管線參數 +

尚未選取任何參數

设定计算目标

- 建立新的
- 输入计算目标名称

設定計算目標 ×

選取現有項目 建立新的

計算目標組態

建議您使用下方預先定義的組態，快速設定訓練計算目標。

組態	vCPU	RAM	儲存體	節點計數
Predefined	2	7GB	8GB	2

不符合您的需求嗎? 請前往 [\[計算\] -> \[訓練叢集\]](#)，以建立自訂計算目標。

新增計算目標名稱 [②](#)



HelloML

儲存

取消

新增数据集

- Datasets -> Samples
- Automobile price data(Raw)
- 拖曳至工作区

HelloML > 設計工具(預覽) > 正在撰寫

Search by name, tags and description

Car Price Prediction

Datasets Modules Models

自動儲存開啟

Modules

- ▶ Data Input and Output (3)
- ▶ Data Transformation (19)
- ▶ Feature Selection (2)
- ▶ Statistical Functions (1)
- ▶ Machine Learning Algorithms (18)
- ▶ Model Training (4)
- ▶ Model Scoring & Evaluation (6)
- ▶ Python Language (2)
- ▶ R Language (1)
- ▶ Text Analytics (7)

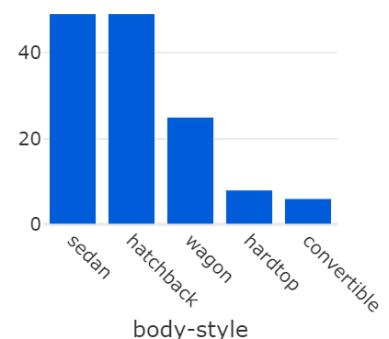
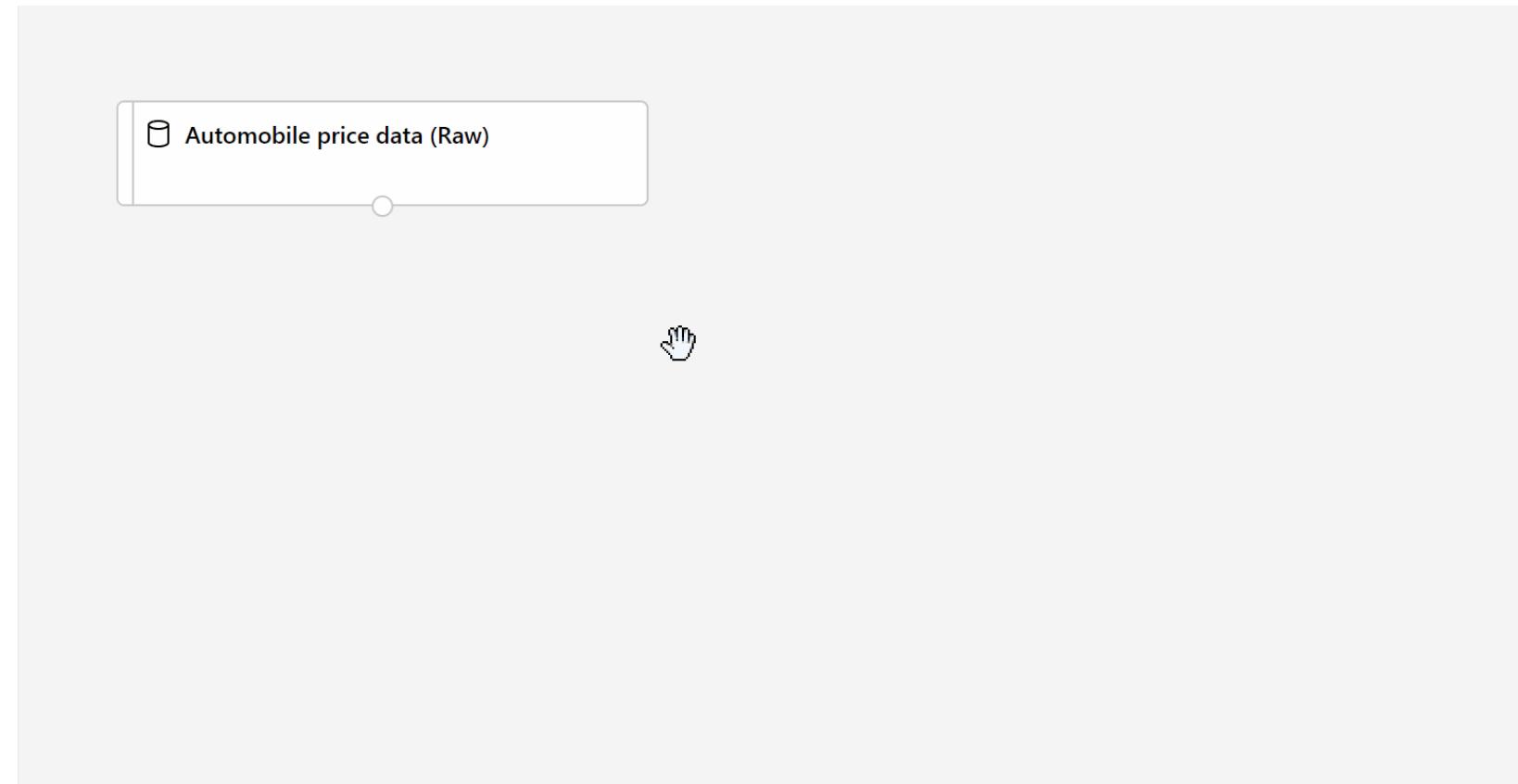
观察数据集

- 点击该数据集区块
- 右方选单分页->输出
- Dataset output -> 可视化

Automobile price data (Raw) 結果視覺效果

資料列 ②	資料行 ②
205	26

num-of-doors	body-style	drive-wheels	engine-location	wheel-base	length
two	convertible	rwd	front	88.6	109.9



检查遗漏值状况

- 发现normalized-losses有41个遗漏值

Automobile price data (Raw) 結果視覺效果

X

資料列 ② 資料行 ②

205 26



增加数据字段选取模块

- Data Transformation -> Select Columns in Dataset
- 拖曳至工作区
- 将目前两个模块连接

HelloML > 設計工具 (預覽) > 正在撰寫

Car Price Prediction

自動儲存開啟

Automobile price data (Raw)

Select Columns Transform

Select Columns in Dataset

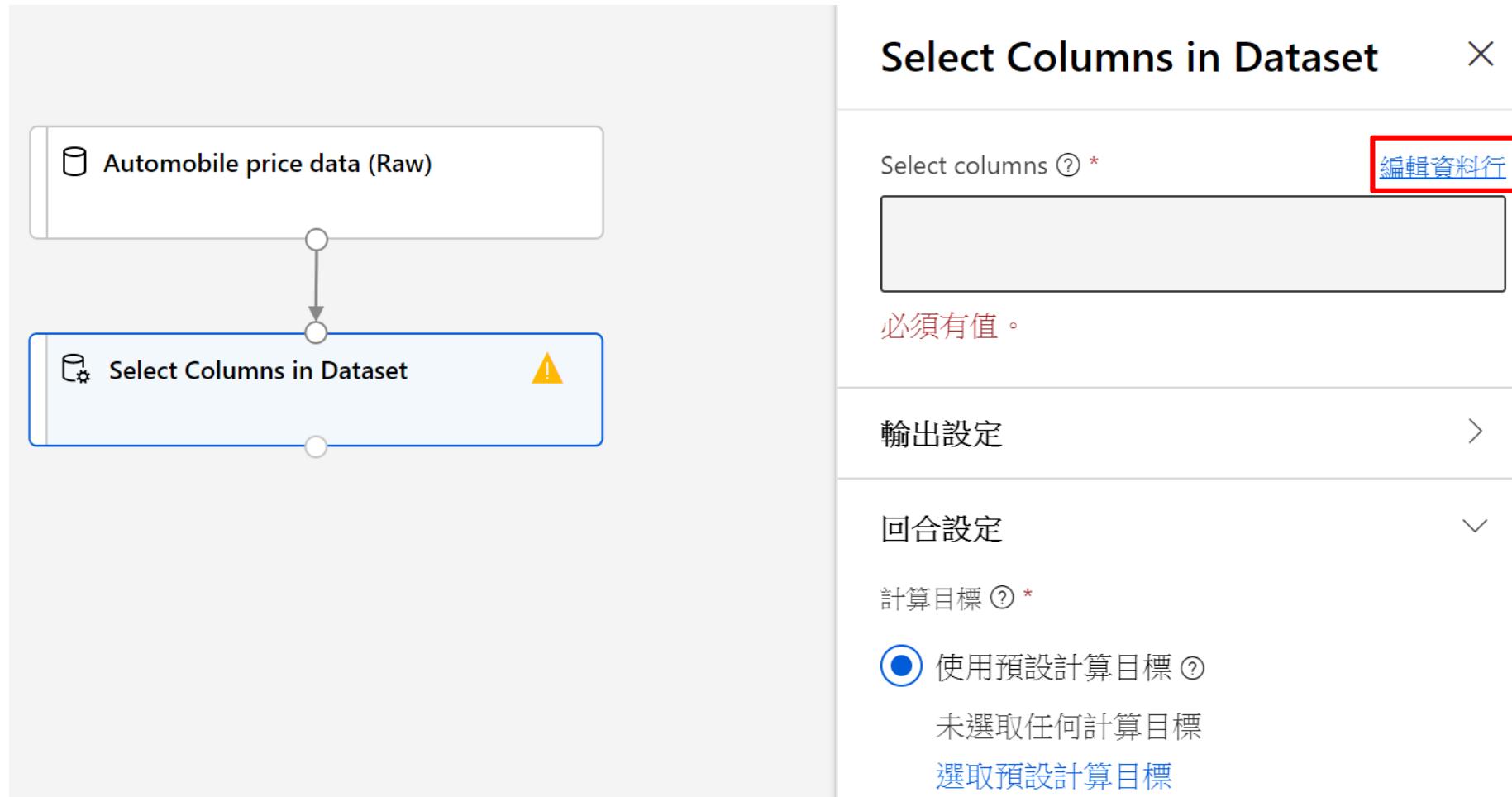
Select Columns in Dataset

Split Data

導覽器

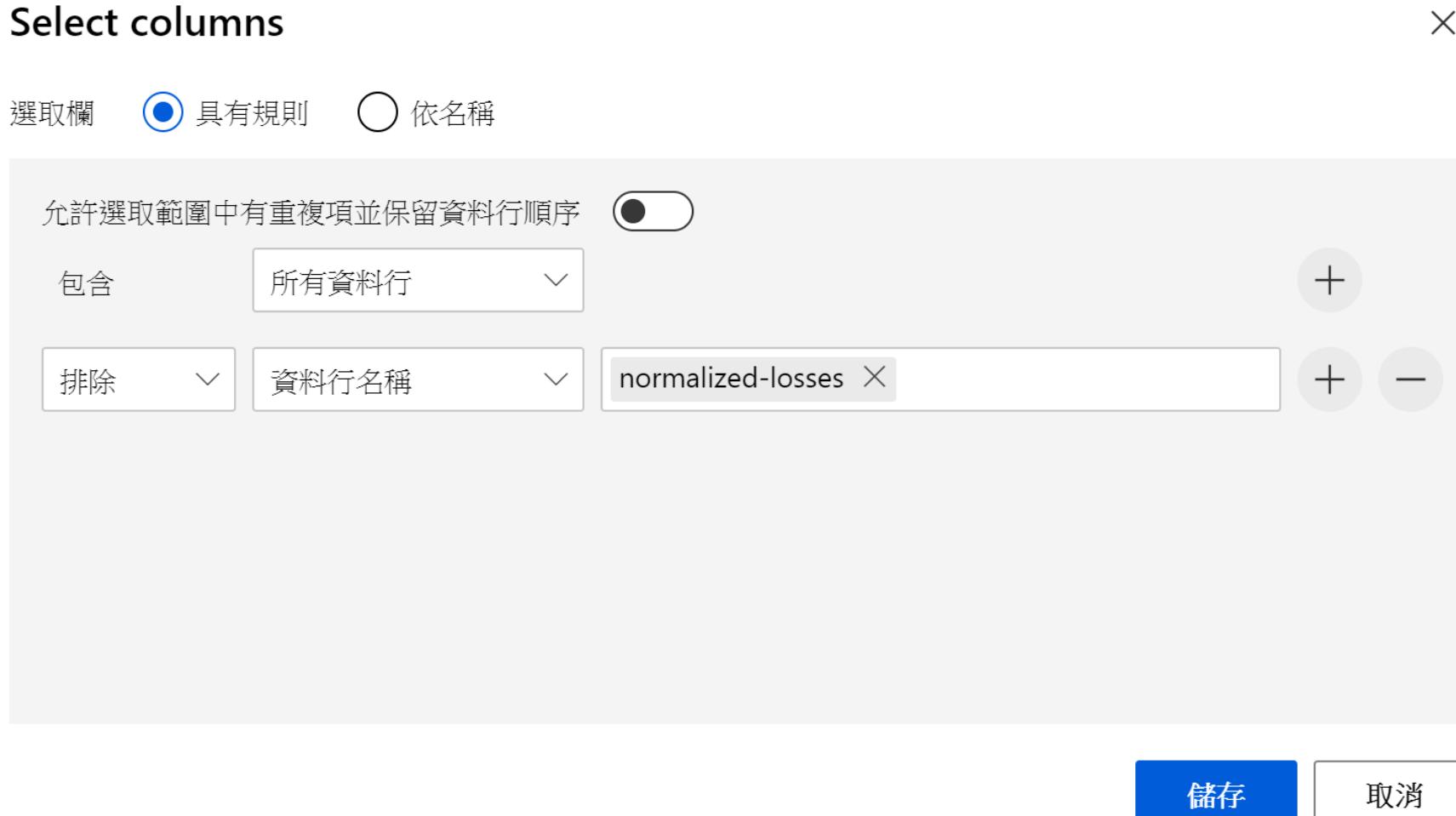
排除normalized-losses字段

- 选取 Select Columns in Dataset
- 右边 -> 编辑数据行



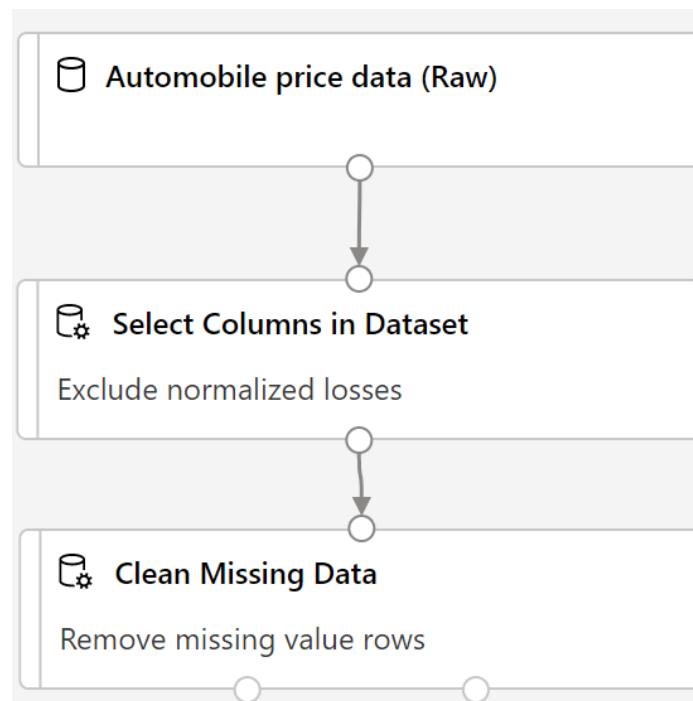
排除normalized-losses字段

- 包含「所有数据行」，「+」
- 「排除」->「数据行名称」-> normalized-losses
- 储存



处理遗漏值数据

- 模块 -> Data Transformation -> Clean Missing Data
- 拖曳至工作区
- 将「Select Columns in Dataset」与「Clean Missing Data」连接
- 选取「Clean Missing Data」，右边点击「编辑数据行」-> 包含「所有数据行」，储存
- 右下方的Cleaning mode，设定为「Remove entire row」
- 右下方批注区块输入Remove missing value rows
- 也可以顺便补一下上一个的批注
 - Exclude normalized losses



Clean Missing Data

Columns to be cleaned ② * 編輯資料行
所有資料行

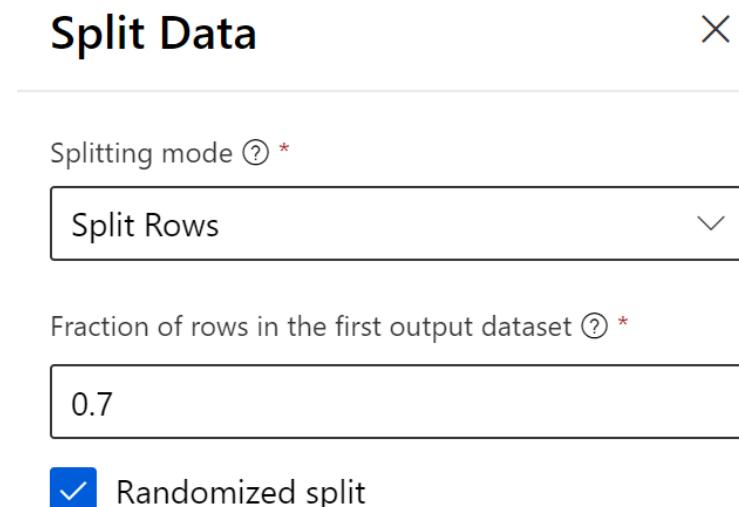
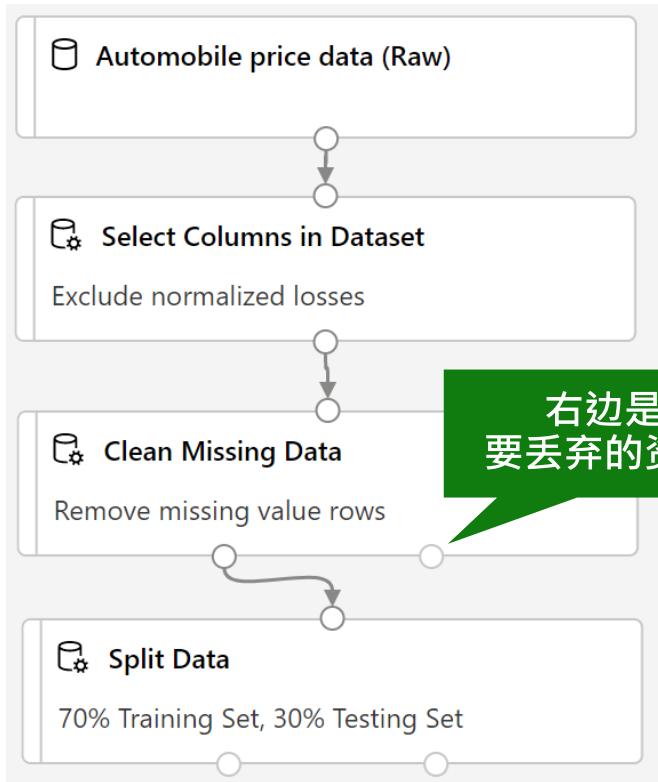
Minimum missing value ratio ② *
0.0

Maximum missing value ratio ② *
1.0

Cleaning mode ② *
Remove entire row

资料切割

- 模块 -> Data Transformation -> Split Data
- 拖曳至工作区
- 将「Clean Missing Data」左边的节点往「Split Data」连接
- 选取「Split Data」，右方比例设定为0.7
- 填写批注「70% Training Set, 30% Testing Set」

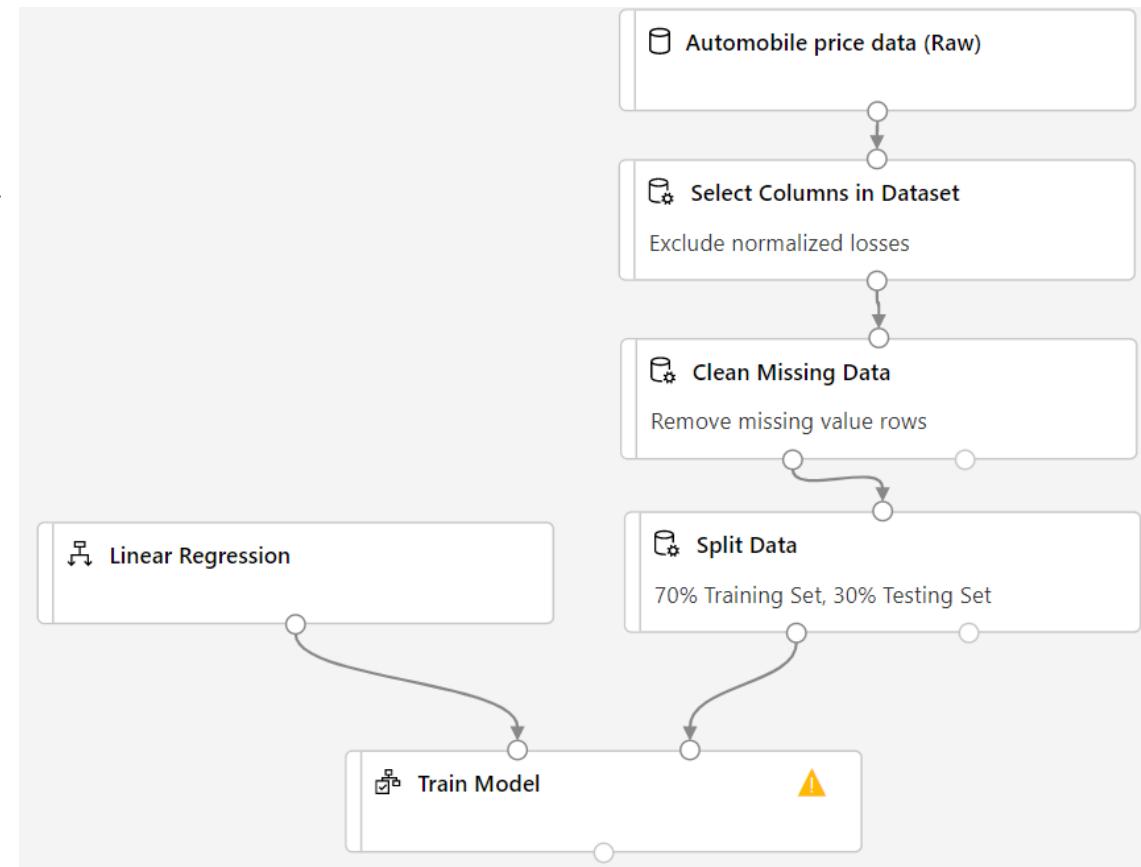


模型训练

- 模块 -> Machine Learning Algorithms -> Regression -> Linear Regression，拖曳至工作区
- 模块 -> Model Training -> Train Model，拖曳至工作区
- 将Linear Regression连接至Train Model的左上角节点
- 将Split Data左下方节点连接至Train Model的右上角节点
- 设定训练目标
 - 选取Train Model，右边编辑数据行
 - 选取单一数据行「数据行名称」「price」

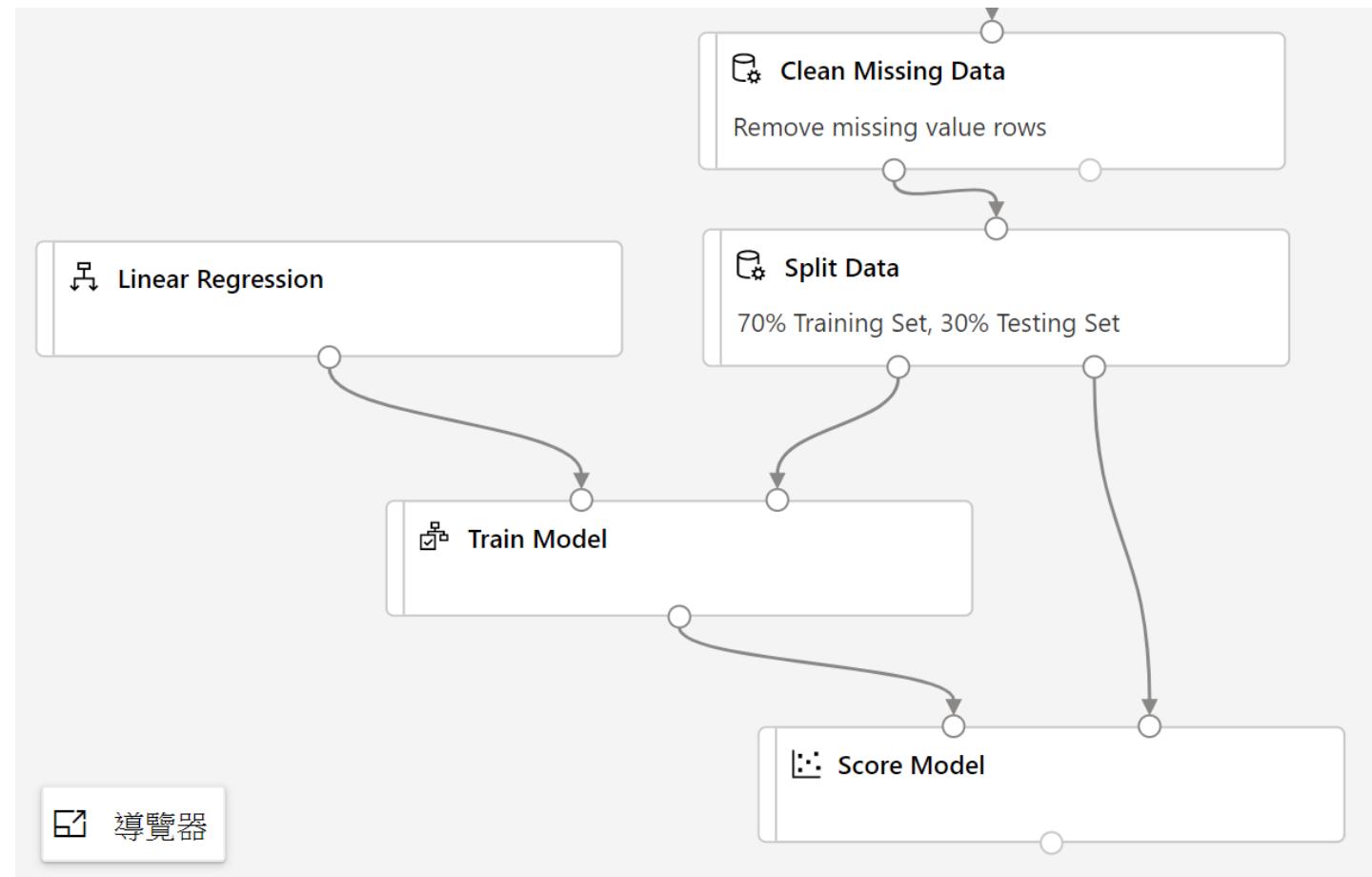
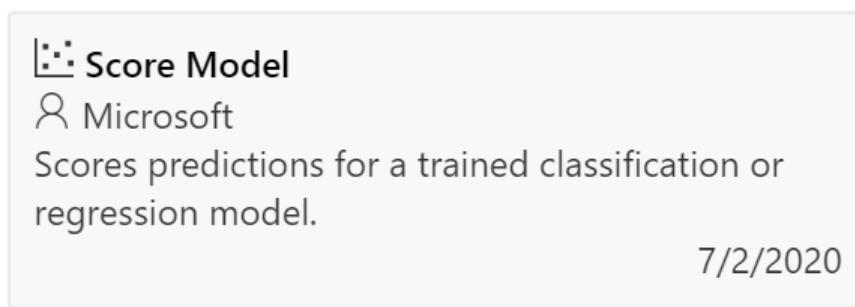
Label column

選取單一資料行 資料行名稱 ▾ price X



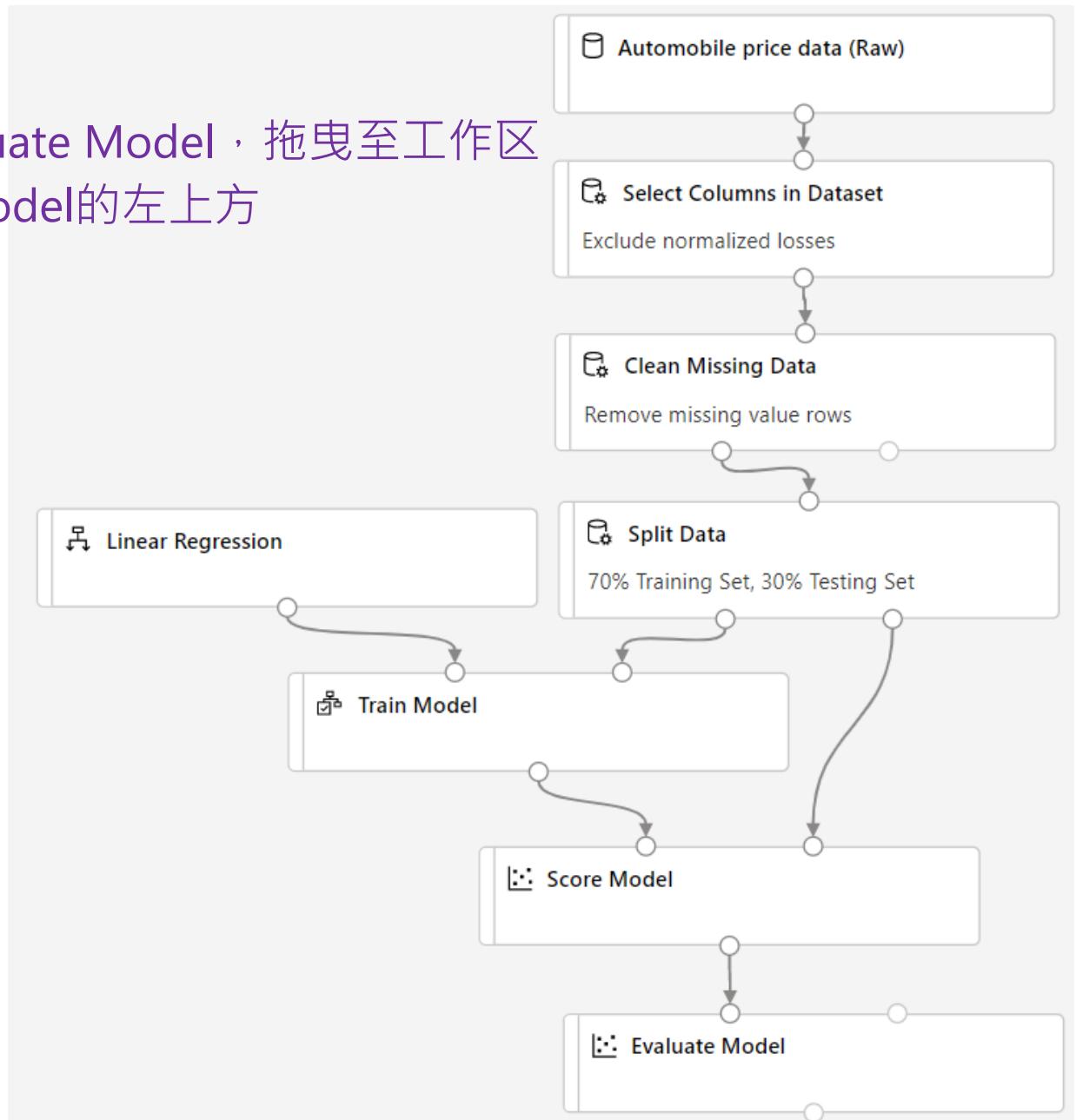
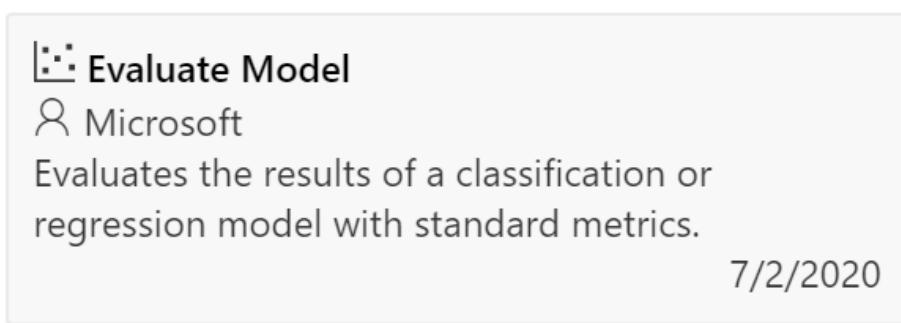
模型预测结果

- 模块 -> Model Scoring & Evaluation -> Score Model，拖曳至工作区
- 将Train Model与Score Model的左上角节点连接
- 将Split Data的右下方节点与Score Model的右上角节点连接



模型评估

- 模块 -> Model Scoring & Evaluation -> Evaluate Model，拖曳至工作区
- 将Score Model下方的节点连接至Evaluate Model的左上方



提交运行!

- 右上角 -> 提交 -> 实验「建立新的」
- 输入实验名称
 - Hello_Car_Price_Predict
- 提交

設定管線回合

×

實驗

選取現有項目 建立新的

新增實驗名稱 *

Hello_Car_Price_Predict



回合描述 *

Car Price Prediction

計算目標

預設值

HelloML

提交

取消

查看结果

- 选取Score Model区块，右方「输出+记录文件」分页，数据输出 -> 点击可视化图示
- 查看最右方的Scored Labels，与price进行比对

Score Model

參數 輸出 + 記錄檔 詳細資料 計量 子回合

資料輸出 隱藏資料輸出 ^

Scored dataset  

其他輸出

price	Scored Labels
9639	10909.222803
6377	5982.362604
18950	18032.474669
12945	6698.976263
11248	8388.865106
9988	8099.639332
7895	9710.382318

评估结果

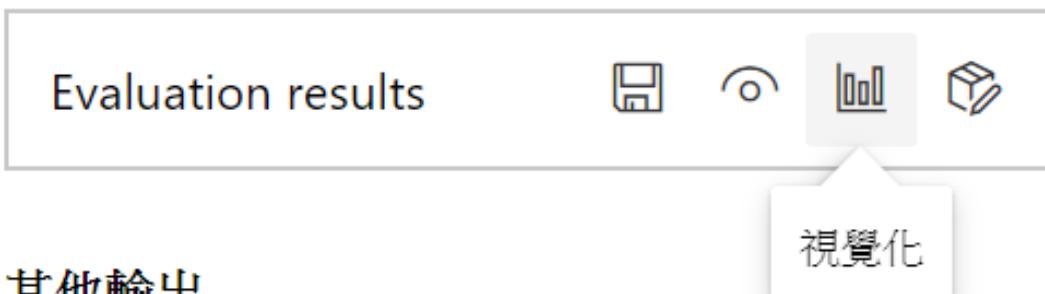
- 选取Evaluate Model区块，右方「输出+记录文件」分页，数据输出 -> 点击可视化图示
 - Coefficient_of_Determination(r2_score) : 0.8682

Evaluate Model

參數 輸出 + 記錄檔 詳細資料 計量 子回合

資料輸出 隱藏資料輸出 ^

Evaluation results



其他輸出

部署机器学习模型 – Machine Learning Designer

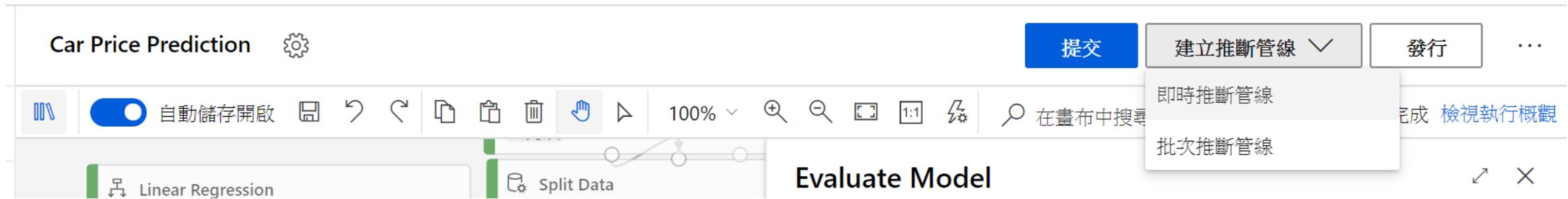
- 可实时产生推论结果



Source : abeyon.com

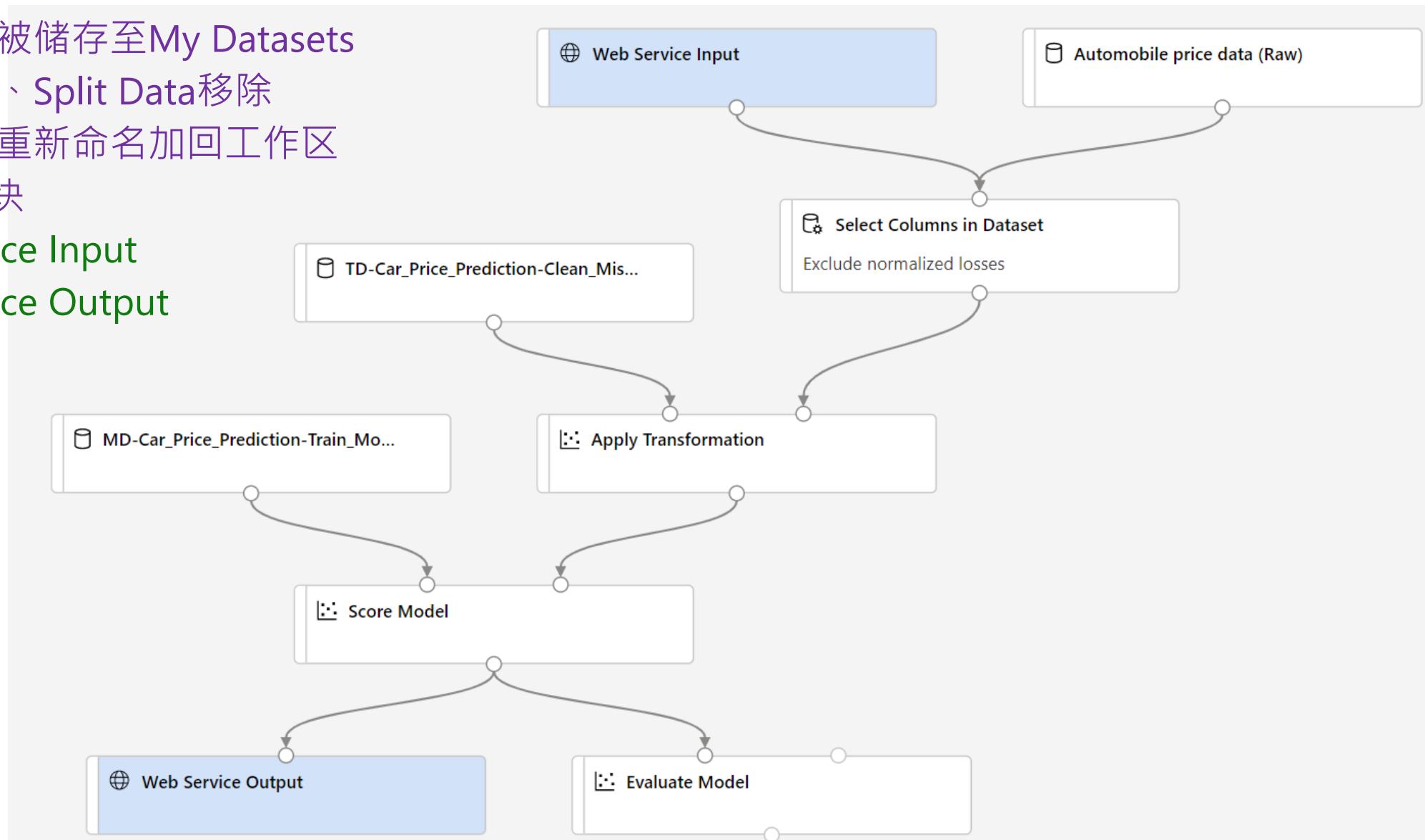
建立实时推论管线

- Real-time inference pipeline



自动完成以下任务

- Train Model被储存至My Datasets
- Train Model、Split Data移除
- Train Model重新命名加回工作区
- 增加两个模块
 - Web Service Input
 - Web Service Output



确认后提交、部署

- 按下提交，选择前一阶段相同的实验
- 完成后，按下部署

The screenshot shows the Azure ML Studio interface with the following details:

- Title Bar:** Car Price Prediction-real time inference
- Top Right Buttons:** 提交 (Submit), 部署 (Deploy), and three dots for more options.
- Toolbar:** Includes icons for back, forward, search, and zoom (90%, 1:1).
- Search Bar:** 在畫布中搜尋 (Search in canvas) with a magnifying glass icon.
- Main Area:** 設定管線回合 (Set Pipeline Iteration).
 - Experiment Selection:** 實驗 (Experiment) section with two radio buttons:
 - 選取現有項目 (Select Existing Project) (selected)
 - 建立新的 (Create New)
 - Existing Experiment:** 現有實驗 * (Existing Experiment *) dropdown menu showing "Hello_Car_Price_Predict".
 - Iteration Description:** 回合描述 * (Iteration Description *) input field containing "Car Price Prediction-real time inference".
- Compute Target:** 計算目標 (Compute Target) section showing "預設值" (Default Value) and "HelloML".
- Bottom Buttons:** 提交 (Submit) and 取消 (Cancel) buttons.

设定实时端点

- 点击计算，新增推论从集、完成后重新整理页面

設定即時端點

您對於部署尚無任何計算目標。請瀏覽至[計算](#)，並建立*推斷叢集*部署的計算目標。
建立之後，[重新整理](#)，此快顯視窗可查看最近新增的計算目標。

新增推斷叢集

計算名稱 * [i](#)

HelloCompute

Kubernetes 服務

[建立新的](#) [使用現有項目](#)

區域 *

東亞

虛擬機器大小 * [i](#)

Standard_D2 2 核心, 7 GB (RAM), 100 GB (磁碟)

叢集用途

生產環境 開發/測試

節點數 * [i](#)

1

網路組態 [i](#)

[基本](#) [進階](#)

啟用 SSL 設定 [i](#)

[下載自動化的範本](#)

[部署](#) [取消](#) [建立](#)

设定实时端点

- 部署新的实时端点、选择刚刚建立好的计算目标，按下部署，完成后点击检视实时端点

設定即時端點 ×

部署新的即時端點 取代現有的即時端點

即時端點名稱 *
car-price-prediction-real-time-i

端點描述 (選用)

計算目標

現有的計算目標 重新整理

計算目標名稱	節點計數	區域	狀態 ↓
HelloCompute	① 1	eastasia	Succeeded

部署 取消

部署: 正在從模型建立即時端點

部署: 正在等候建立即時端點

部署: 成功 檢視即時端點

测试实时端点

- 进入测试分页、可修改各项属性值，按下测试

HelloML > 端點 > car-price-prediction-real-time-i

car-price-prediction-real-time-i

詳細資料 测試 取用

輸入要測試即時端點的資料 **測試** 測試結果 **已剖析** 原始

WebServiceInput0 **WebServiceOutput0**

symboling	key	value
3	symboling	3
1	make	alfa-romero
alfa-romero	fuel-type	gas
gas	aspiration	std
std	num-of-doors	two
	body-style	convertible
	drive-wheels	rwd
	engine-location	front
	wheel-base	88.6
	length	168.8
	Scored Labels	13935.989246512887

预测：年收入是否大于五万美金

- 使用Azure Machine Learning Designer



Source : conflictdivorcedenver.com

Microsoft Azure Machine Learning

- 设计工具 -> 立即启动

Microsoft Azure Machine Learning

HelloML > 首頁

歡迎使用工作室!

建立新的

筆記本

利用 Python SDK 編寫程式碼，並執行範例實驗。

立即啟動

自動化 ML (預覽)

使用目標計量自動訓練及調整模型。

立即啟動

設計工具 (預覽)

將介面從預備資料，拖放到部署模型。

立即啟動

<https://ml.azure.com/>

Designer

- 新增易于使用的预建模块

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽)

設計工具 (預覽)

新增管線

易於使用的預建模組 ②

- HelloML
- 設計工具 (預覽)
- 新增
- 首頁
- 作者
- 筆記本
- Automated ML (preview)
- 設計工具 (預覽) (selected)
- 資產
- 資料集
- 實驗

修改项目名称

- Adult_Income_Prediction

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽) > 正在撰寫

Search by name, tags and description

Adult_Income_Prediction

提交

新增 首頁

作者 筆記本 自動化 ML (預覽)

設計工具 (預覽) 資產

資料集 實驗 管線 模型 端點 管理 計算

Datasets Modules Models

自動儲存開啟

Modules

- ▶ Data Input and Output (3)
- ▶ Data Transformation (19)
- ▶ Feature Selection (2)
- ▶ Statistical Functions (1)
- ▶ Machine Learning Algorithms (18)
- ▶ Model Training (4)
- ▶ Model Scoring & Evaluation (6)
- ▶ Python Language (2)
- ▶ R Language (1)

設定

預設計算目標 ②

選取計算目標以執行管線。

未選取任何計算目標
選取計算目標

管線參數

尚未選取任何參數

預設輸出設定 ②

選取預設資料存放區

运算资源设定

- 点击专案名称旁的「齿轮」
- 右下方点击「选取计算目标」

Microsoft Azure Machine Learning

HelloML > 設計工具 (預覽) > 正在撰寫

Car Price Prediction 

提交 發行 ...

自動儲存開啟

100% 1:1

設定

預設計算目標 ⑦

⚠ 選取計算目標以執行管線。

未選取任何計算目標 

管線參數 +

尚未選取任何參數

新增 首頁 作者 筆記本 自動化 ML (預覽) 設計工具 (預覽) 資產 資料集 實驗 管線 模型 端點

設定計算目標

- 建立新的并输入计算目标名称
- 或选择已经存在的

設定計算目標

選取現有項目 建立新的

計算目標組態

建議您使用下方預先定義的組態，快速設定訓練計算目標。

組態	vCPU	RAM	儲存體	節點計數
Predefined	2	7GB	8GB	2

不符合您的需求嗎？請前往 [計算] -> [訓練叢集]，以建立自訂計算目標。

新增計算目標名稱 [?](#)



HelloML

儲存

取消

設定計算目標

選取現有項目 建立新的

現有的計算目標

[重新整理](#)

計算目標名稱	可用節點數/節點... HelloML HelloAutoML	區域 westcentr... westcentr...	狀態 Succeeded Succeeded
HelloML	0 / 2	westcentr...	Succeeded
HelloAutoML	0 / 1	westcentr...	Succeeded

儲存

取消

新增数据集

- Datasets -> Samples
- Adult Census Income
- 拖曳至工作区

The screenshot shows the Azure Machine Learning studio interface. On the left, there's a sidebar with a search bar and three tabs: 'Datasets' (selected), 'Modules', and 'Models'. Below the tabs, there are two sections: 'Datasets' (containing 2 datasets) and 'Samples' (containing 13 samples). A sample named 'Adult Census Income Binary Classification dataset' is highlighted with a red box. To the right, the main workspace is titled 'Adult_Income_Prediction'. It shows a list of datasets and modules. One dataset, 'Adult Census Income Binary Classification dataset', is selected and shown in a preview pane. On the far right, detailed information about this dataset is displayed, including its name ('Adult Census Income Binary Classification'), type ('GlobalDataset'), description ('Census Income dataset'), data type ('DataFrameDirectory'), storage location ('azureml_globaldatasets'), and relative path ('GenericCSV/Adult_Census_Income_Binary_').

Search by name and description

Adult_Income_Prediction

Datasets Modules Models

Datasets

Datasets (2)

Samples (13)

Adult Census Income Binary Classification dataset

Automobile price data (Raw)

CRM Appetency Labels Shared

CRM Churn Labels Shared

Adult Census Income Binary Classification dataset

Adult Census Income Binary Classification dataset

參數 輸出

資料集名稱
Adult Census Income Binary Classification

資料來源類型
GlobalDataset

描述
Census Income dataset

資料類型
DataFrameDirectory

資料存放區名稱
azureml_globaldatasets

相對路徑
GenericCSV/Adult_Census_Income_Binary_

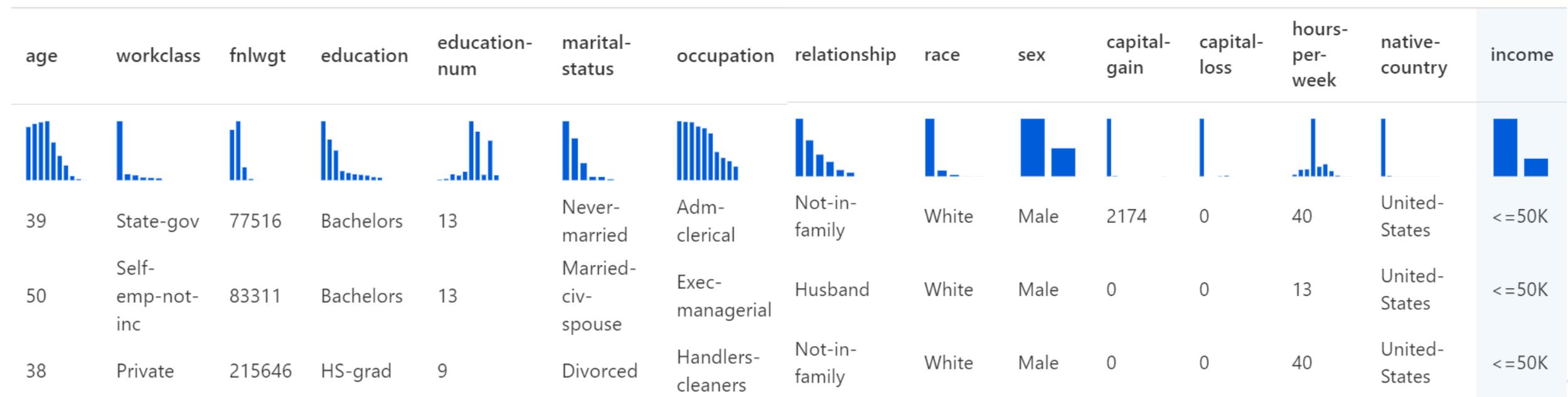
观察数据集

- 点击该数据集区块
- 右方选单分页->输出
- Dataset output -> 可视化

Adult Census Income Binary Classification dataset 結果視覺效果

資料列 ② 資料行 ②

32,561 15



检查遗漏值状况

- 这三个有遗漏值

workclass

統計資料

平均數

-

中位數

-

最小值

-

最大值

-

標準差

-

唯一值

8

遺漏值

1836

功能類型

String Feature

occupation

統計資料

平均數

-

中位數

-

最小值

-

最大值

-

標準差

-

唯一值

14

遺漏值

1843

功能類型

String Feature

native-country

統計資料

平均數

-

中位數

-

最小值

-

最大值

-

標準差

-

唯一值

41

遺漏值

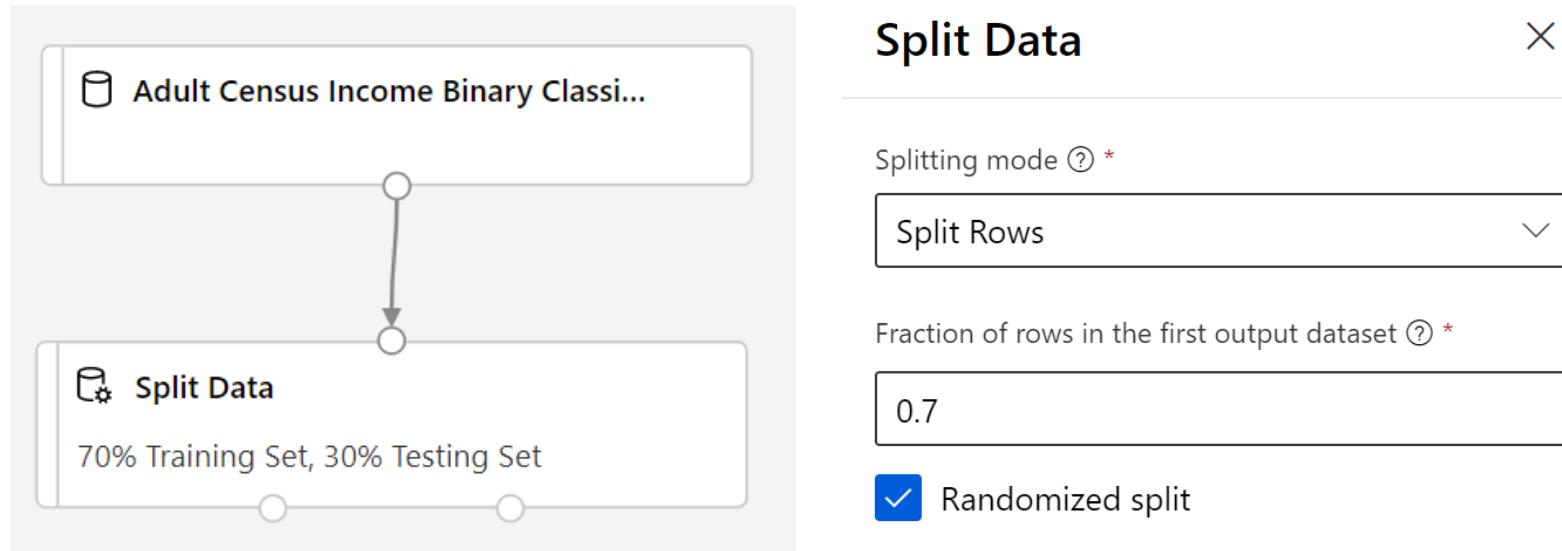
583

功能類型

String Feature

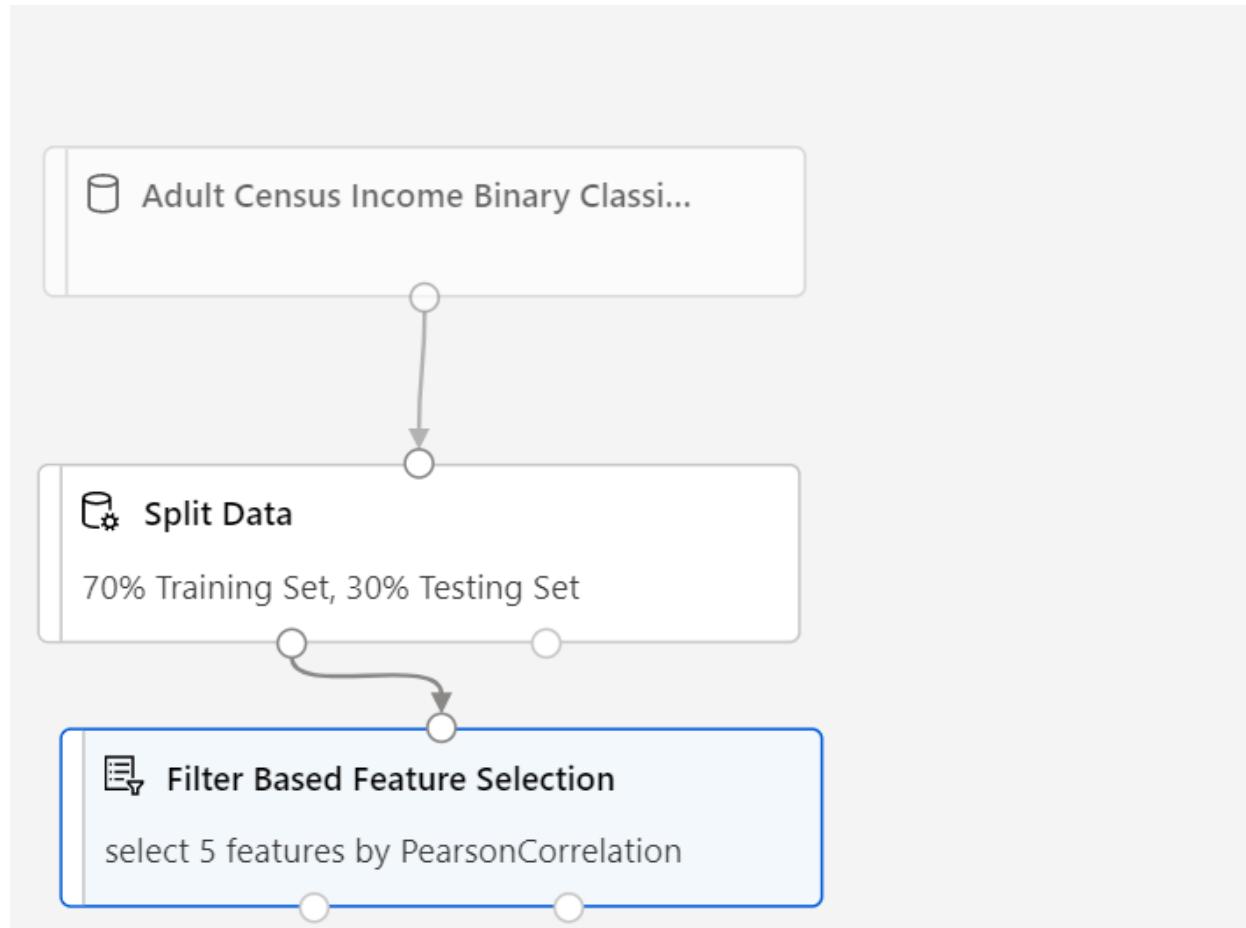
资料切割

- 模块 -> Data Transformation -> Split Data
- 拖曳至工作区
- 将「Clean Missing Data」左边的节点往「Split Data」连接
- 选取「Split Data」，右方比例设定为0.7
- 填写批注「70% Training Set, 30% Testing Set」



特征选取

- 目标为income
- 使用PearsonCorrelation选择5个特征



Filter Based Feature Selection

Operate on feature columns only

Target column ⓘ *

資料行名稱: income

Number of desired features ⓘ *

5

Feature scoring method ⓘ *

PearsonCorrelation

模型训练

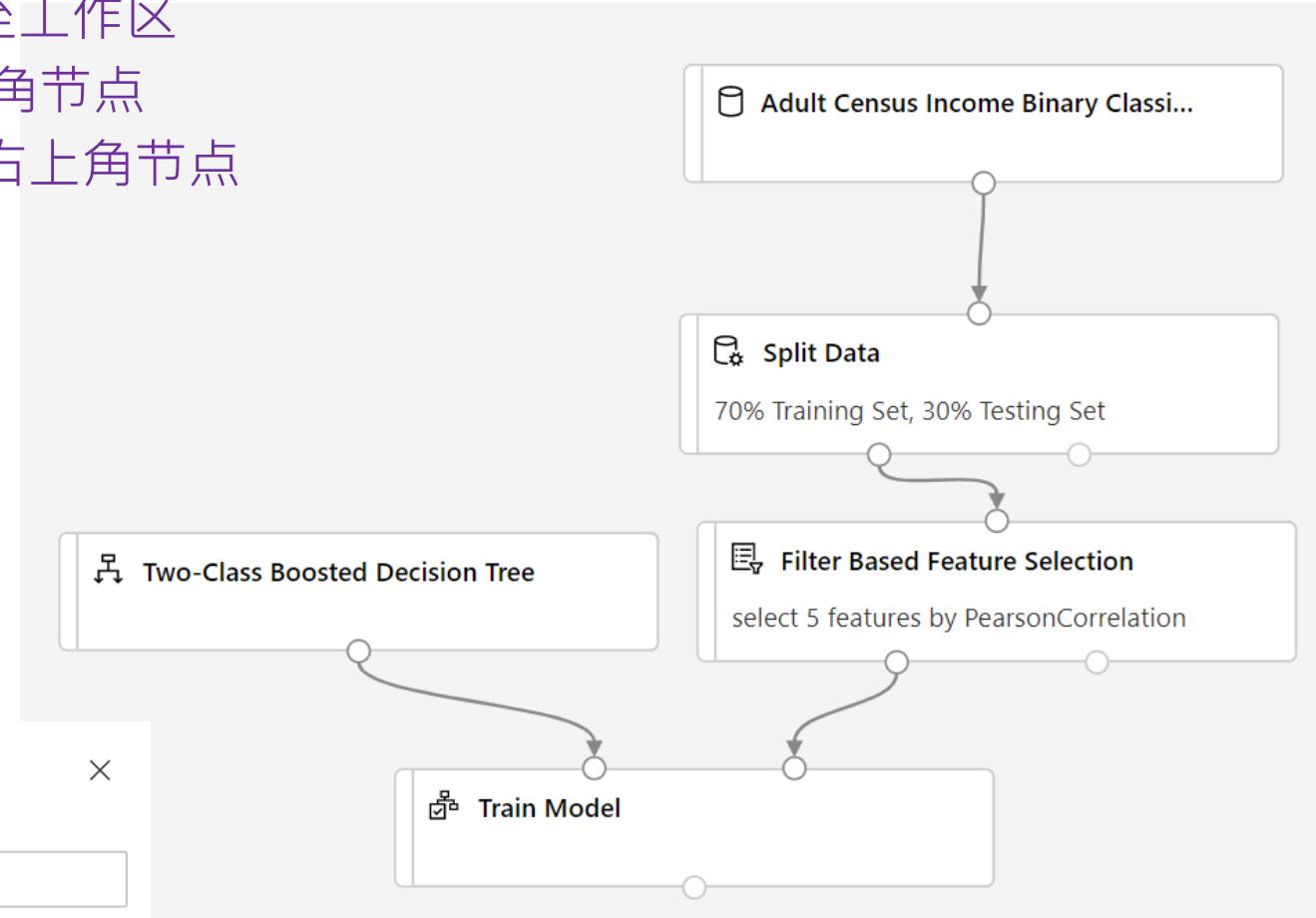
- 模块 -> Machine Learning Algorithms -> Classification -> Two-Class Boosted Decision Tree，拖曳
- 模块 -> Model Training -> Train Model，拖曳至工作区
- 将Two-Class Boosted连接至Train Model的左上角节点
- 将Filter Based左下方节点连接至Train Model的右上角节点
- 设定训练目标
 - 选取Train Model，右边编辑数据行
 - 选取单一数据行「数据行名称」「income」

Label column

选取单一資料行

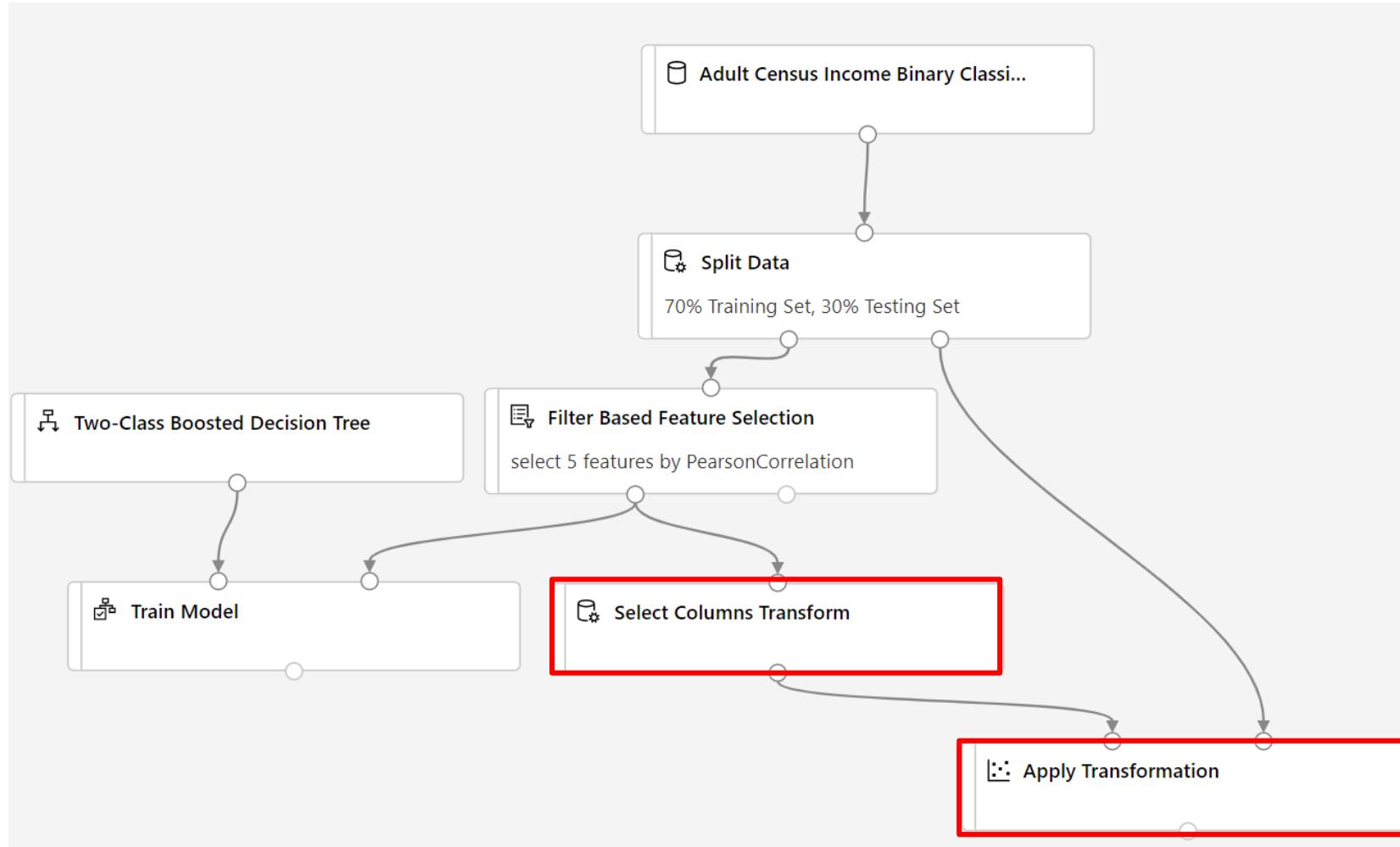
Column names

income X



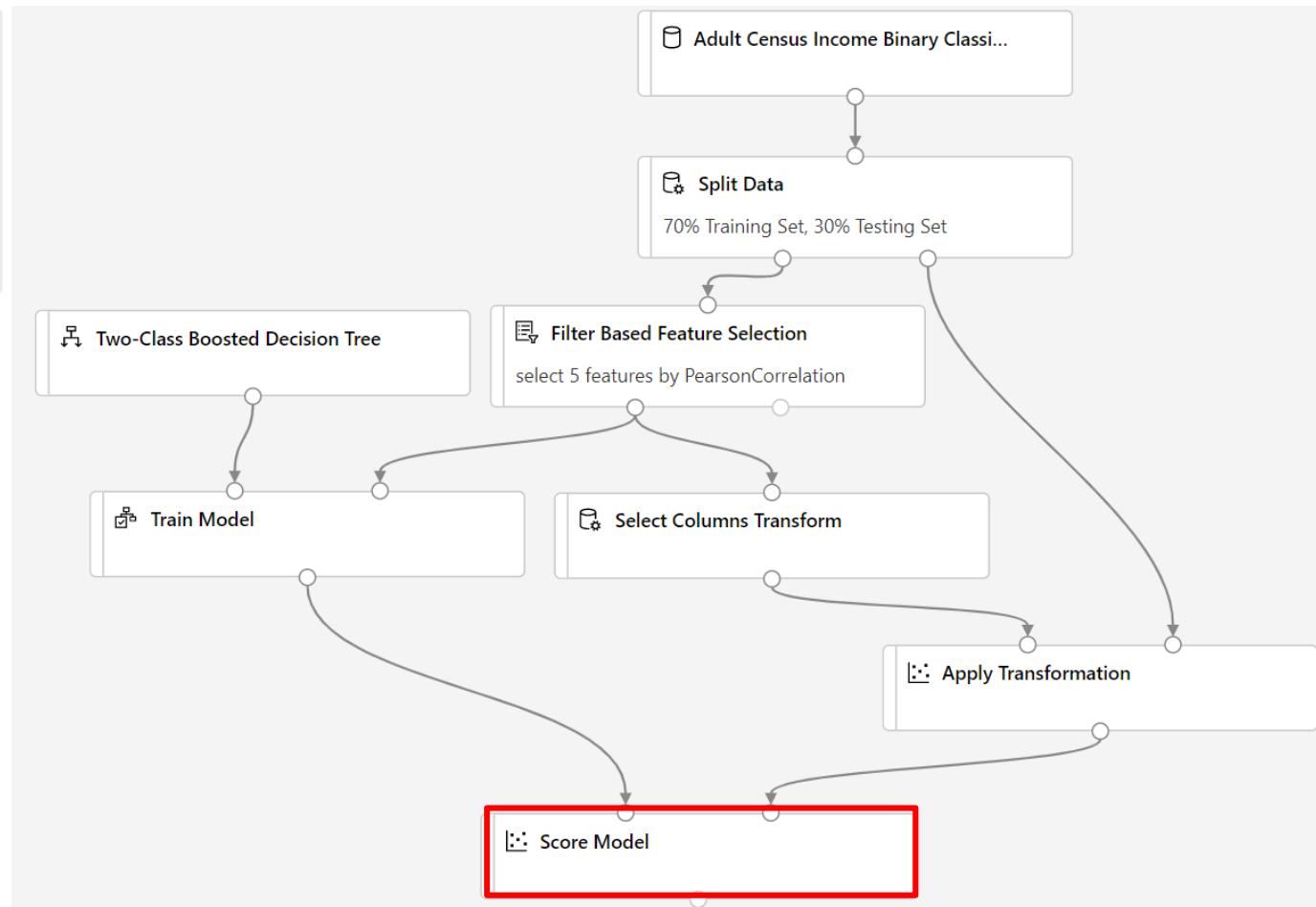
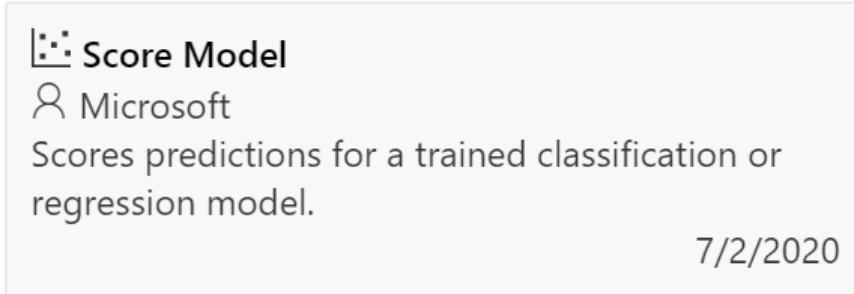
Data Transformation

- 加入Select Columns Transform与Apply Transformation两个模块
- 把Filter Based Feature Selection所选择的Feature也套用在Testing Dataset上



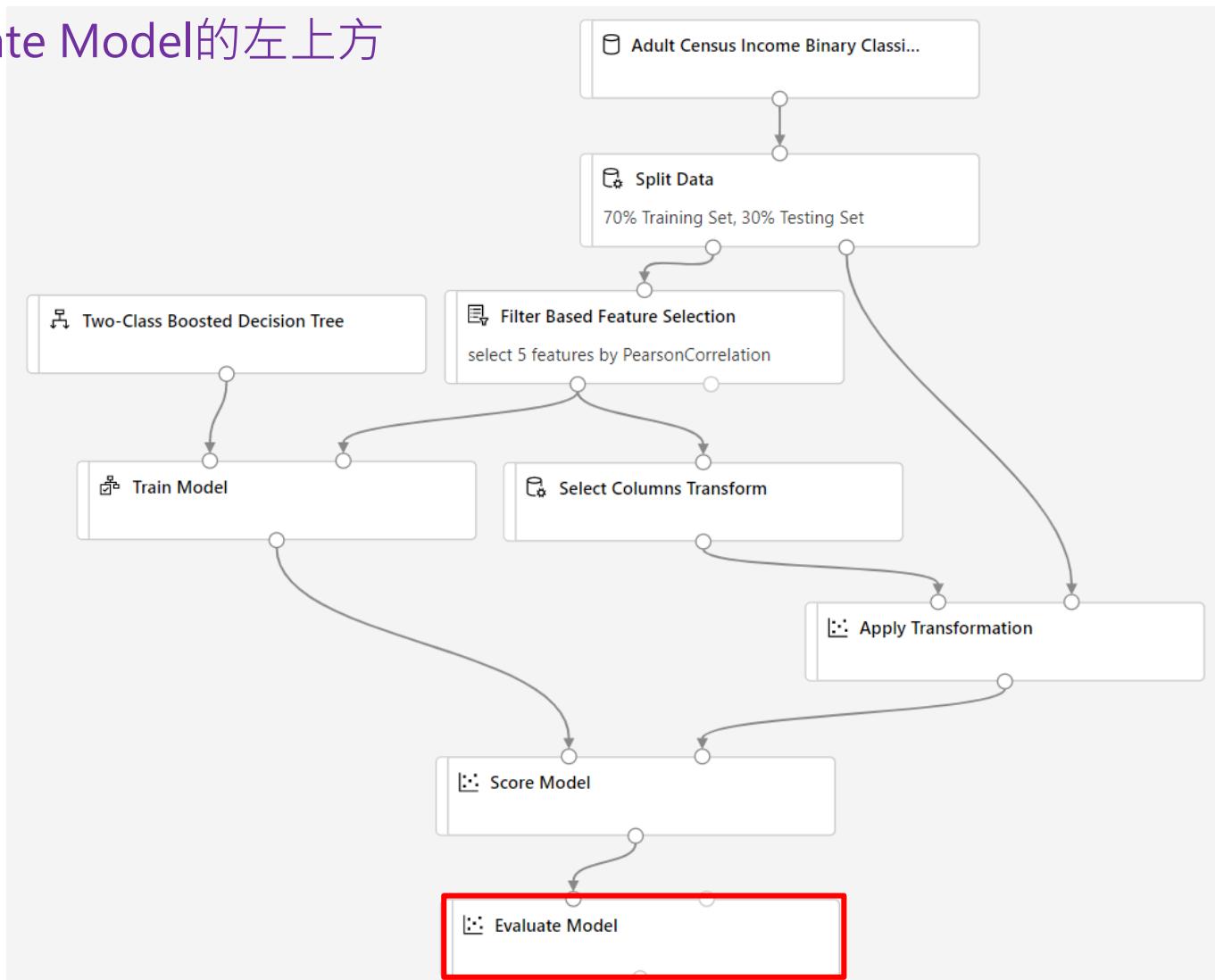
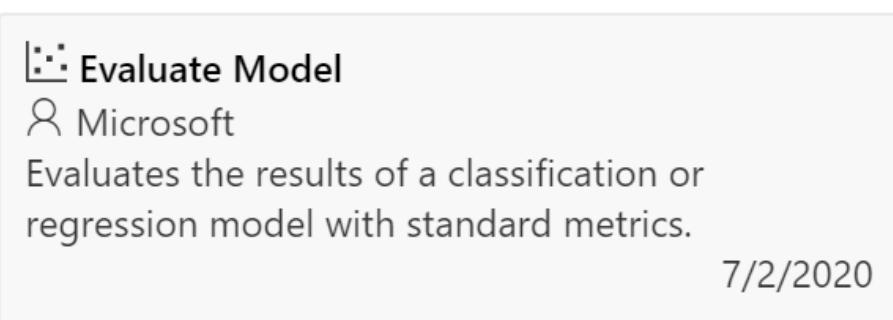
模型预测结果

- 模块 -> Model Scoring & Evaluation -> Score Model，拖曳至工作区
- 将Train Model与Score Model的左上角节点连接
- 将Apply Transformation与Score Model的右上角节点连接



模型评估

- 模块 -> Model Scoring & Evaluation -> Evaluate Model，拖曳至工作区
- 将Score Model下方的节点连接至Evaluate Model的左上方



提交运行!

- 右上角 -> 提交 -> 实验 -> 建立新的
- 输入新的实验名称
- 提交

設定管線回合

X

實驗



選取現有項目



建立新的

新增實驗名稱 *



Adule_Income_Prediction

回合描述 *

Adult_Income_Prediction

計算目標

預設值

HelloML

提交

取消

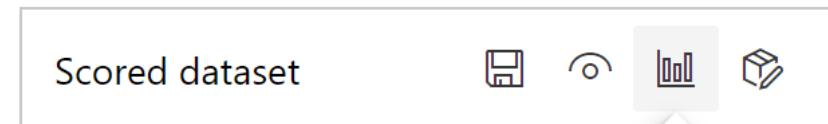
查看结果

- 选取Score Model区块，右方「输出+记录文件」分页，数据输出 -> 点击可视化图示
- 查看最右方的Scored Labels，与income进行比对

Score Model

參數 輸出 + 記錄檔 詳細資料 計量 子回合

資料輸出 隱藏資料輸出 ^



其他輸出

Score Model 結果視覺效果

資料列 ② 資料行 ②

9,768 8

age	education-num	capital-gain	capital-loss	hours-per-week	income	Scored Labels	Scored Probabilities
36	11	0	0	40	<=50K	<=50K	0.213274
34	9	0	0	40	<=50K	<=50K	0.112404
52	15	0	0	50	>50K	>50K	0.826978
27	15	0	0	40	<=50K	<=50K	0.163372
62	14	0	0	70	<=50K	>50K	0.528947
38	13	0	0	40	>50K	<=50K	0.383485
31	10	0	0	40	<=50K	<=50K	0.143985
65	14	0	0	aa	>50K	>50K	0.216022

评估结果

- 选取Evaluate Model区块，右方「输出+记录文件」分页，数据输出 -> 点击可视化图示
 - 观察各项数值表现

Evaluate Model



Azure Machine Learning – 自動化ML(UI)

首頁 > 新增 >

機器學習

Microsoft



機器學習

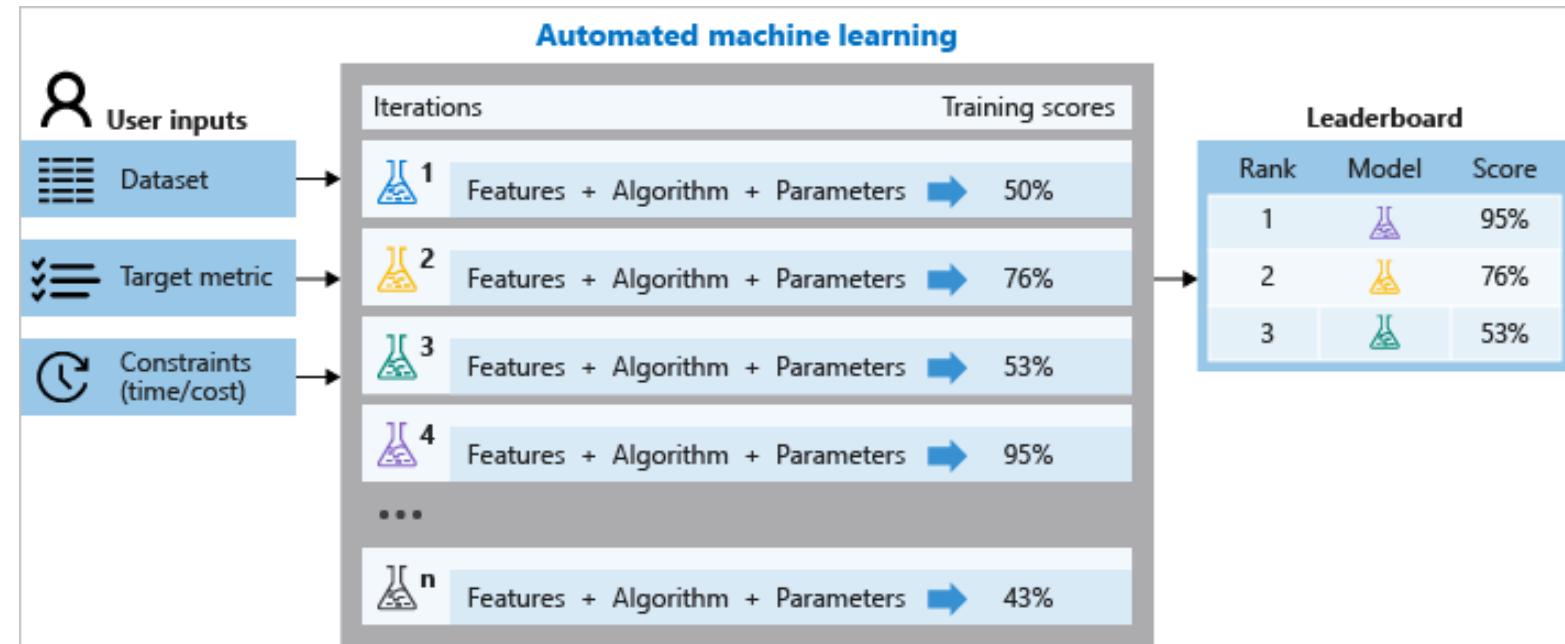
已儲存

Microsoft

建立

概觀

方案



Azure Machine Learning 可為開發人員及資料科學家提供各種具有生產力的機器學習模型建置、訓練與部署體驗。建立 Azure Machine Learning 工作區即可訓練、管理及部署機器學習實驗及 Web 服務。

實用的連結

文件

定價詳細資料

Azure AI 資源庫

Azure Machine Learning – 自动化ML(UI)

- 省略标签步骤，进行检阅+建立

首頁 > 新增 > 機器學習 >

機器學習

建立機器學習工作區

基本 標籤 檢閱 + 建立

專案詳細資料

選取用以管理部署資源及成本的訂用帳戶。使用像資料夾這樣的資源群組來安排及管理您的所有資源。

工作區詳細資料

指定工作區的名稱、區域和版本。

工作區名稱 * ⓘ

HelloAutoML 

區域 * ⓘ

東亞 

工作區版本 * ⓘ

企業 

判斷客戶是否會申請定期存款



Automated ML

- 新增自动化ML回合

The screenshot shows the Microsoft Azure Machine Learning studio interface. The left sidebar contains navigation links such as '新增' (Add), '首頁' (Home), '作者' (Author), '筆記本' (Notebook), 'Automated ML (preview)' (which is highlighted with a red box), '設計工具 (預覽)' (Design Tools (Preview)), '資料集' (Datasets), '實驗' (Experiments), '管線' (Pipelines), '模型' (Models), and '端點' (Endpoints). The main content area is titled 'Automated ML (preview)' and displays the message: '無需撰寫任何程式碼，即可讓自動化 ML 依據您的資料訓練及尋找最佳模型。深入了解自動化 ML' (No code required to let automated ML train your data and find the best model. Learn more about automated ML). A red box highlights the '+ 新增自動化 ML 回合' (Create new automated ML iteration) button. Below it, a message says '沒有任何最近的自動化 ML 回合可顯示。' (No recent automated ML iterations are displayed.) and '按一下 [新增自動化 ML 回合] 建立您的第一個回合' (Click [Create new automated ML iteration] to build your first iteration). A link '深入了解如何建立自動化 ML 回合' (Learn more about how to build an automated ML iteration) is also present. At the bottom, there's a '文件' (File) section with a link '概念: 什麼是自動化 ML?' (Concept: What is automated ML?).

<https://ml.azure.com/>

选取数据集

- 建立数据集 -> 来自本机档案

建立新的自動化 ML 回合

選取資料集

設定執行

工作類型與設定

選取資料集

從下面的清單中選取資料集，或建立新的資料集。自動化 ML 目前只支援用表格式資料撰寫執行。

+ 建立資料集 | 僅顯示支援的資料集

來自本機檔案

來自資料存放區

來自 Web 檔案

來自開放資料集

資料集類型

建立日期

已修改

搜尋以篩選項目...



沒有任何可顯示的 資料集

上一步

下一個

取消

填写数据集名称 BankMarketing

從本機檔案建立資料集

基本資訊

名稱 * 資料集版本

資料集類型 *

描述

上一步 下一個 取消

基本資訊
資料存放區和檔案選取
設定與預覽
結構描述
確認詳細資料

默认数据存放区 -> 设定上传路径 -> 下一步

從本機檔案建立資料集

資料存放區和檔案選取

選取或建立資料存放區 *

目前選取的資料存放區: workspaceblobstore (Azure Blob 儲存體) (預設)

先前建立的資料存放區

建立新的資料存放區

為您的資料集選取檔案 *

建立資料集之後，這些檔案將會上傳至您的預設 Blob 儲存體，並可在工作區中使用。支援的檔案類型包含: 分隔檔(也就是 csv、tsv)、Parquet、JSON Lines 與純文字。

瀏覽 已選取 1 個檔案。大小總計為 3.958 MiB。已上傳 0/1 個檔案

檔案名稱	大小 (MiB)	上傳 %	狀態
bankmarketing_train.csv	3.958	<div style="width: 100%;">100%</div>	完成

上傳路徑

dataset 檔案將會上傳至 '\$(上傳路徑)07-08-2020_035156_UTC'

跳過資料驗證 ⓘ

上一步 **下一步** **取消**

设定与预览

- 确认侦测是否正确：
 - 文件格式-分隔符
 - 分隔符-逗号
 - 编码-UTF-8
 - 所有档案都有相同标头
 - 略过资料列-无

設定與預覽

這些設定是自動偵測而來。請驗證選取項目正確，否則請予以更新

檔案格式

分隔符號

分隔符號 範例

逗號 Field1,Field2,Field3

編碼

UTF-8

資料行標頭

所有檔案都有相同的標頭

略過資料列

無

識..	age	job	marital	education	defa
1	57	technician	married	high.school	no
2	55	unknown	married	unknown	unknow
3	33	blue-collar	married	basic.9y	no
4	36	admin.	married	high.school	no

上一步

下一個

取消

选择要实验的字段

- 将 day_of_week 关闭

Create dataset from local files

Basic info

Settings and preview

Schema

Confirm details

Schema

Include	Column name	Properties	Type
<input checked="" type="checkbox"/>	Path	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	age	Not applicable to selecte...	Integer
<input checked="" type="checkbox"/>	job	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	marital	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	education	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	default	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	housing	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	loan	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	contact	Not applicable to selecte...	String
<input checked="" type="checkbox"/>	month	Not applicable to selecte...	String

Back Next Cancel

确认建立

從本機檔案建立資料集

確認詳細資料

資料集類型 表格式
檔案設定
檔案格式 分隔符號
分隔符號 逗號
編碼 UTF-8
資料行標頭 所有檔案都有相同的標頭
略過資料列 無

在建立後分析此資料集

[上一步](#) [建立](#)

选取数据集完成

- 勾选BankMarketing -> 下一个

HelloML > Automated ML (preview) > 啟動回合

✓ 成功: 已成功建立 BankMarketing 資料集

建立新的自動化 ML 回合

選取資料集

從下面的清單中選取資料集，或建立新的資料集。自動化 ML 目前只支援用表格式資料撰寫執行。

+ 建立資料集 | 僅顯示支援的資料集 | 搜尋以篩選項目...

資料集名稱

BankMarketing

選取資料集

設定執行

工作類型與設定

上一步

下一步

設定執行

- 输入实验名称(HelloAutoML)、目标数据行(y)、选取计算丛集(建立新的计算)

建立新的自動化 ML 回合

設定執行

設定實驗。從現有實驗中選取名稱或定義新名稱，然後選取要使用的目標資料行及訓練計算。[深入了解如何設定實驗](#)

資料集
BankMarketing ([檢視資料集](#))

實驗名稱 *

選取現有的 新建

新增實驗名稱

目標資料行 * [i](#)

選取計算叢集 * [i](#)

[建立新的計算](#) [重新整理計算](#)

[上一步](#) [下一個](#)



建立新的計算

- 計算名稱
- 虛擬機器類型
- 虛擬機器大小
- 节点数目下限
- 节点数目上限

新增計算叢集 [\(i\)](#)

計算名稱 * [\(i\)](#)

區域 * [\(i\)](#)

虛擬機器類型 *

虛擬機器優先順序 * [\(i\)](#)

專用 低優先順序

虛擬機器大小 * [\(i\)](#)

Standard_DS12_v2 4 核心, 28 GB (RAM), 56 GB (磁碟)

節點數目下限 * [\(i\)](#)

節點數目上限 * [\(i\)](#)

相應減少之前的閒置秒數 * [\(i\)](#)

[下載自動化的範本](#) [建立](#) [取消](#)

設定執行

- 確認計算叢集後 -> 下一個

建立新的自動化 ML 回合

設定執行

設定實驗。從現有實驗中選取名稱或定義新名稱，然後選取要使用的目標資料行及訓練計算。深入了解如何設定實驗 [\[\]](#)

資料集
BankMarketing ([檢視資料集](#))

實驗名稱 * [\[\]](#)

選取現有的 新建

新增實驗名稱 [\[\]](#)
HelloAutoML

目標資料行 * [\[\]](#)
y

選取計算叢集 * [\[\]](#)
HelloAutoML

[\[\]](#) 建立新的計算 [\[\]](#) 重新整理計算

[上一步](#) [下一個](#) [取消](#)

选取工作类型

- 分类

建立新的自動化 ML 回合

- 選取資料集
- 設定執行
- 工作類型與設定

選取工作類型

為實驗選取機器學習服務工作類型。如有需要，還可使用其他設定來微調實驗。



分類

預測目標資料行中數種類別的其中一種。是/否、藍、紅、綠。



迴歸

預測連續數值



時間序列預測

根據時間預測值

[檢視其他組態設定](#)

[檢視特徵化設定](#)

上一步

完成

取消

选取工作类型

- 分类

建立新的自動化 ML 回合

選取資料集

設定執行

工作類型與設定

選取工作類型

為實驗選取機器學習服務工作類型。如有需要，還可



分類

預測目標資料行中數種類別的其中一種。是



啟用深度學習



迴歸

預測連續數值



時間序列預測

根據時間預測值

檢視其他組態設定

檢視特徵化設定

上一步

完成

其他組態

主要計量

精確度

解釋最佳模型

封鎖的演算法

自動化 ML 在訓練期間不會使用的演算法清單。

結束準則

訓練作業時間 (小時)

1

計量分數閾值

計量分數閾值

驗證

驗證類型

自動

並行

並行反覆次數上限

1

取消

完成 -> 等待自动化ML进行

- 自动依算法精确度进行排序

持續時間

45 分 9.645 秒

最佳模型摘要

演算法名稱

[VotingEnsemble](#)

精確度

0.91927 [檢視所有其他計量](#)

取樣

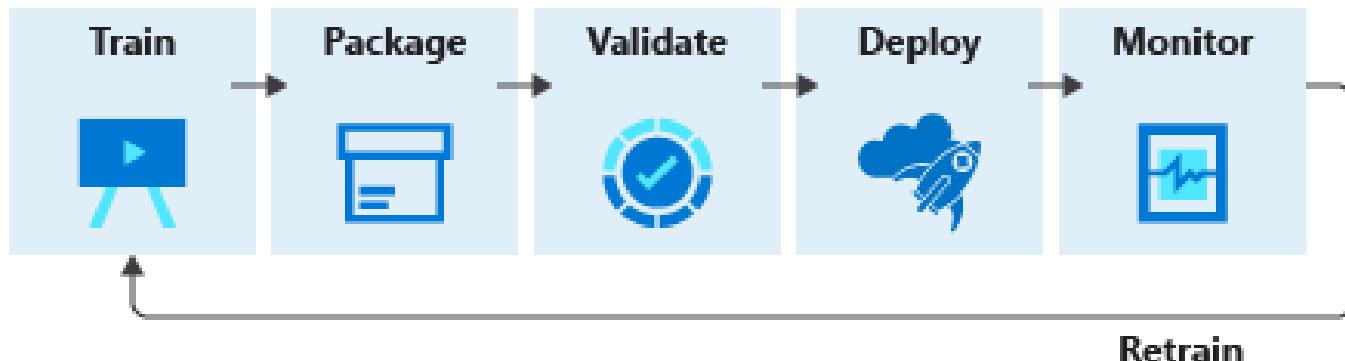
100% [\(i\)](#)

演算法名稱	解釋	精確度 ↓
VotingEnsemble	檢視說明	0.91927
SparseNormalizer, XGBoostClassifier		0.91411
StackEnsemble		0.91290
SparseNormalizer, XGBoostClassifier		0.91259
MaxAbsScaler, LightGBM		0.91199

Azure Machine Learning 微软 Azure 机器学习

- 云端环境
- 可以进行模型的训练/部署/自动化/管理/追踪
- 适用于
 - 传统机器学习 / 深度学习 / 监督式学习 / 非监督式学习
- 使用弹性
 - 可自行撰写Python/R 或 使用Azure ML 图形化界面

Azure Machine Learning Model Workflow



Azure Machine Learning Designer

- 微软 Azure 机器学习设计师
- 最新推出，以拖拉方式建立机器学习流程

Jupyter Notebook

- 可以在云端上使用，上面有许多范例也可以自己从头建立

VS Code Extension 插件

- 可结合本地端运行

CLI Extension

- 以指令列方式使用

Reinforcement learning

- 实验中，使用Ray RLLib



Reactor



developer.microsoft.com/reactor/
@MSFTReactor on Twitter

议程结束
感谢聆听



请记得填写课程回馈问卷
<https://aka.ms/ReactorFeedback>

© 2019 Microsoft Corporation. All rights reserved. The text in this document is available under the Creative Commons Attribution 3.0 License, additional terms may apply. All other content contained in this document (including, without limitation, trademarks, logos, images, etc.) are not included within the Creative Commons license grant. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it. Some examples are for illustration only and are fictitious. No real association is intended or inferred. Microsoft makes no warranties, express or implied, with respect to the information provided here.