

# AzureMachineLearning

## ハンズオン入門

株式会社ISAO MSPプロジェクト

秋山 康平

# はじめる前に

- 以下の準備はできていますか？
  - PC
  - ネット環境
  - ブラウザ
  - Microsoftアカウント

# 目次

- 自己紹介
- Microsoft Azure 概要
- Azure Machine Learning 概要
- ハンズオン

# 自己紹介

- 2017年1月からISAO入社
- ブログ  
<http://kheiakiyama.hateblo.jp/>
- Azure 歴それなり
- 機械学習歴<10人日
- ゼロから始めるDeepLearning4章で挫折

# 機械学習の学習(僕の場合)

- 「機械学習」でググる
- 数学の知識が必要と知る
- 微分積分を学び直し始める
- python を学び始める

压倒的 yak shaving 感

# プログラマの学習

- HelloWorld
- マルバツゲーム、シューティングゲーム
- 徐々に大きめのアプリへ
- 第2言語を学ぶ

# HelloWorld としての AzureMachineLearning



# Microsoft Azure 概要

今日の内容向けの最低限

# Microsoft Azure 概要

- ポータル

<https://portal.azure.com>

- 認証は Microsoftアカウント
- 請求単位は Azureサブスクリプション

# Azure Machine Learning

## 概要

# Azure Machine Learning 概要

- 公式

<https://azure.microsoft.com/ja-jp/services/machine-learning/>

- 概要

<https://docs.microsoft.com/ja-jp/azure/machine-learning/machine-learning-studio-overview-diagram>

Machine Learning in ML Studio

Anomaly Detection

- One-class Support Vector Machine
- Principal Component Analysis-based Anomaly Detection
- Time Series Anomaly Detection\*

Classification

Two-class Classification

- Averaged Perceptron
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- Support Vector Machine

Multi-class Classification

- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-all

Clustering

- K-means Clustering

Recommendation

- Matchbox Recommender

Regression

- Bayesian Linear Regression
- Boosted Decision Tree
- Decision Forest
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

Statistical Functions

- Descriptive Statistics
- Hypothesis Testing T-Test
- Linear Correlation
- Probability Function Evaluation

Text Analytics

- Feature Hashing
- Named Entity Recognition
- Vowpal Wabbit

Computer Vision

- OpenCV Library

Data/Model Visualization

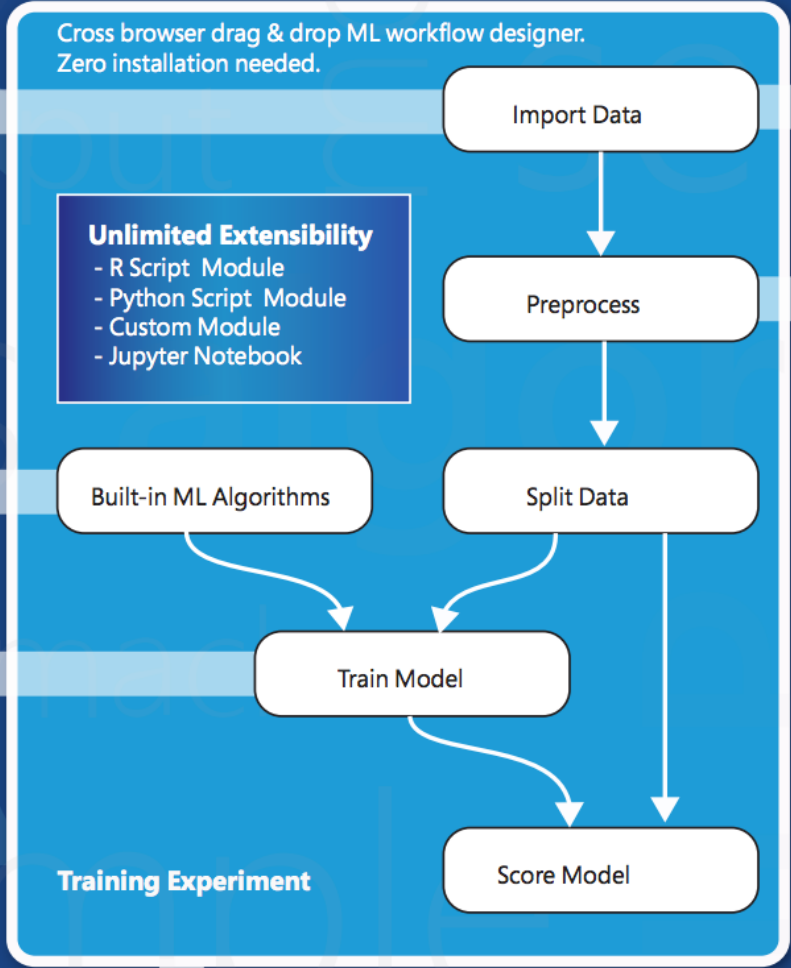
- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree\*

Training

- Cross Validation
- Retraining
- Parameter Sweep

<https://studio.azureml.net>

Guest Access Workspace: Free trial access without logging in.  
Free Workspace: Free persisted access, no Azure subscription needed.  
Standard Workspace: Full access with SLA under an Azure subscription.



Data Source

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW\*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server\*
- Web URL (HTTP)

Data Format

- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Data Preparation

- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQLite Transformation
- Synthetic Minority Oversampling Technique

Enterprise Grade Cloud Service

- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance\*

Community

- Gallery (<http://gallery.azureml.net>)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support

\* Feature Coming Soon

One-click Operationalization

Predictive Experiment

Make Prediction with Elastic APIs

- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API



Azure Machine Learning Studio Capabilities Overview



# Azure Machine Learning 概要

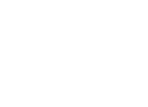
- 課金体系は2つ
  - Microsoftアカウントで使える無料プラン
  - Azureサブスクリプションで使う有料プラン
- Pricing
  - <https://azure.microsoft.com/ja-jp/pricing/details/machine-learning/>

ここからハンズオン



# ハンズオン

- 今回のゴール
  - 日本の祝日かどうかを判定するAIの作成
    - たとえば 2017/1/1 は祝日、  
2017/1/2 は祝日ではない
  - 50年分のデータを元に分析器を作成
  - WebAPIとして呼び出す



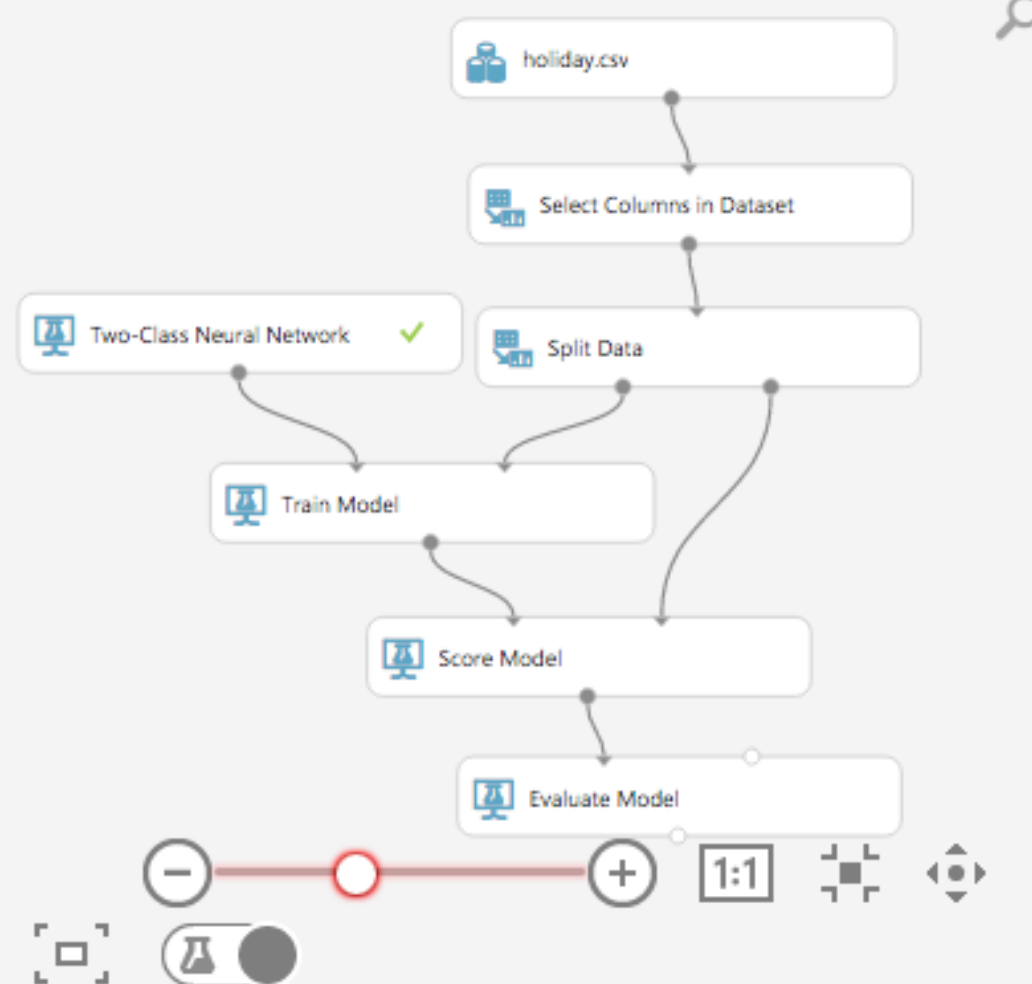
Search experiment items

- ▶ Saved Datasets
- ▶ 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

## Experiment created ...

In draft

Draft saved at 0:18:06



Properties Project

### Experiment Properties

START TIME	3/13/2017 ...
END TIME	3/13/2017 ...
STATUS CODE	InDraft
STATUS DETAILS	None

### Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN


SET UP WEB SERVICE


PUBLISH TO GALLERY


# ハンズオン


- Microsoft Azure Machine Learning Studio  
<https://studio.azureml.net/>  
を開く
- 右上の Sign In からログイン





 PROJECTS


 EXPERIMENTS

 WEB SERVICES

 NOTEBOOKS

 DATASETS

 TRAINED MODELS

 SETTINGS

## experiments

MY EXPERIMENTS   SAMPLES

		NAME	AUTHOR	STATUS	LAST E... ↓	PROJ...	
No experiments found							

0 items selected



# ハンズオン

- <https://goo.gl/Mdzn3p>  
をダウンロード
- DATASETS を開き、NEW から csv ファイルを  
アップロード



PROJECTS



EXPERIMENTS



WEB SERVICES



NOTEBOOKS



DATASETS



TRAINED MODELS



SETTINGS

## datasets

MY DATASETS

SAMPLES

	NAME	SUBMITTED BY	DESCRIPTION	DATA TYPE	CREATED	SIZE	PROJECT	
<input type="checkbox"/>	holiday.csv	kheiakiyama		GenericCSV	3/12/2017 10:58:...	310.99 KB	None	

+ NEW



DOWNLOAD



DELETE



OPEN IN



GENERATE DATA



ADD TO PROJECT

# ハンズオン

- EXPERIMENTS を開き、NEW から Blank Experiment をクリック  
適当な名前をつける
- Saved Datasets -> My Datasets から holiday.csv を配置
- 📖 配置した item の移動や拡大・縮小や視点の移動に慣れておきましょう



Search experiment items



Saved Datasets

My Datasets

holiday.csv



Samples

Data Format Conversions

Data Input and Output

Data Transformation

Feature Selection

Machine Learning

OpenCV Library Modules

Python Language Modules

R Language Modules

Experiment created on 2017/3/12

In draft



holiday.csv

Mini Map



holiday.csv



NEW



RUN HISTORY



SAVE



SAVE AS



DISCARD CHANGES



RUN



SET UP WEB  
SERVICE



PUBLISH TO  
GALLERY



# ハンズオン

- 左上の 検索ボックスから Select Columns in Dataset を検索し配置
- Holiday.csv の下から Select Columns に接続
- Select Columns を選択し、Properties -> Launch column selector で title 以外を選択

📖 試しに Holiday.csv の下をクリック、Visualize を開いてみてください

Select Columns の下を見た場合と比べてみてください



select



## Saved Datasets

### Samples

Restaurant customer data

Restaurant feature data

## Data Transformation

### Manipulation

Select Columns in Datas...

Select Columns Transfor...

Experiment created ...

In draft



holiday.csv

Select Columns in Dataset

1

Mini Map

holiday.csv

Select Columns in Dataset

1



Properties

Project



## Select Columns in Dataset

Select columns

**Selected columns:**

**Column names:**

year,month,day,holiday

Launch column selector

Quick Help



+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

# ハンズオン

- 左上の 検索ボックスから Split Data を検索し配置
  - Select Columns の下から Split Data に接続
  - Split Data を選択し、Properties -> Fraction of rows ... で 0.7 を入力
- 📖 今回はデータの70%を学習に、30%を検証に使う

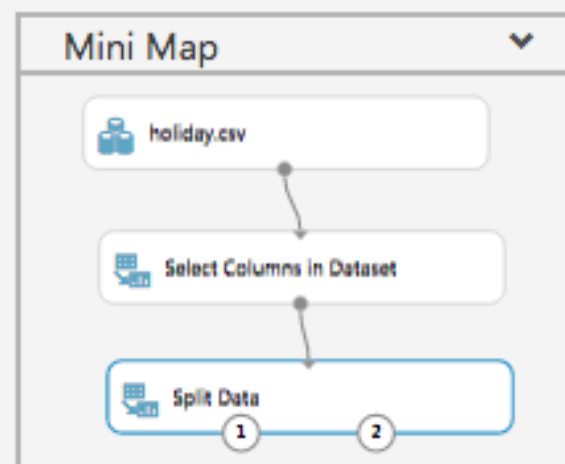
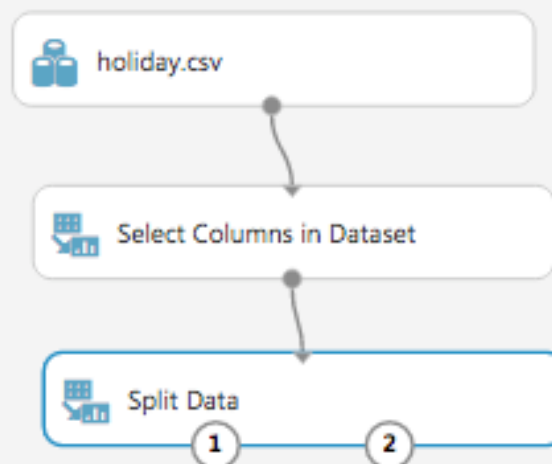


< split 🔍

- Saved Datasets
  - Samples
    - Restaurant ratings
- Data Transformation
  - Sample and Split
    - Split Data

Experiment created ...

In draft



Properties Project

Split Data

Splitting mode

Split Rows

Fraction of rows in the first...

0.7

☒ Randomized split

Random seed

0

Stratified split

False

Quick Help

Split the rows of a dataset into two distinct sets

[\(more help...\)](#)



# ハンズオン

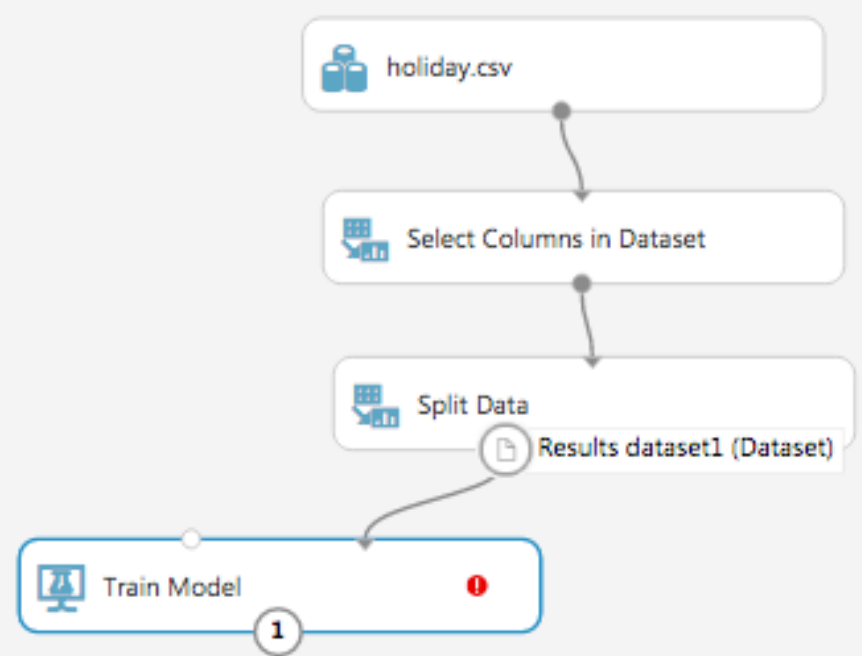
- 左上の 検索ボックスから Train Modelを検索し配置
- Split Data の左下から Train Model の右上に接続
- Train Model を選択し、Properties -> Launch column selector で holiday を選択

📖 今回は教師あり学習



- train
- Sweep Clustering
- Train Anomaly Detection...
- Train Clustering Model
- Train Matchbox Recom...
- Train Model
- Tune Model Hyperpara...
- OpenCV Library Modules
- Pre-trained Cascade Image...
- Text Analytics
- Train Vowpal Wabbit Versi...
- Train Vowpal Wabbit Versi...
- Train Vowpal Wabbit Versi...

Experiment created ... In draft



Properties Project

Train Model

Label column

Selected columns:  
Column names: holiday

Launch column selector

Quick Help



# ハンズオン

- 左上の 検索ボックスから Score Model を検索し配置
- Train Model の下から Score Model 左上に接続
- Split Data の右下から Score Model 右上に接続

 スコアを計測する



score mode

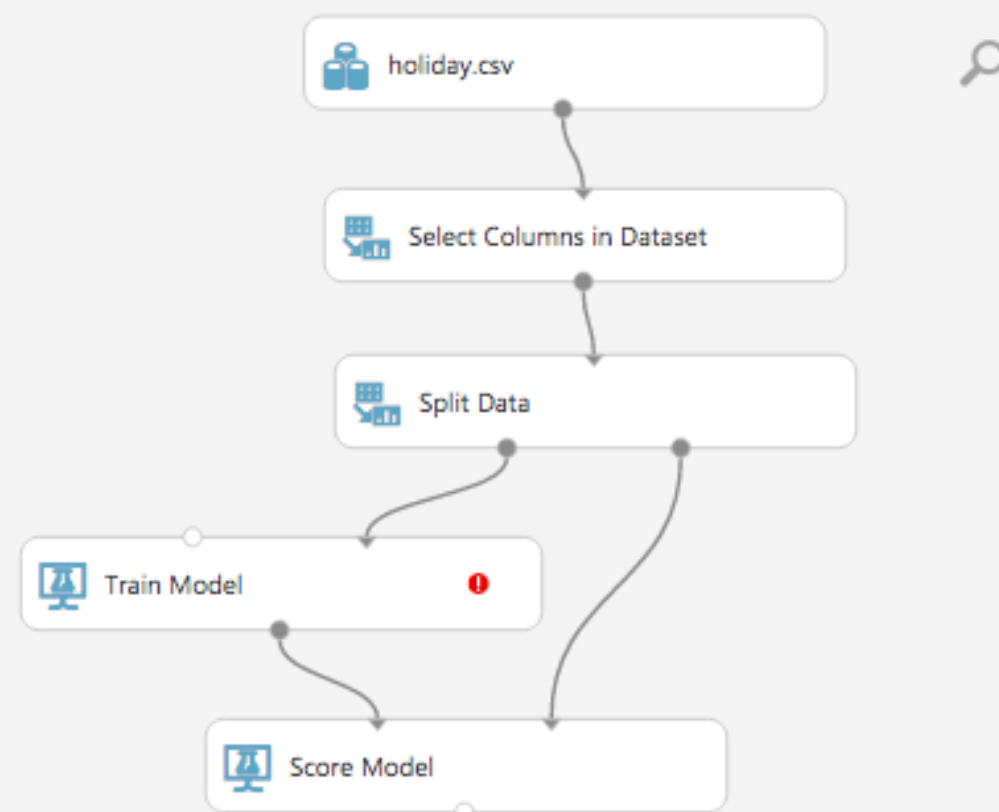
## Machine Learning

### Score

Score Model

Experiment created ...

In draft



Properties Project

### Experiment Properties

STATUS CODE InDraft

### Summary

Enter a few sentences describing your experiment (up to 140 characters).

### Description

Enter the detailed description for your

### Quick Help

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY




# ハンズオン

- 左上の 検索ボックスから Evaluate Model を  
検索し配置
- Score Model の下から Evaluate Model 左上に  
接続

 結果を評価する



< eval 

Machine Learning

Evaluate

Cross Validate Model

Evaluate Model

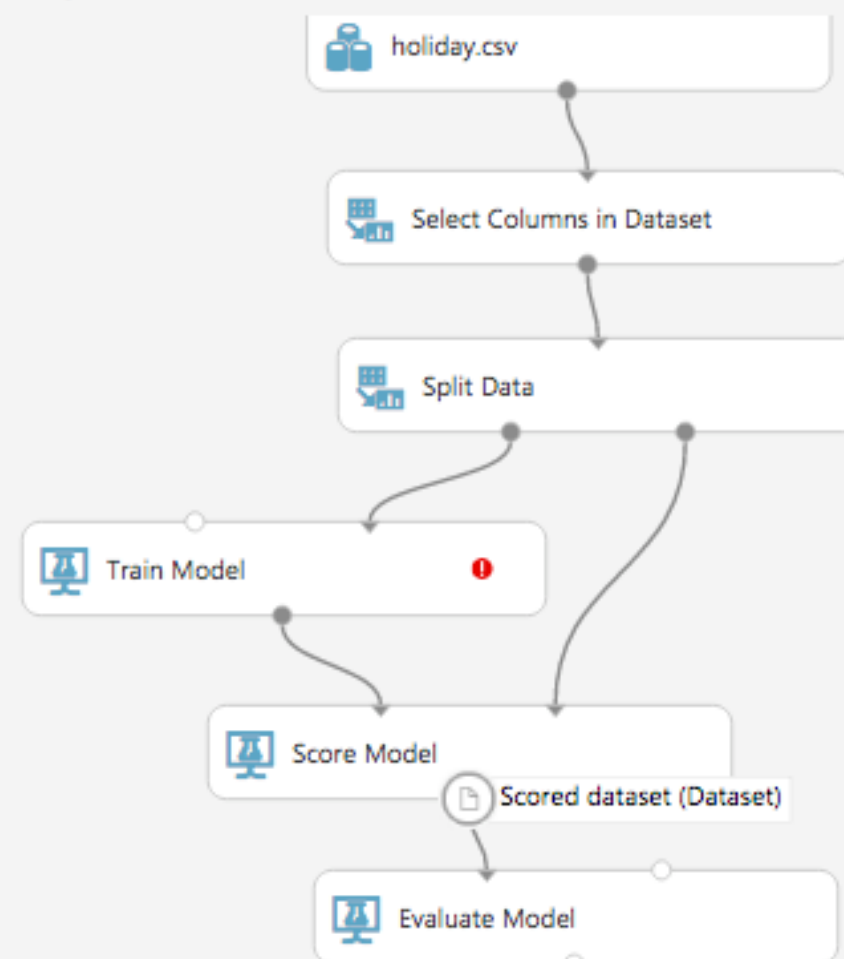
Evaluate Recommender

Statistical Functions

Evaluate Probability Function

# Experiment created ...

In draft



## Properties Project

### Experiment Properties

STATUS CODE InDraft

### Summary

Enter a few sentences describing your experiment (up to 140 characters).

### Description

Enter the detailed description for your

### Quick Help

+ NEW

 RUN HISTORY

 SAVE

 SAVE AS

 DISCARD CHANGES

 RUN

 SET UP WEB SERVICE

 PUBLISH TO GALLERY

# ハンズオン

- いよいよアルゴリズムの選択

チートシートを確認する

今回は2項分類なので、Two-Class ... がよさげです

- (なんらかのアルゴリズム)を検索し、配置
  - 配置したものから Train Model 左上に接続
- 📖 分類、値予測、異常の検出、など何を行いたいかによって使うアルゴリズムは異なります



two-class

Machine Learning

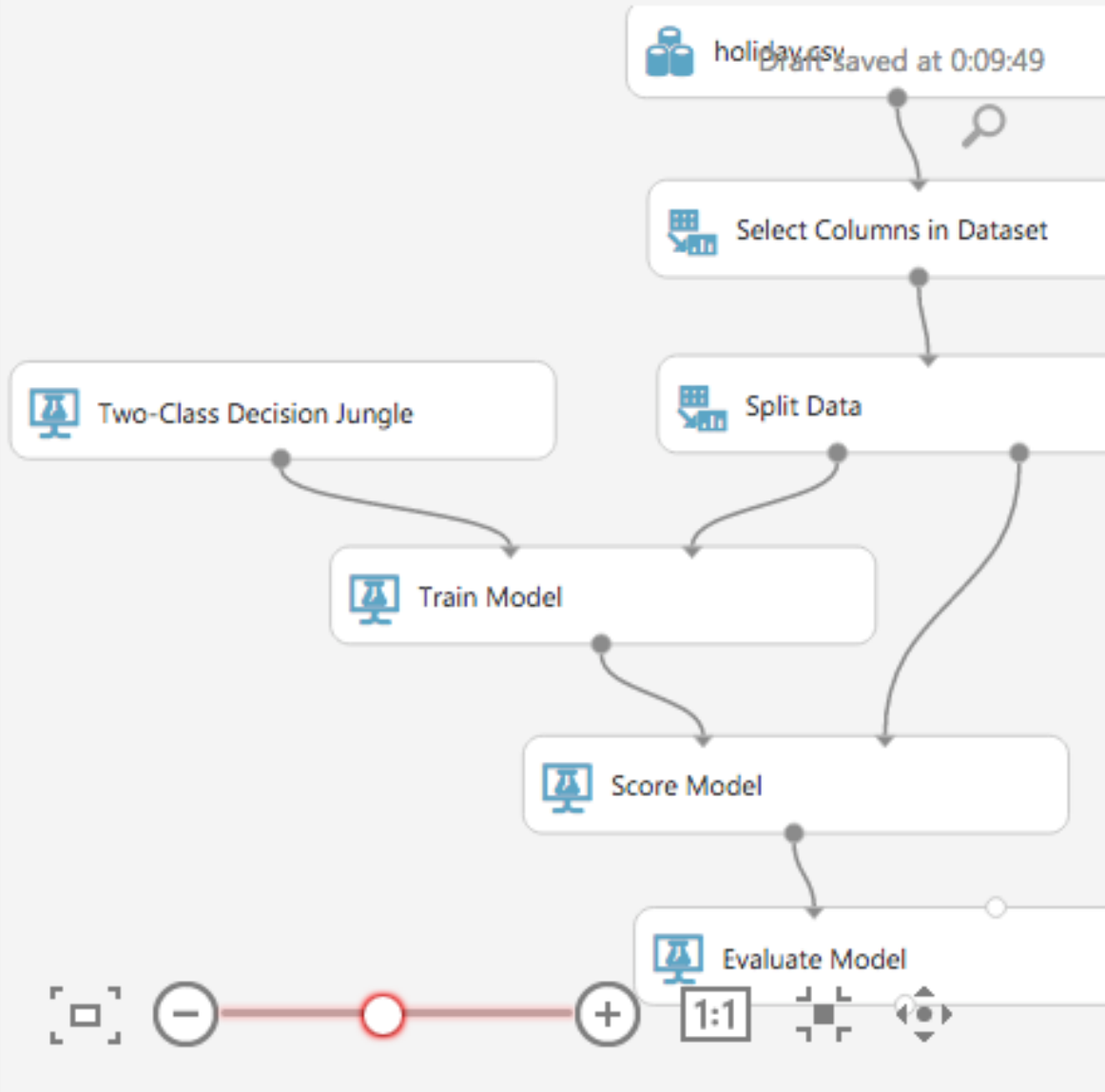
Initialize Model

Classification

- Two-Class Averaged ...
- Two-Class Bayes Poi...
- Two-Class Boosted ...
- Two-Class Decision F...
- Two-Class Decision J...
- Two-Class Locally-D...
- Two-Class Logistic R...
- Two-Class Neural Ne...
- Two-Class Support V...

Experiment created

In draft



Properties Project

Experiment Properties

START TIME	3/13/2017 ...
END TIME	3/13/2017 ...
STATUS CODE	InDraft
STATUS DETAILS	None

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help



# ハンズオン

- 学習をスタートします
  - 画面下の Run をクリックします
  - Evaluate Model の下をクリックし、Visualize で結果を確認します
  - Train Model の下をクリックし、Visualize でどういうモデルが作成されたかを確認できます
- 📖 結果を見て、あとは試行錯誤の繰り返しです



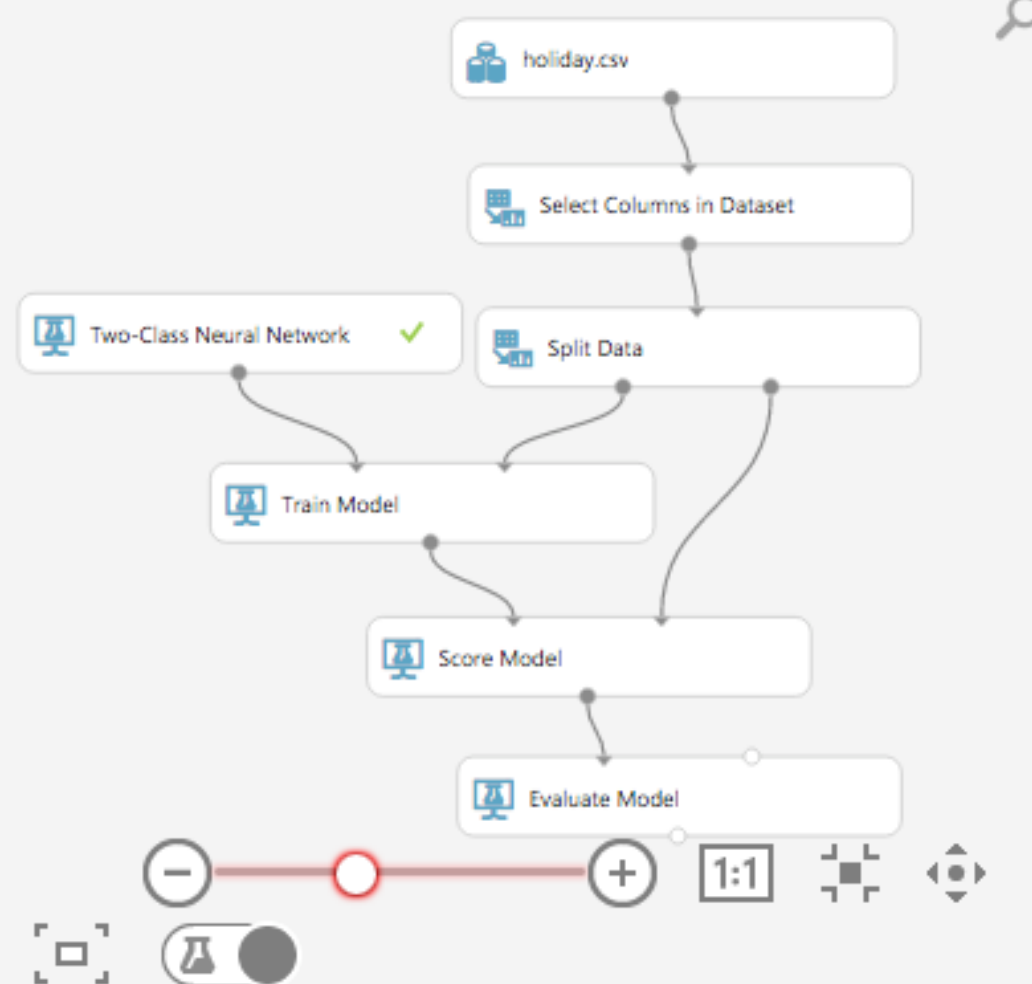
Search experiment items

- ▶ Saved Datasets
- ▶ 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

## Experiment created ...

In draft

Draft saved at 0:18:06



Properties Project

### Experiment Properties

START TIME	3/13/2017 ...
END TIME	3/13/2017 ...
STATUS CODE	InDraft
STATUS DETAILS	None

### Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help

+ NEW

RUN HISTORY

SAVE

SAVE AS

DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY



Experiment created Finished running ✓ Properties Project

Experiment created on 2017/3/12 > Evaluate Model > Evaluation results



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
144	74	0.986	0.980	0.5	0.989
False Positive	True Negative	Recall	F1 Score		
3	5258	0.661	0.789		
Positive Label	Negative Label				
True	False				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	109	0	0.020	0.980	0.667	1.000	0.500	0.980	1.000	0.000
(0.800,0.900]	19	1	0.024	0.983	0.738	0.992	0.587	0.983	1.000	0.000

Breast cancer data



NEW



RUN HISTORY



SAVE



SAVE AS



DISCARD CHANGES



RUN



SET UP WEB SERVICE



PUBLISH TO GALLERY

# ハンズオン

- SetUp WebService からWebAPI へデプロイします
- Predictive Web Service でWebAPIリクエストの入出力を設定します





Search experiment items

Saved Datasets

My Datasets

holiday.csv

Samples

Adult Census Income Bi...

Airport Codes Dataset

Automobile price data (...)

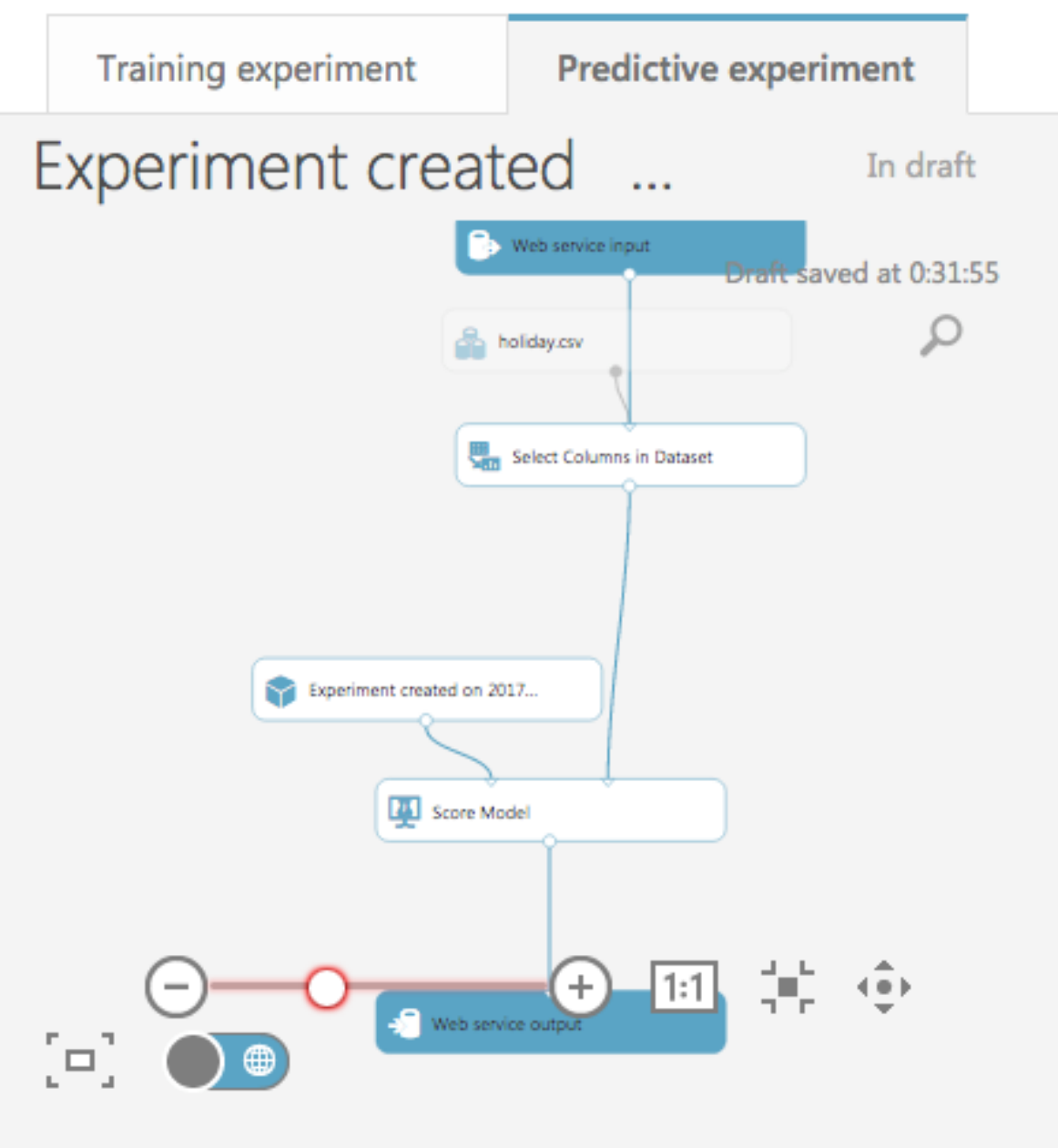
Bike Rental UCI dataset

Bill Gates RGB Image

Blood donation data

Book Reviews from Am...

Breast cancer data



Properties | Project

Experiment Properties

START TIME	3/13/2017 ...
END TIME	3/13/2017 ...
STATUS CODE	InDraft
STATUS DETAILS	None

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help

# ハンズオン

- 1度 Run してから Web Deploy します
  - すぐにデプロイ完了します
  - WebServiceの管理画面はプレビュー機能の新UIのほうが便利です
- 📖 Swagger API もサポートしています



# experiment created on 2017/3/12 [predictive exp.]

DASHBOARD CONFIGURATION

General

**New Web Services Experience** preview

Published experiment

[View snapshot](#)

[View latest](#)

Description

No description provided for this web service.

API key

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	<a href="#">Test</a> <a href="#">Test <small>preview</small></a>	<a href="#">Excel 2013 or later</a>   <a href="#">Excel 2010 or ear</a>	3/13/2017 12:27:36 AM
BATCH EXECUTION	<a href="#">Test <small>preview</small></a>	<a href="#">Excel 2013 or later workbook</a>	3/13/2017 12:27:36 AM



POST

/execute?api-version=2.0&format=swagger

Parameters

Execution request

body

```
{
  "Inputs": {
    "input1": [
      {
        "year": 1,
        "month": 1,
        "day": 1,
        "holiday": false,
        "title": ""
      }
    ]
  },
  "GlobalParameters": {}
}
```

The API version to use<sup>REQUIRED</sup>

query

string

時間があれば補足

# AMLの中で行われたこと

year	month	day	holiday
2016	1	1	TRUE
2017	1	1	TRUE
2018	1	1	TRUE
...	1	1	TRUE?

# AMLの中で行われたこと

year	month	day	holiday
2016	1	11	TRUE
2017	1	9	TRUE
2017	1	11	FALSE
2018	1	8	TRUE
2019	1	9	???

# AMLの中で行われたこと

- Excute Python(R) Script で自作コードも可

<https://goo.gl/Z2D0fr>

```
1  import pandas as pd
2  import datetime
3
4  def azureml_main(dataframe1 = None, dataframe2 = None):
5      ret = []
6      for index, row in dataframe1.iterrows():
7          date = datetime.date(int(row['year']),int(row['month']),int(row['day']))
8          ret.append({
9              'year':row['year'],
10             'month':row['month'],
11             'day':row['day'],
12             'holiday':row['holiday'],
13             'week':(row['day'] / 7) + 1,
14             'dayofweek':date.weekday()})
15      return pd.DataFrame(ret),
```



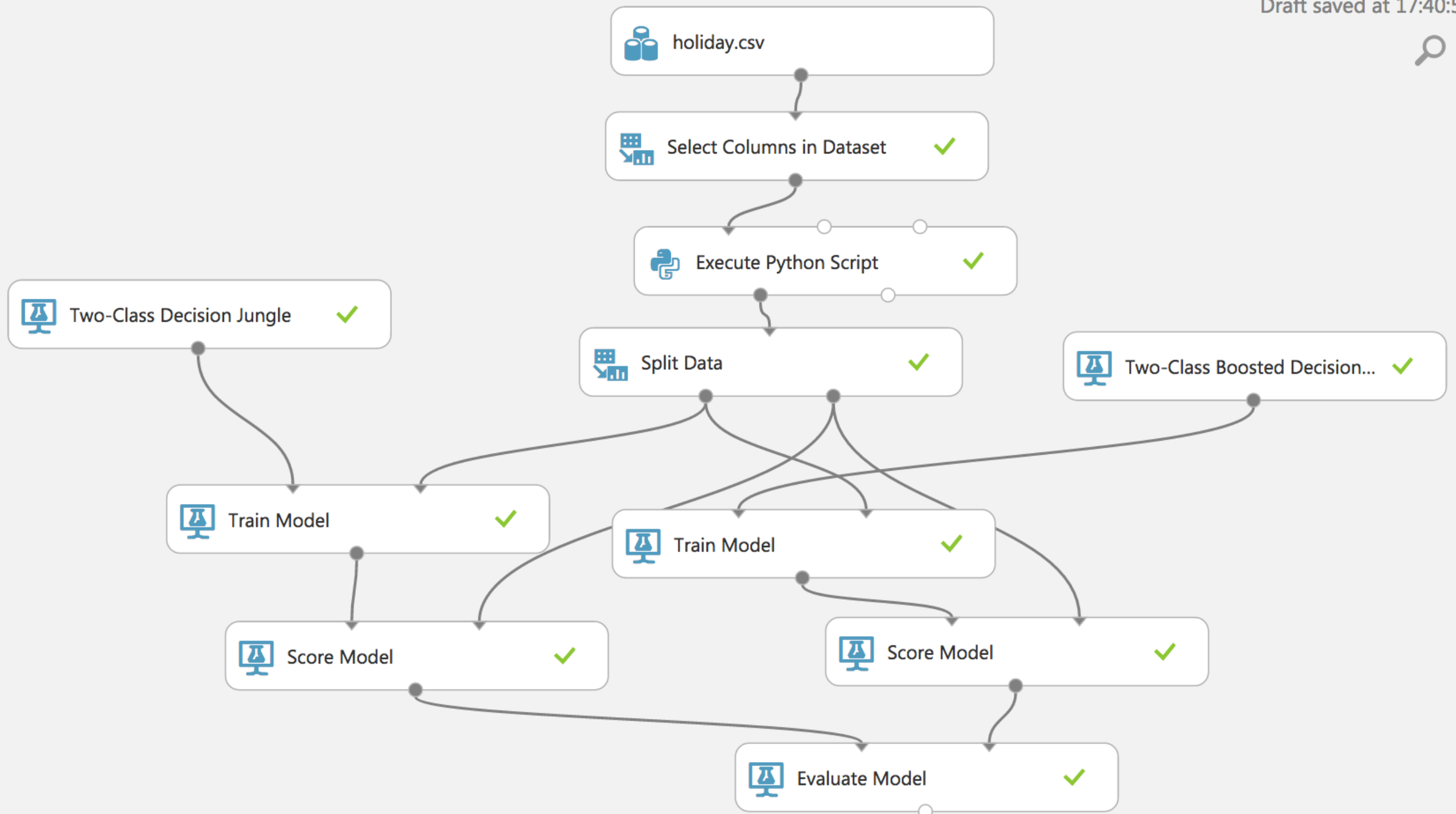
# AMLの中で行われたこと

year	month	day	dayof week	week	holiday
2016	1	11	1	2	TRUE
2017	1	9	1	2	TRUE
2017	1	11	3	2	FALSE
2018	1	8	1	2	TRUE
2019	1	9	3	2	FALSE?

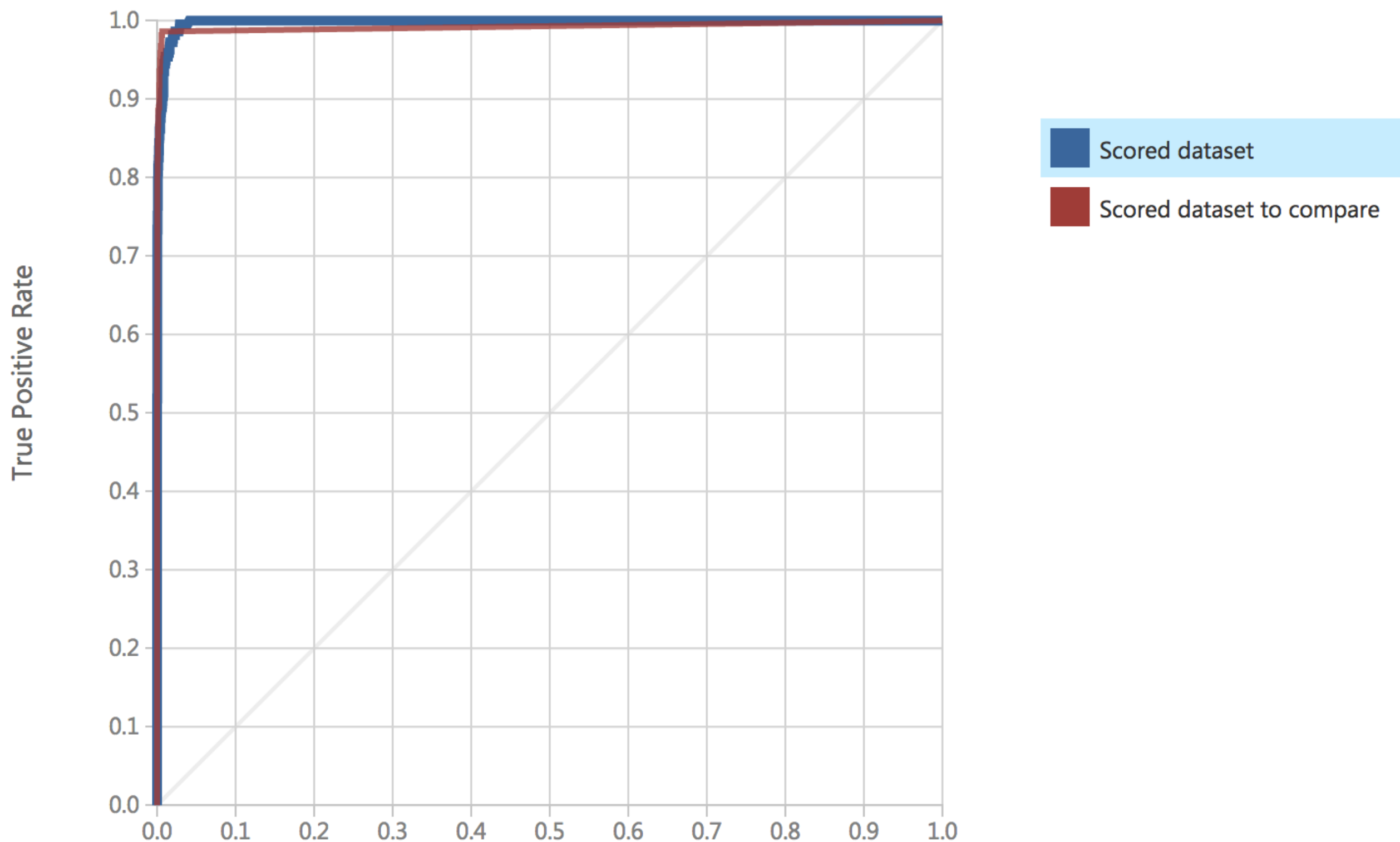
他にもできること

# アルゴリズムの結果比較

Draft saved at 17:40:58



ROC PRECISION/RECALL LIFT

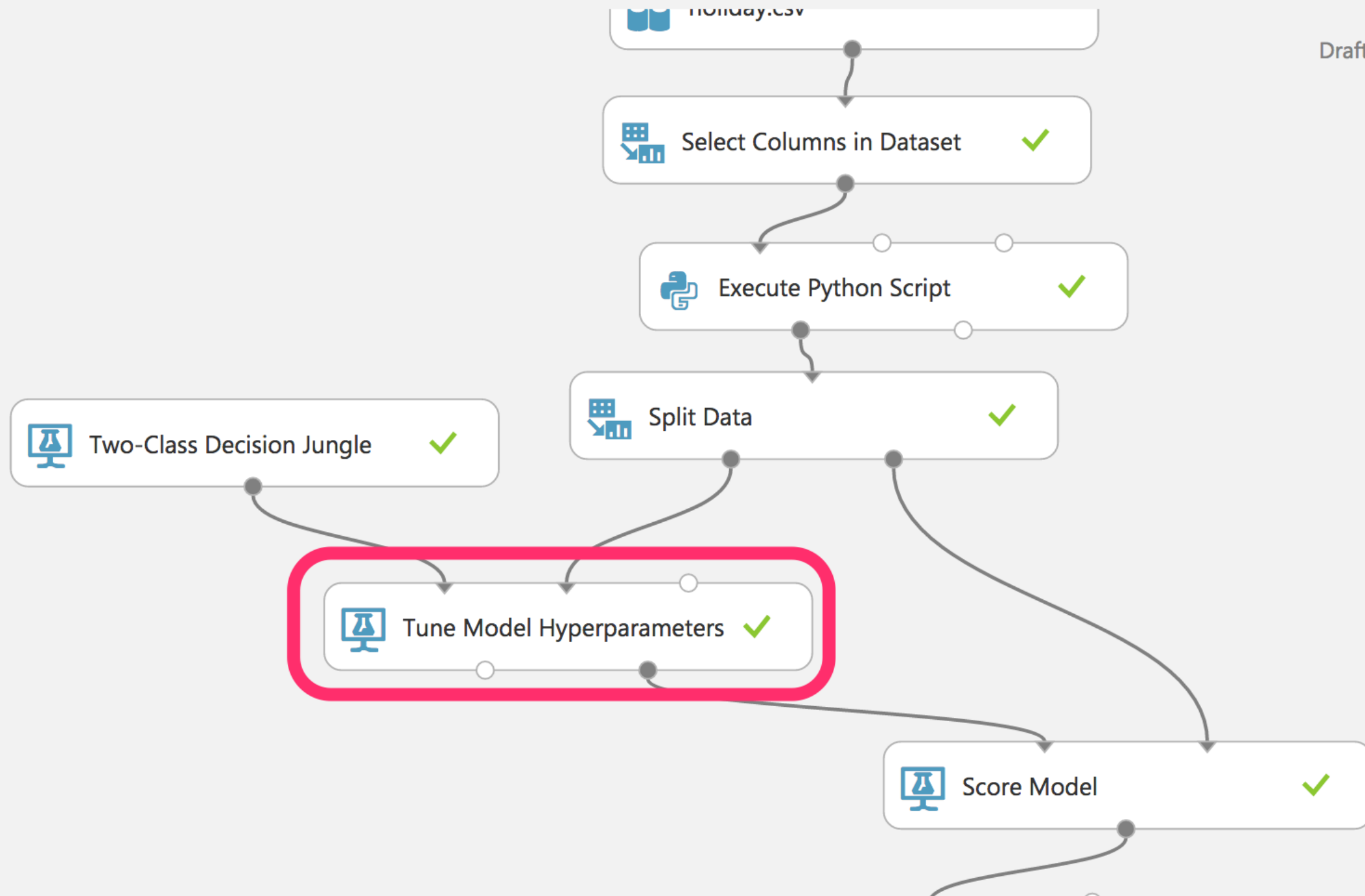


# パラメータの自動チューニング

meetup-0323 - tuning

Finished running ✓

Draft saved at 17:46:10



ここまでハンズオン

おつかれさまでした

# まとめ

- Azure Machine Learning は非エンジニアでも使える
- 機械学習でも少しずつ動くようになるフィードバックループで学習スピードを高めよう
- 適切なデータを集めることが大事



Machine Learning in ML Studio

Anomaly Detection

- One-class Support Vector Machine
- Principal Component Analysis-based Anomaly Detection
- Time Series Anomaly Detection\*

Classification

Two-class Classification

- Averaged Perceptron
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- Support Vector Machine

Multi-class Classification

- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-all

Clustering

- K-means Clustering

Recommendation

- Matchbox Recommender

Regression

- Bayesian Linear Regression
- Boosted Decision Tree
- Decision Forest
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

Statistical Functions

- Descriptive Statistics
- Hypothesis Testing T-Test
- Linear Correlation
- Probability Function Evaluation

Text Analytics

- Feature Hashing
- Named Entity Recognition
- Vowpal Wabbit

Computer Vision

- OpenCV Library

Data/Model Visualization

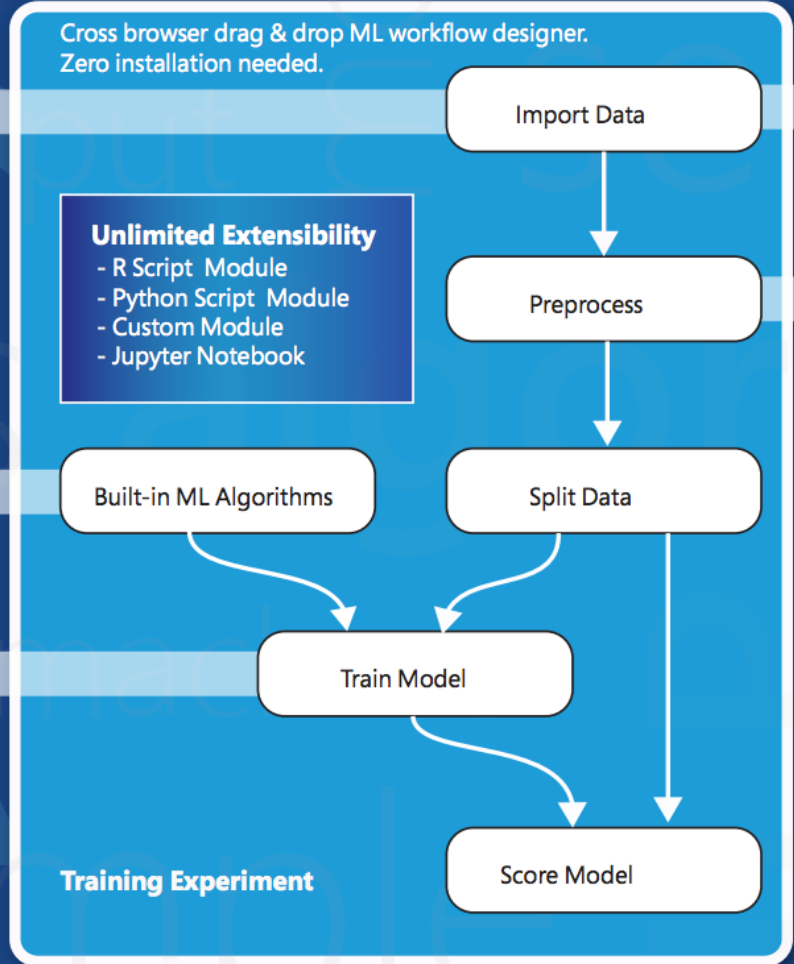
- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree\*

Training

- Cross Validation
- Retraining
- Parameter Sweep

<https://studio.azureml.net>

Guest Access Workspace: Free trial access without logging in.  
Free Workspace: Free persisted access, no Azure subscription needed.  
Standard Workspace: Full access with SLA under an Azure subscription.



Data Source

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW\*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server\*
- Web URL (HTTP)

Data Format

- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Data Preparation

- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQLite Transformation
- Synthetic Minority Oversampling Technique

Enterprise Grade Cloud Service

- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance\*

Community

- Gallery (<http://gallery.azureml.net>)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support

\* Feature Coming Soon

One-click Operationalization

Predictive Experiment

Make Prediction with Elastic APIs

- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API



Azure Machine Learning Studio Capabilities Overview



# Azure関連サービス

- AzureStorage  
Blob, Table(KVS)
- AzureDataFactory  
サービス間のデータ移行支援
- AzureFunctions  
サーバーレスで RestAPI を実装
- CognitiveServices  
画像認識、音声認識などのWebAPI群

# 参考資料

# 参考

- 公式:よくある質問

<https://docs.microsoft.com/ja-jp/azure/machine-learning/machine-learning-faq>

- 非公式:よくある質問

<https://blogs.msdn.microsoft.com/kosasaki/2016/04/13/azure-machine-learning-%E3%82%92%E5%88%A9%E7%94%A8%E3%81%99%E3%82%8B%E3%81%8A%E5%AE%A2%E6%A7%98%E3%81%8B%E3%82%89%E3%81%AE%E3%82%88%E3%81%8F%E3%81%82%E3%82%8B%E8%B3%AA%E5%95%8F/>

# 参考

- 公式:Microsoft Azure Machine Learning のアルゴリズムの選択方法

<https://docs.microsoft.com/ja-jp/azure/machine-learning/machine-learning-algorithm-choice>

- 非公式:日本語版アルゴリズムチートシート

<https://ty-memo.azurewebsites.net/azure/microsoft-azure-machine-learning-%E6%97%A5%E6%9C%AC%E8%AA%9E%E7%89%88%E3%82%A2%E3%83%AB%E3%82%B4%E3%83%AA%E3%82%BA%E3%83%A0%E3%83%81%E3%83%BC%E3%83%88%E3%82%B7%E3%83%BC%E3%83%88%E3%82%92%E5%85%AC/>

# 参考

- 公式:学習アルゴリズム

<https://msdn.microsoft.com/library/en-us/Dn905812.aspx>

- Azure Machine Learning 分析アルゴリズムの選択

<http://qiita.com/hiiyan0402/items/3e1584b25bfb6f453f52>

- 機械学習アルゴリズム解説スライドまとめ

<https://freelance.levtech.jp/guide/detail/41/>

- 代表的な機械学習手法一覧

<http://qiita.com/tomomoto/items/b3fd1ec7f9b68ab6dfe2>