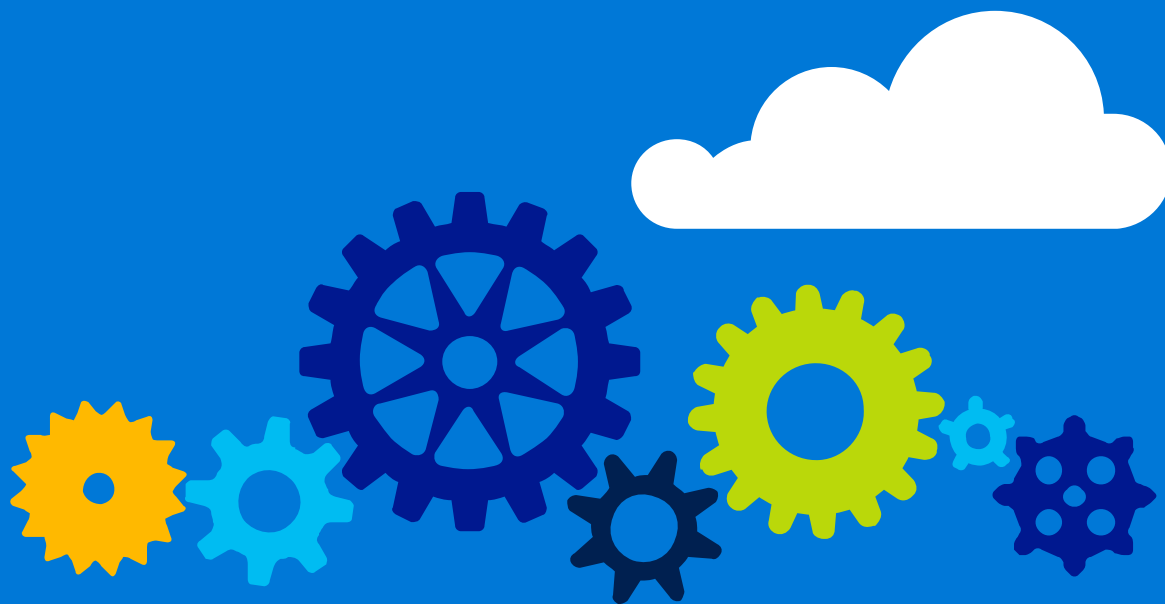


予兆保全のサンプル紹介

Version 1.0



本ドキュメントの目的

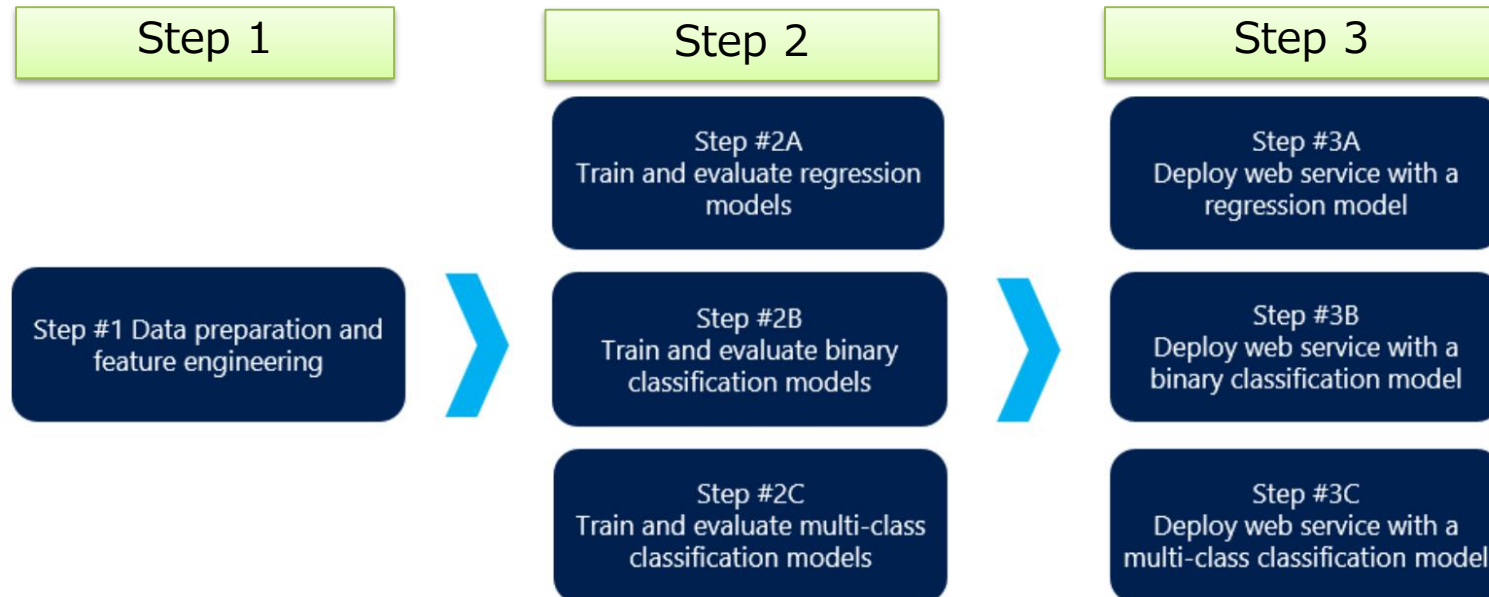
- ・ 予兆保全のサンプルをご紹介します
 - ・ 本サンプルの「解説」ではありません
 - ・ 詳細は記載した URL のページを参考にしてください
- ・ 予兆保全のサンプルは、以下を目的としたものです
 - ・ データサイエンスの一般的なプロセスを理解します
 - ・ 機械学習のアプローチによる故障予測の方法を示します

予兆保全のサンプル URL 一覧

- ・ 以下の7ページに分かれているが、説明内容は全て同じ。「Open in Studio」をクリックして開く「Experiments」が異なる。
 - ・ Step 1 of 3 - data preparation and feature engineering
 - ・ <https://gallery.cortanaintelligence.com/Experiment/df7c518dcba7407fb855377339d6589f>
 - ・ Step 2A of 3 - train and evaluate regression models
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2A-of-3-train-and-evaluate-regression-models-2>
 - ・ Step 2B of 3 - train and evaluate binary classification models
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2B-of-3-train-and-evaluate-binary-classification-models-2>
 - ・ Step 2C of 3, train and evaluation multi-class classification models
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2C-of-3-train-and-evaluation-multi-class-classification-models-2>
 - ・ Step 3A of 3, deploy web service with a regression model
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3A-of-3-deploy-web-service-with-a-regression-model-2>
 - ・ Step 3B of 3, deploy web service with a binary classification model
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3B-of-3-deploy-web-service-with-a-binary-classification-model-2>
 - ・ Step 3C of 3, deploy web service with a multi-class classification model
 - ・ <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3C-of-3-deploy-web-service-with-a-multi-class-classification-model-2>

サンプルの構造 – 3個のStep

- Step1 : Data preparation and feature engineering
 - ・ データの準備と特徴量エンジニアリング
- Step2 : Train and evaluate model
 - ・ 学習と評価のモデリング
 - ・ Regression モデル／Binary classification モデル／Multi-class classification モデル
- Step3 : Deploy web service
 - ・ 出来上がったモデルをWebサービスとして配置



サンプルで利用するデータ

- Prognostics Center of Excellence
 - <https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/>
 - NASAが予兆アルゴリズム (prognostic algorithms) 開発のために公開しているデータリポジトリ
- Turbofan Engine Degradation Simulation Data Set
 - <https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/#turbofan>
 - 4種類のデータセット (FD001~FD004) のうち、1つ (FD001) を利用

Step1 :

Data preparation and feature engineering

Step1 のプロセス

- ・ Step1-1 : データの読み込み
- ・ Step1-2 : ラベルの作成
- ・ Step1-3 : 特徴量エンジニアリング

Step1-1：3種類のデータを利用

- ・ 100個のデバイスの時系列データ
 - ・ 各デバイスの各時点における状態が1レコードとして記録されている
- ・ 3種類のデータ
 - ・ ①トレーニングデータ（Training data）：train_FD001.txt
 - ・ モデリングを作成するために利用
 - ・ レコードが途絶えたタイミングが故障を示す
 - ・ ②テストデータ（Testing data）：test_FD001.txt
 - ・ モデルをテストするために利用
 - ・ トレーニングデータと同じ構造であるが、故障を示すイベントは含まれない
つまり、レコードが途絶えたタイミングは故障を意味しない
 - ・ ③解答データ（Ground truth data）：RUL_FD001.txt
 - ・ ②のデバイスが実際に故障した際のデータ
 - ・ 予測モデルを②でテストする際、③の解答にどれだけ近いかという観点で評価

Step1-1：データスキーマ

- ① トレーニングデータと②テストデータのスキーマ

Index	Data fields	Type	Descriptions	
1	id	Integer	aircraft engine identifier, range [1, 100]	デバイスのID
2	cycle	Integer	time, in cycles	時間の流れ
3	setting1	Double	operational setting 1	
4	setting2	Double	operational setting 2	
5	setting3	Double	operational setting 3	
6	s1	Double	sensor measurement 1	21個のセンサーから取得した値
7	s2	Double	sensor measurement 2	
...	...			
26	s21	Double	sensor measurement 21	

Step1-1 : データの見方

Sample training data

~20k rows,
100 unique engine id

[①トレーニングデータ]

id=1のデバイスはcycle=192で終了している。ここで故障したことを意味する。

id	cycle	setting1	setting2	setting3	s1	s2	s3	...	s19	s20	s21
1	1	-0.0007	-0.0004	100	518.67	641.82	1589.7		100	39.06	23.419
1	2	0.0019	-0.0003	100	518.67	642.15	1591.82		100	39	23.4236
1	3	-0.0043	0.0003	100	518.67	642.35	1587.99		100	38.95	23.3442
...	...										
1	191	0	-0.0004	100	518.67	643.34	1602.36		100	38.45	23.1295
1	192	0.0009	0	100	518.67	643.54	1601.41		100	38.48	22.9649
2	1	-0.0018	0.0006	100	518.67	641.89	1583.84		100	38.94	23.4585
2	2	0.0043	-0.0003	100	518.67	641.82	1587.05		100	39.06	23.4085
2	3	0.0018	0.0003	100	518.67	641.55	1588.32		100	39.11	23.425
...	...										
2	286	-0.001	-0.0003	100	518.67	643.44	1603.63		100	38.33	23.0169
2	287	-0.0005	0.0006	100	518.67	643.85	1608.5		100	38.43	23.0848

Sample testing data

~13k rows,
100 unique engine id

[②テストデータ]

id=1のデバイスはcycle=31で終了している。しかし、ここで故障したことを意味ものではない。

id	cycle	setting1	setting2	setting3	s1	s2	s3	...	s19	s20	s21
1	1	0.0023	0.0003	100	518.67	643.02	1585.29		100	38.86	23.3735
1	2	-0.0027	-0.0003	100	518.67	641.71	1588.45		100	39.02	23.3916
1	3	0.0003	0.0001	100	518.67	642.46	1586.94		100	39.08	23.4166
...	...										
1	30	-0.0025	0.0004	100	518.67	642.79	1585.72		100	39.09	23.4069
1	31	-0.0006	0.0004	100	518.67	642.58	1581.22		100	38.81	23.3552
2	1	-0.0009	0.0004	100	518.67	642.66	1589.3		100	39	23.3923
2	2	-0.0011	0.0002	100	518.67	642.51	1588.43		100	38.84	23.2902
2	3	0.0002	0.0003	100	518.67	642.58	1595.6		100	39.02	23.4064
...	...										
2	48	0.0011	-0.0001	100	518.67	642.64	1587.71		100	38.99	23.2918
2	49	0.0018	-0.0001	100	518.67	642.55	1586.59		100	38.81	23.2618
3	1	-0.0001	0.0001	100	518.67	642.03	1589.92		100	38.99	23.296
3	2	0.0039	-0.0003	100	518.67	642.23	1597.31		100	38.84	23.3191
3	3	0.0006	0.0003	100	518.67	642.98	1586.77		100	38.69	23.3774
...	...										
3	125	0.0014	0.0002	100	518.67	643.24	1588.64		100	38.56	23.227
3	126	-0.0016	0.0004	100	518.67	642.88	1589.75		100	38.93	23.274

Sample ground truth data

100 rows

[③回答データ]

id=1~100に対応した100レコードが存在。各デバイスが故障したcycleを示す。

RUL
112
98
69
82
91

Step1-2 : 3種類の予測を実施する

- ・ 予測A. デバイスは、何サイクルで故障するか？
 - ・ サイクル数（数値）を予測する必要がある
 - ・ 回帰モデル（Regression models）
- ・ 予測B. デバイスは、 $w1$ サイクル以内に故障するか？
 - ・ 例えば 30（= $w1$ ）サイクル以内に故障するのかどうか？故障する or 故障しないの2択で予測する必要がある
 - ・ 2分類モデル（Binary classification）
- ・ 予測C. デバイスは、 $w0$ or $w1$ サイクル以内に故障するか？
 - ・ 例えば 15（= $w2$ ）もしくは、30（= $w1$ ）サイクル以内に故障するのかどうか？15サイクル以内／30サイクル以内／故障しないの3択で予測する必要がある
 - ・ 多分類モデル（Multi-class classification）

Step1-2 : ラベルの作成

- ①トレーニングデータと③解答データにラベルを作成
 - RUL : 予測Aのために作成
 - Label1 : 予測Bのために作成
 - Label2 : 予測Cのために作成

			新規作成			w0 = 15 w1 = 30
id	cycle	...	RUL	label1	label2	
	1	1	191	0	0	
	1	2	190	0	0	
	1	3	189	0	0	
	1	4	188	0	0	
...			
	1	160	32	0	0	w1
	1	161	31	0	0	
	1	162	30	1	1	
	1	163	29	1	1	
	1	164	28	1	1	
	1	165	27	1	1	
	1	166	26	1	1	
	1	167	25	1	1	
	1	168	24	1	1	
	1	169	23	1	1	
	1	170	22	1	1	
	1	171	21	1	1	
	1	172	20	1	1	
	1	173	19	1	1	
	1	174	18	1	1	
	1	175	17	1	1	
	1	176	16	1	1	w0
	1	177	15	1	2	
	1	178	14	1	2	
	1	179	13	1	2	
	1	180	12	1	2	
	1	181	11	1	2	
	1	182	10	1	2	
	1	183	9	1	2	
	1	184	8	1	2	
	1	185	7	1	2	
	1	186	6	1	2	
	1	187	5	1	2	
	1	188	4	1	2	
	1	189	3	1	2	
	1	190	2	1	2	
	1	191	1	1	2	
	1	192	0	1	2	

Step1-3：特徴量を追加

- ・ 特徴量とは、モデルに入力するインプット
- ・ 本サンプルにおける既存の特徴量は「Step1-1：データスキーマ」のとおり
 - ・ id, cycle, setting1-setting3, s1-s21
- ・ さらに、新しく以下の特徴量を追加する（w=5）
 - ・ a1-a21：当該レコードを含めて過去 w サイクル分のレコードの平均値
 - ・ sd1-sd21：当該レコードを含めて過去 w サイクル分のレコードの標準偏差

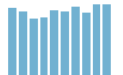


id	cycle	s3	a3	sd3
				
1	1	1589.7	1589.7	0
1	2	1591.82	1590.76	1.499066
1	3	1587.99	1589.836667	1.918654
1	4	1582.79	1588.075	3.855909
1	5	1582.85	1587.03	4.075678
1	6	1584.47	1585.984	3.885831
1	7	1592.32	1586.084	4.075375

Diagram illustrating the calculation of a3 (average) and sd3 (standard deviation) for cycle 5, using the previous 5 cycles (cycles 1-5) as the window (w=5).

For cycle 5, the values of s3 for cycles 1 through 5 are: 1589.7, 1591.82, 1587.99, 1582.79, and 1582.85. These values are used to calculate the average (a3) and standard deviation (sd3) for cycle 5.

The calculated values for cycle 5 are: a3 = 1587.03 (average) and sd3 = 4.075678 (standard deviation).

Step1-3：その他の処理

- ・ 不要な特徴量を削除
 - ・ 本サンプルでは id 列を削除
- ・ 正規化（normalization）の実施
 - ・ 特徴量の尺度を揃える
- ・ ②テストデータで不要なレコードを削除
 - ・ 各デバイスの最大サイクルレコードのみを残す
 - ・ 結果、100レコードとなる

Step2 :
Train and evaluate model

Step 2A :
train and evaluate
regression models



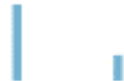

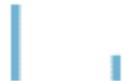


Step2A-1：事前作業

- ・ Label（目的変数）に “RUL” を設定

(本ドキュメントでは説明対象外となります。評価結果のみを記載します)






Step2A : Regression models の評価

- Decision Forest Regression / Boosted Decision Tree Regression モデルの予測性能が相対的に高い
- 20~30のズレ幅で予測できる

Algorithms	Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
						
Decision Forest Regression	462.724302192819	21.141062	30.147662	0.574998	0.526317	0.473683
Boosted Decision Tree Regression	Infinity	21.283579	29.615866	0.578874	0.507913	0.492087
Poisson Regression	NA	23.243022	29.973356	0.632167	0.520249	0.479751
Neural Network Regression	NA	75.505023	86.185055	2.053597	4.301346	-3.301346

Step2B : Binary classification の評価

- ・ w1 (=30) サイクル以内に故障するかどうかは、90%以上の確率で予測可能

Algorithms	Accuracy	Precision	Recall	F-Score
				
Logistic Regression	0.92	0.947368	0.72	0.818182
Boosted Decision Tree	0.91	0.9	0.72	0.8
Decision Forest	0.92	0.947368	0.72	0.818182
Neural Network	0.94	0.952381	0.8	0.869565

Step2C : Multi-class classification の評価

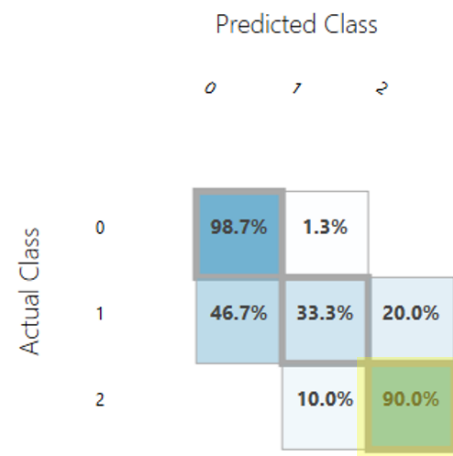
- ・ w0 (=15) サイクル以内に故障するかどうかは、90～100%以上の確率で予測可能

Multiclass Logistic Regression

Metrics

Overall accuracy	0.88
Average accuracy	0.92
Micro-averaged precision	0.88
Macro-averaged precision	0.792622
Micro-averaged recall	0.88
Macro-averaged recall	0.74

Confusion Matrix

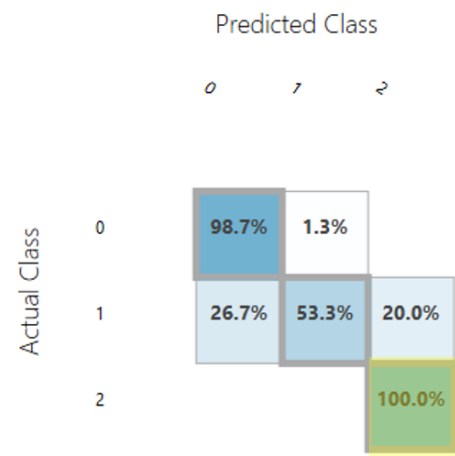


Multiclass Neural Network

Metrics

Overall accuracy	0.92
Average accuracy	0.946667
Micro-averaged precision	0.92
Macro-averaged precision	0.868946
Micro-averaged recall	0.92
Macro-averaged recall	0.84

Confusion Matrix



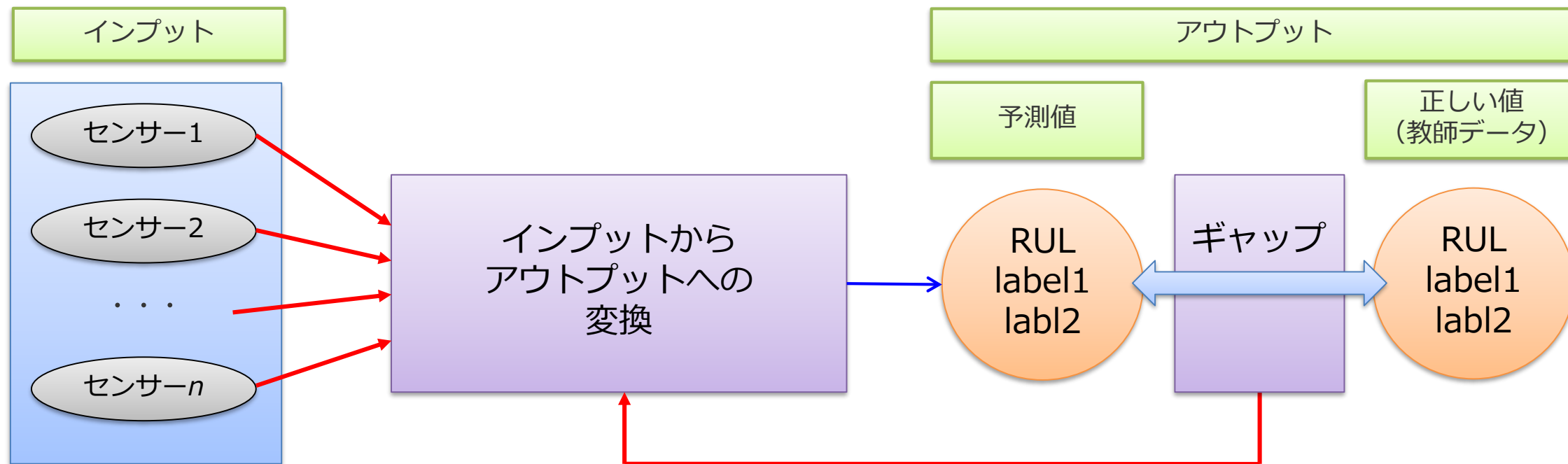
Step3 :
Deploy web service

(本ドキュメントでは説明対象外となります。)

Appendix

Appendix 1.

機械学習の用語の整理



イメージ

用語

特徴量、特徴ベクトル

説明変数

モデル

機械学習アルゴリズム

識別クラス、クラスラベル、ラベル

目的変数、ターゲット変数

Appendix 2.

Microsoft Azure Machine Learning Studio の起動

- ・ 各ページの「Open in Studio」をクリック
- ・ 任意のマイクロソフトアカウントでサインインして利用することが可能

EXPERIMENT

Predictive Maintenance: Step 1 of 3, data preparation and feature engineering

By AzureML Team for Microsoft • April 27, 2015

♡ 7 likes

Twitter LinkedIn Email


Summary

This experiment demonstrates the steps in building a predictive maintenance solution.

Description

Predictive Maintenance Template

Predictive maintenance encompasses a variety of topics, including but not limited to: failure prediction, failure diagnosis (root cause analysis), failure detection, failure type classification, and recommendation of mitigation or maintenance actions after failure. As part of the Azure Machine Learning offering, Microsoft provides a template that helps data scientists easily build and deploy a predictive maintenance solution.



Open in Studio

+ Add to Collection

ここをクリック

Appendix 3.

予兆保全に関する、その他の参考資料一覧

- Cortana Intelligence Solution Template Playbook for predictive maintenance in aerospace and other businesses
 - <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/cortana-analytics-playbook-predictive-maintenance>
- Predictive Maintenance Modelling Guide
 - <https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Modelling-Guide-Experiment-1>
- Predictive Maintenance Modelling Guide R Notebook
 - <https://gallery.cortanaintelligence.com/Notebook/Predictive-Maintenance-Implementation-Guide-R-Notebook-2>
- Predictive Maintenance Modelling Guide Python Notebook
 - <https://gallery.cortanaintelligence.com/Notebook/Predictive-Maintenance-Modelling-Guide-Python-Notebook-1>
- Deep Learning Basics for Predictive Maintenance
 - <https://gallery.cortanaintelligence.com/Tutorial/Deep-Learning-Basics-for-Predictive-Maintenance>