

# ノンコーディングで行う機械入門

Lesson 1 機械学習概要、ソフトウェアRapidMiner Studio概要と事例デモ・実習

Lesson 2 分類1: 主要なアルゴリズム説明と応用事例デモ・実習

Lesson 3 分類2: データ前処理と後処理、教師データと

テストデータの分割による分類問題の実習

**Lesson 4 分類3: 交差検証、最適アルゴリズム探索の実習**

Lesson 5 回帰: 主要なアルゴリズム説明と実習

Lesson 6 (応用) 時系列データの機械学習

Lesson 7 (応用) Extensionによる機能拡張と画像の分類

Lesson 8 自ら学ぶ: RapidMiner のウェブサイトの活用

# 今回取り扱うデータセット

天体の構成がパルサー星かどうかを判別する → True or False の**2値問題**

- (1) Mean of the integrated profile
- (2) Standard deviation of the integrated profile
- (3) Excess kurtosis of the integrated profile
- (4) Skewness of the integrated profile
- (5) Mean of the DM-SNR curve
- (6) Standard deviation of the DM-SNR curve
- (7) Excess kurtosis of the DM-SNR curve
- (8) Skewness of the DM-SNR curve



**上記8つの説明変数を基に、パルサーかどうか(目的変数)の判別を行う。**

→ 機械学習を行う上では**その分野の専門知識は一切なくても行える。**

(つまり、経験や勘に頼らなくてもよい。時には専門知識による常識が、機械学習の妨げになることもある)

# 機械学習プロセス

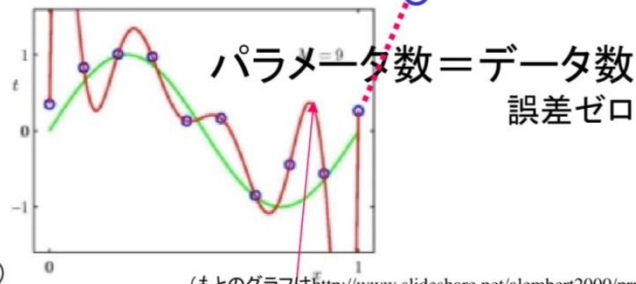
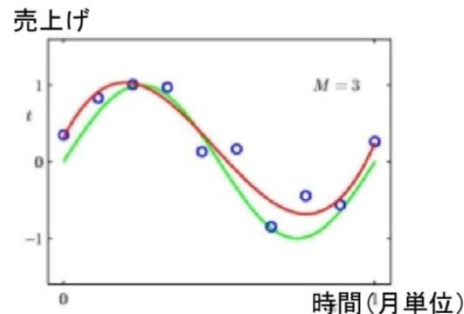
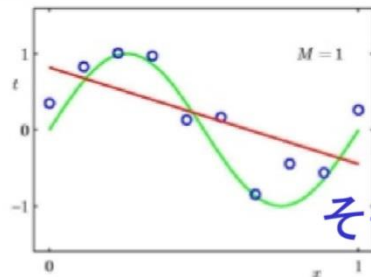
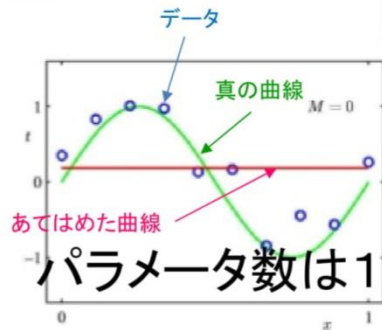


# (復習) 過学習

表現能力の高いモデルには過学習 (over fitting) が避けられない

## 過学習

- 与えられたデータにモデル(回帰曲線)が完全に一致すること
- 過去データ(既知の現象)の説明力は最高だが、未来データ(未知の現象)の予測力は最悪！



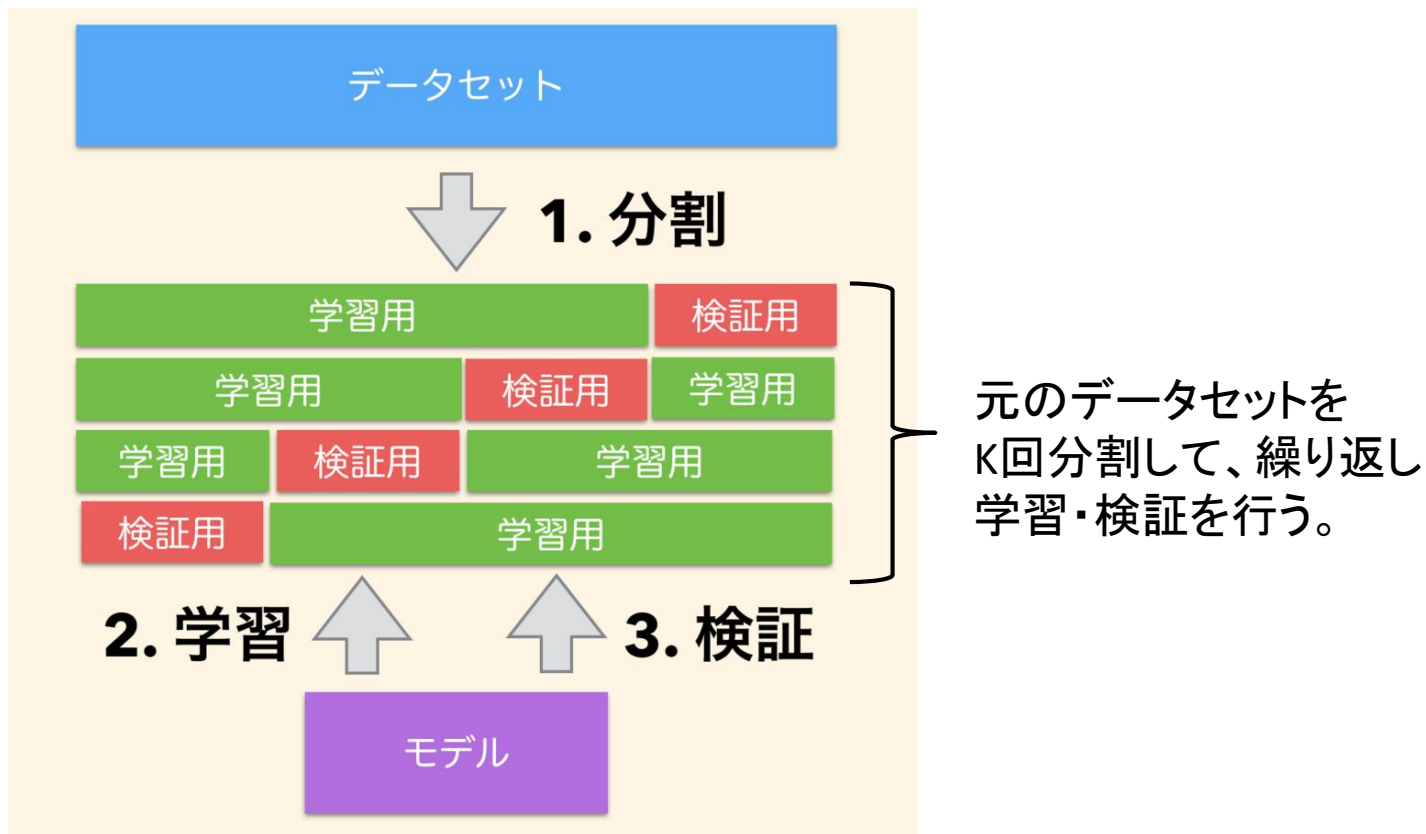
(もとのグラフは<http://www.slideshare.net/alembert2000/prml-at-1>より)

内挿の失敗:こんな予測を信じますか？

回帰曲線で起こる過学習は、モデルを複雑にすると、他の分類器でも起きうる。

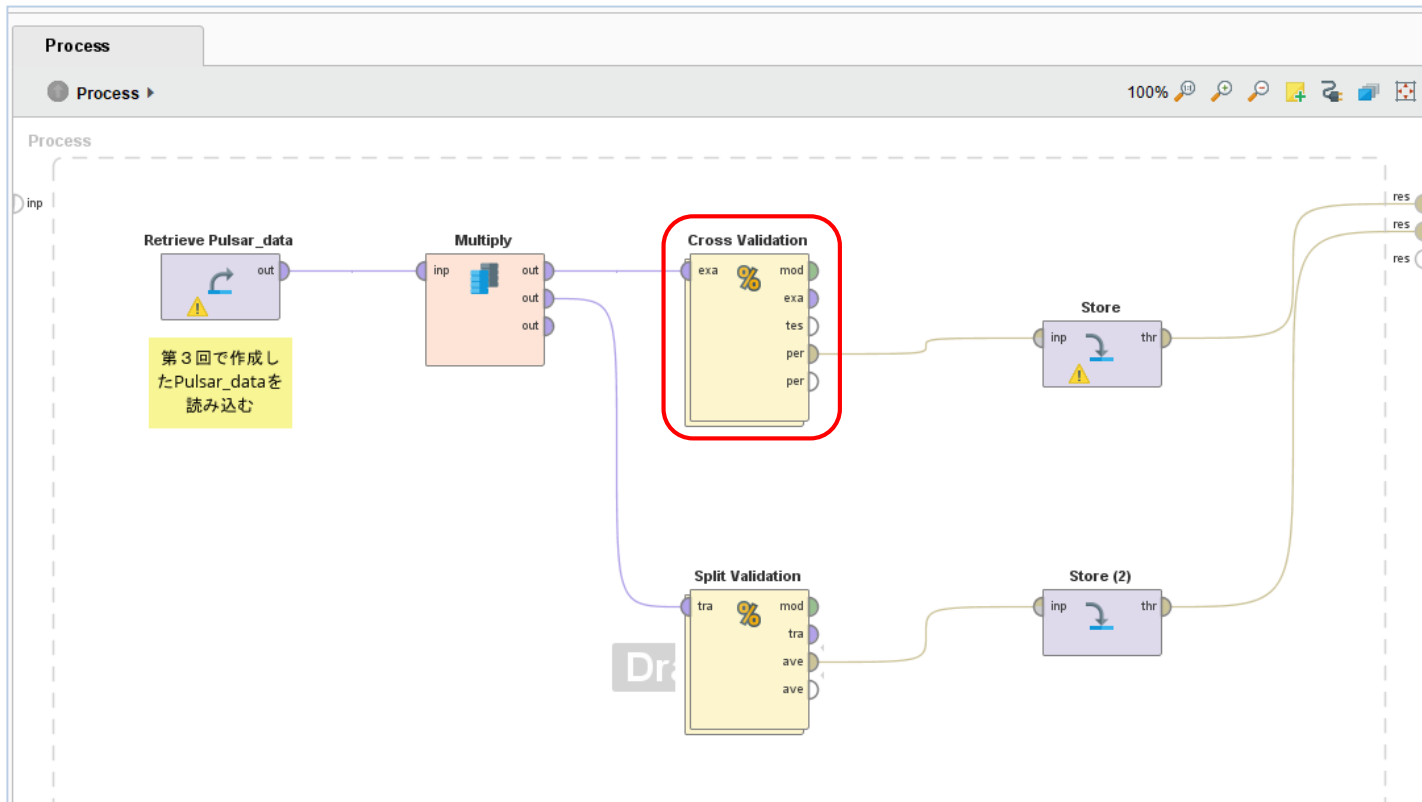
# 過学習と実際の精度検証

**交差検証** (Cross Validation) = k-分割交差検証



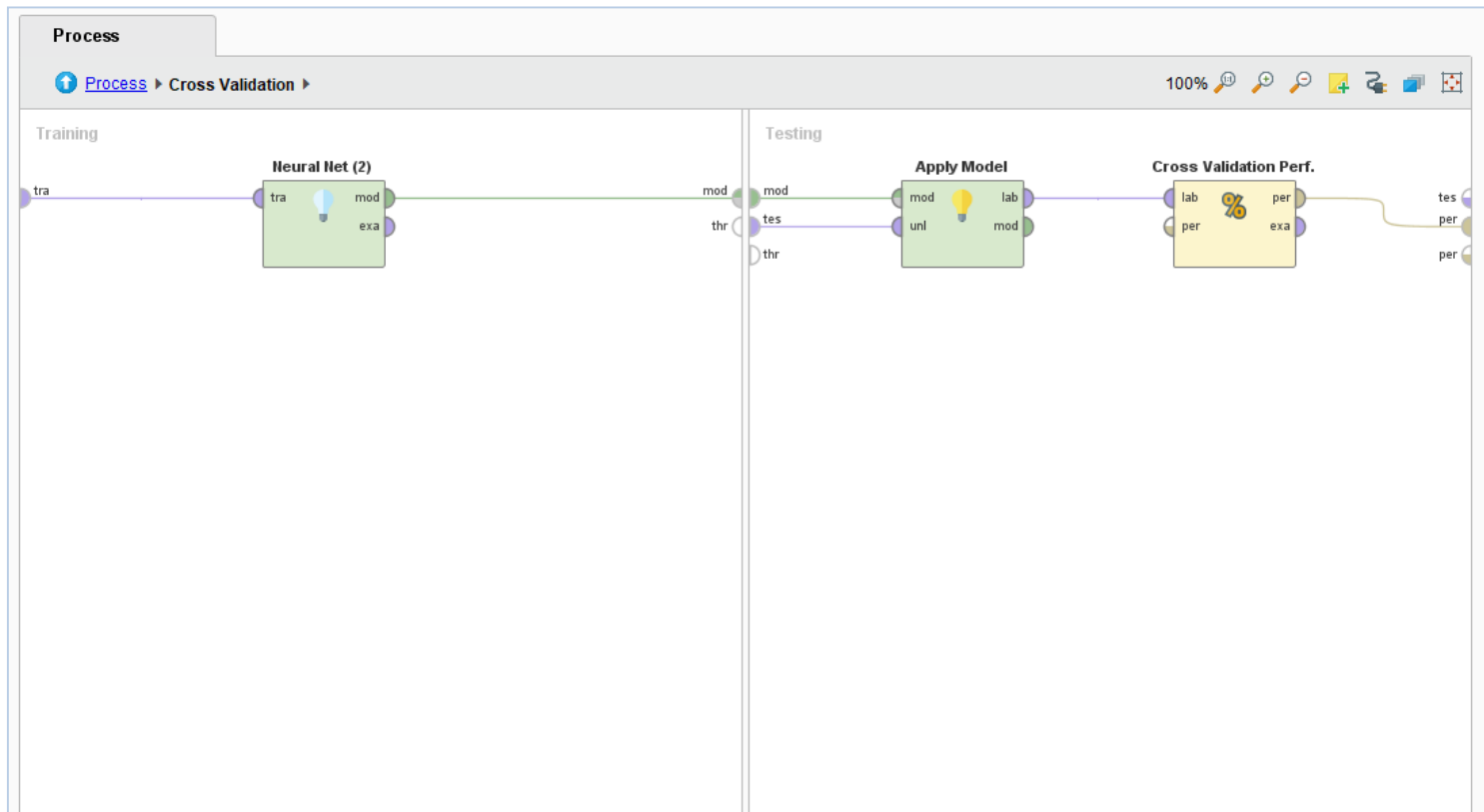
# 機械学習デモ＋実習

## 1. 交差検証 (NN, RF)



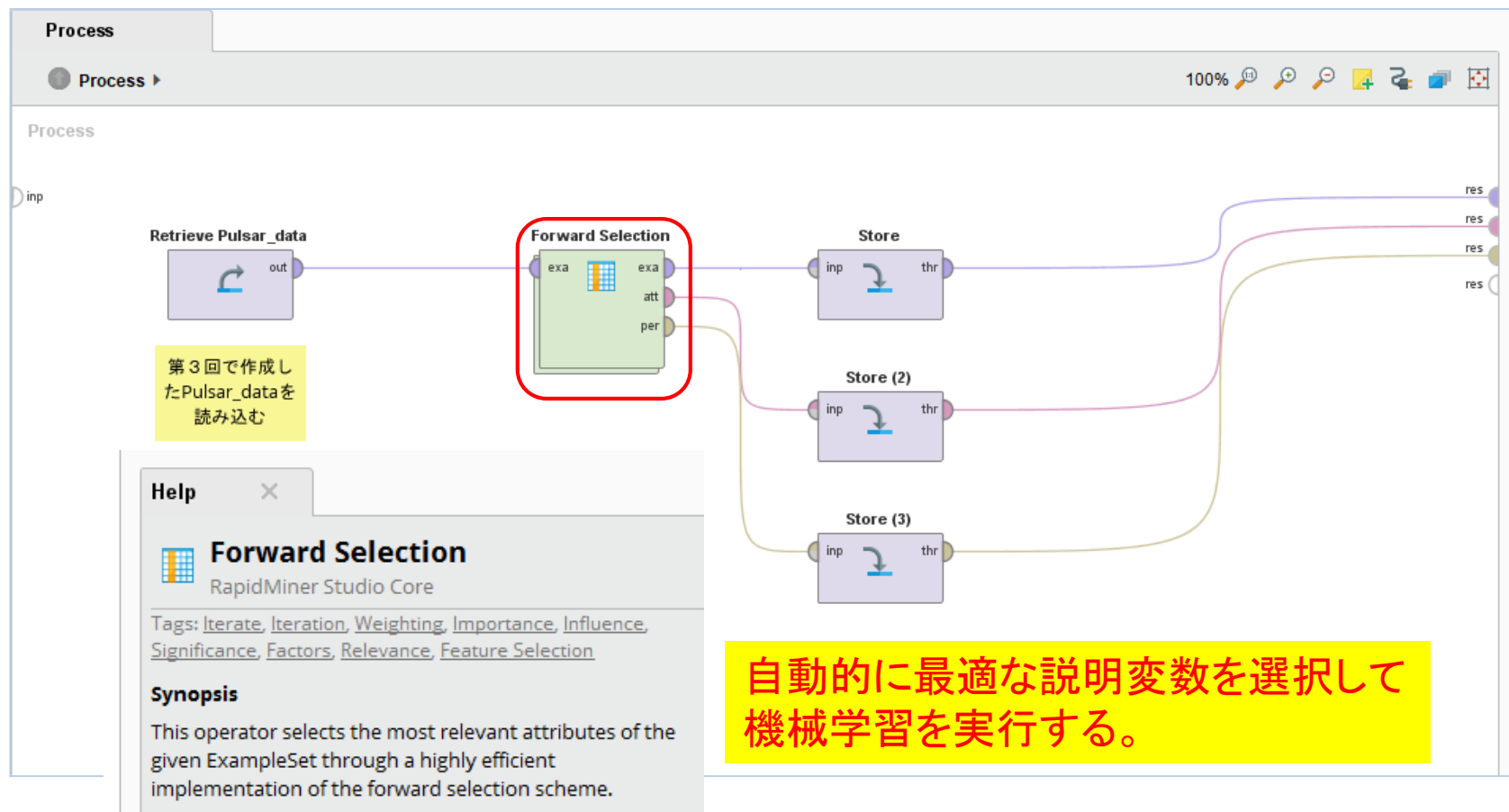
# 機械学習デモ＋実習

## 1. 交差検証 (NN, RF)



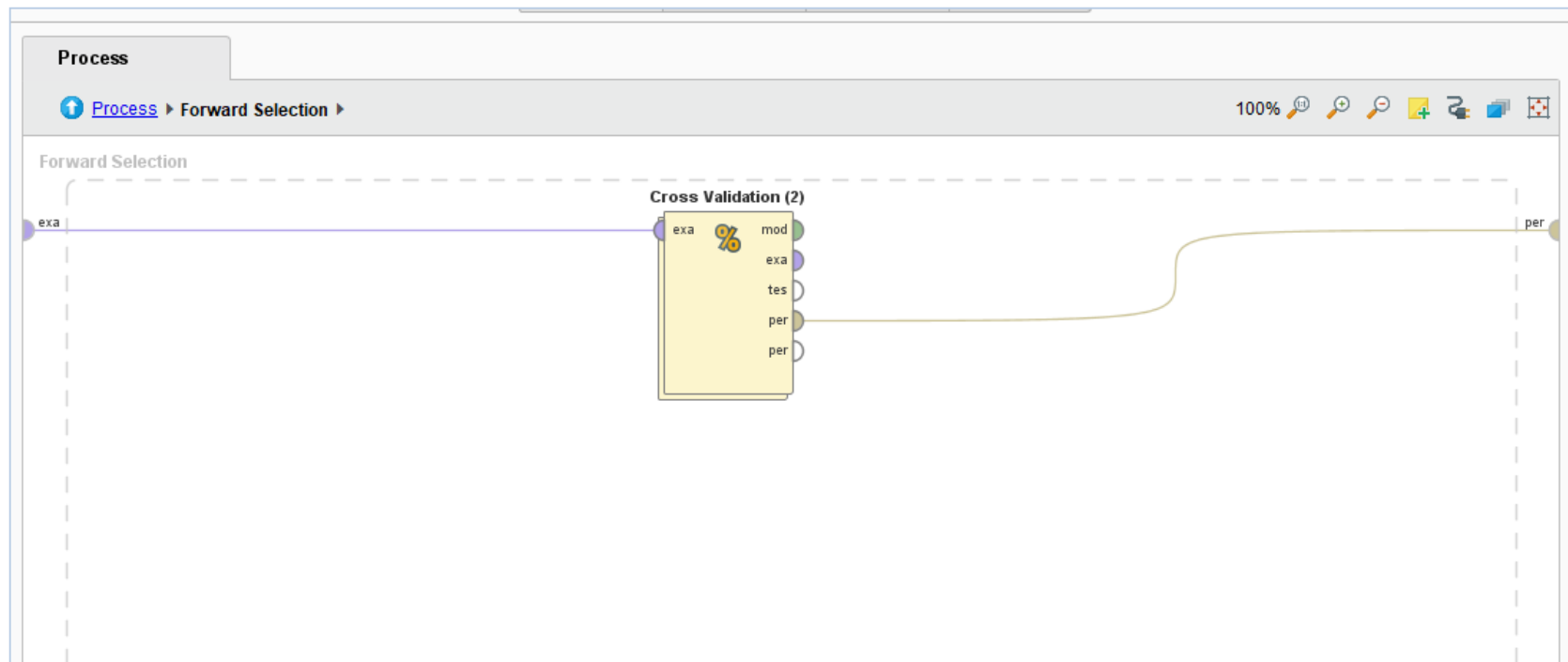
# 機械学習デモ＋実習

## 2. Forward Selection



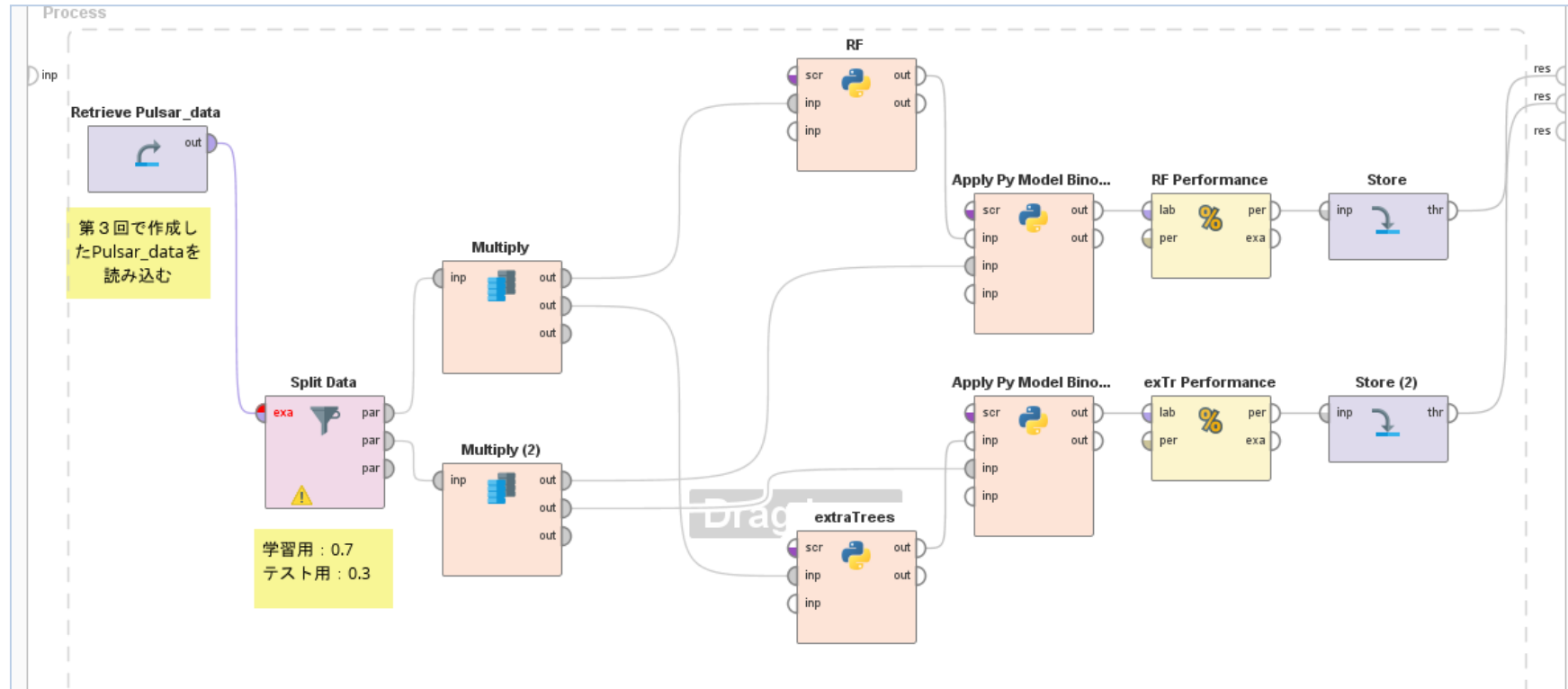


## 2. Forward Selection



# 機械学習デモ＋実習

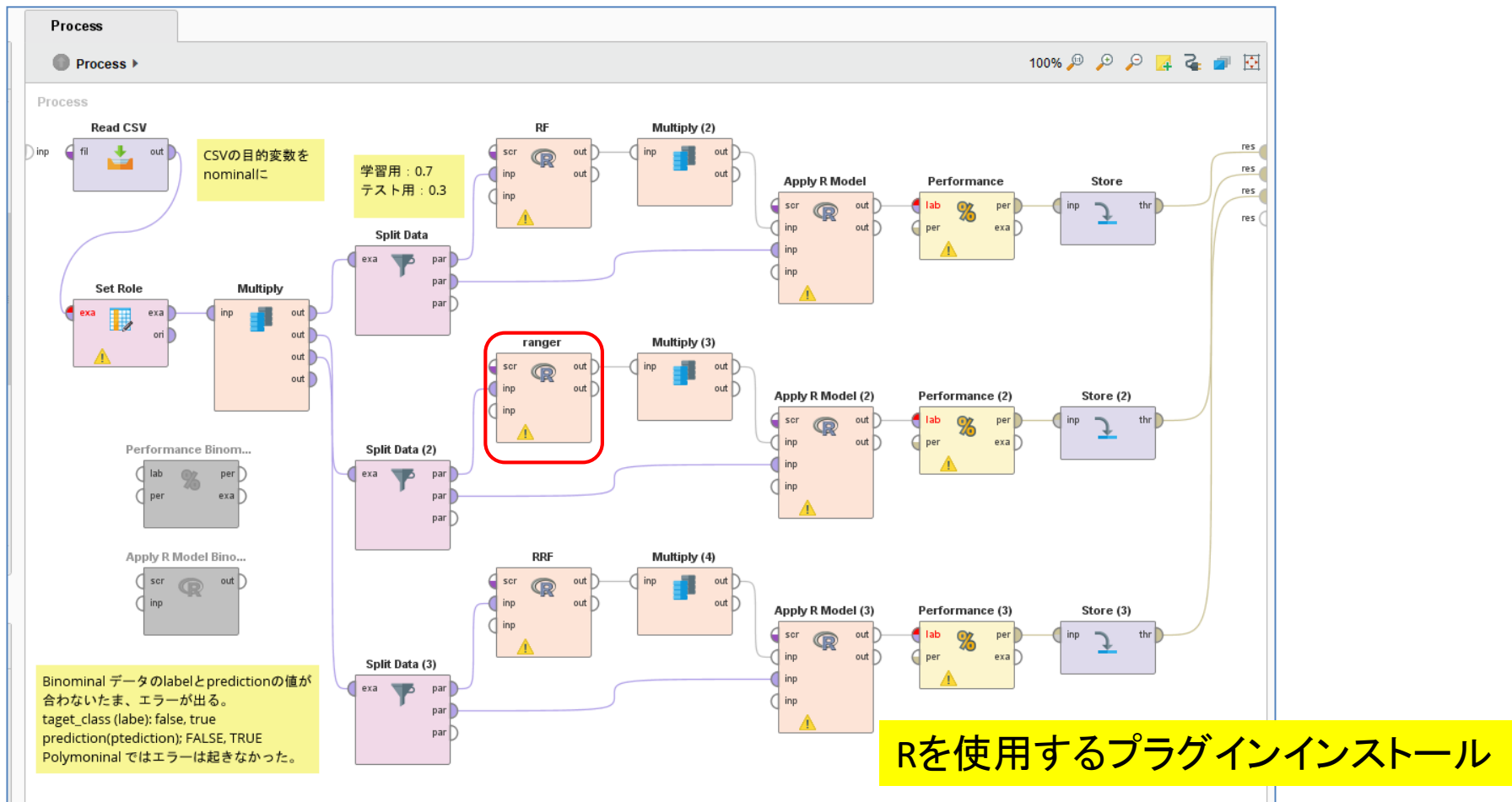
## 3. Python Script



Python Script を使用するプラグインインストール

# 機械学習デモ+実習

## 4. R Script



# 機械学習デモ＋実習

## 4. R Script 結果

### R Script ranger

accuracy: 98.42%			
	true nonPulsar	true Pulsar	class precision
pred. nonPulsar	4851	58	98.82%
pred. Pulsar	27	434	94.14%
class recall	99.45%	88.21%	

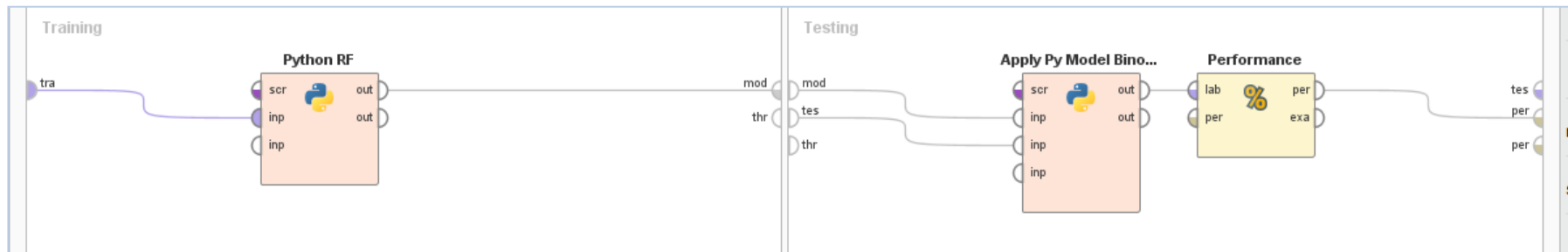
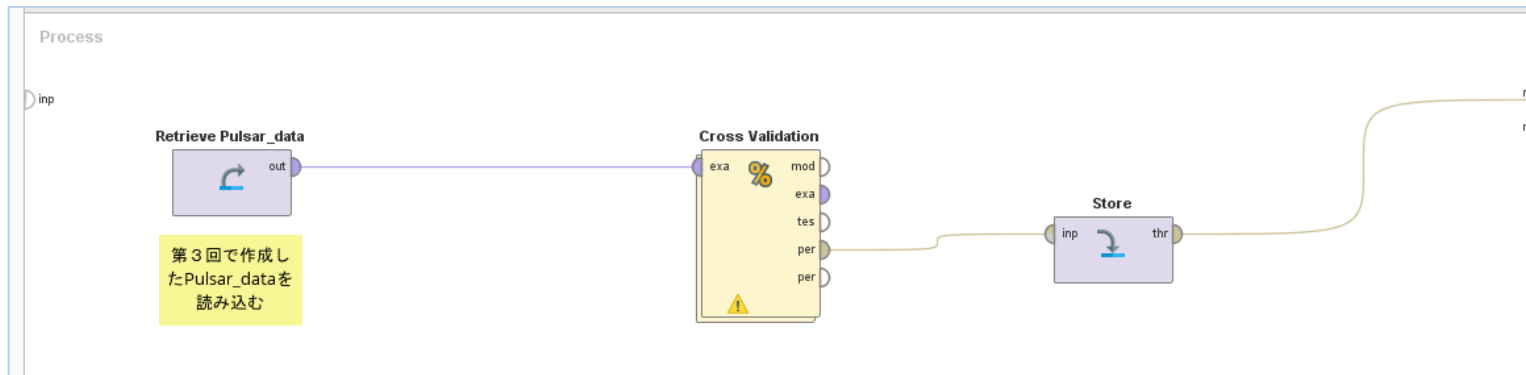
### RapidMiner Random Forest (3回目)

accuracy: 98.16%			
	true false	true true	class precision
pred. false	4872	81	98.36%
pred. true	18	398	95.67%
class recall	99.63%	83.09%	

計算速度: RapidMiner > Python >> R

# 機械学習デモ＋実習

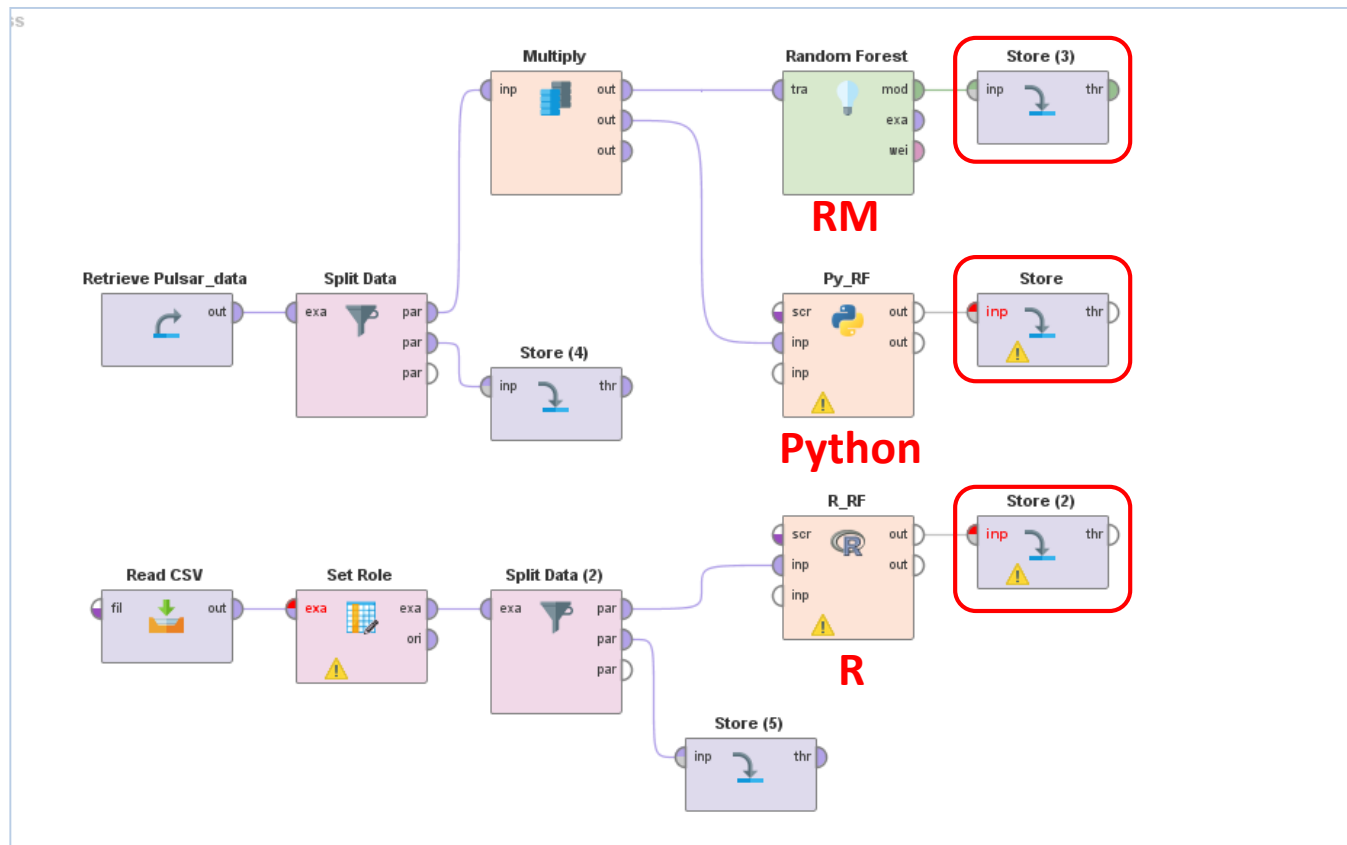
## 5. Cross Validation of Python Script



PythonやRのスクリプトもRMの分類器と同様に交差検証ができる

## 6. Modelの保存

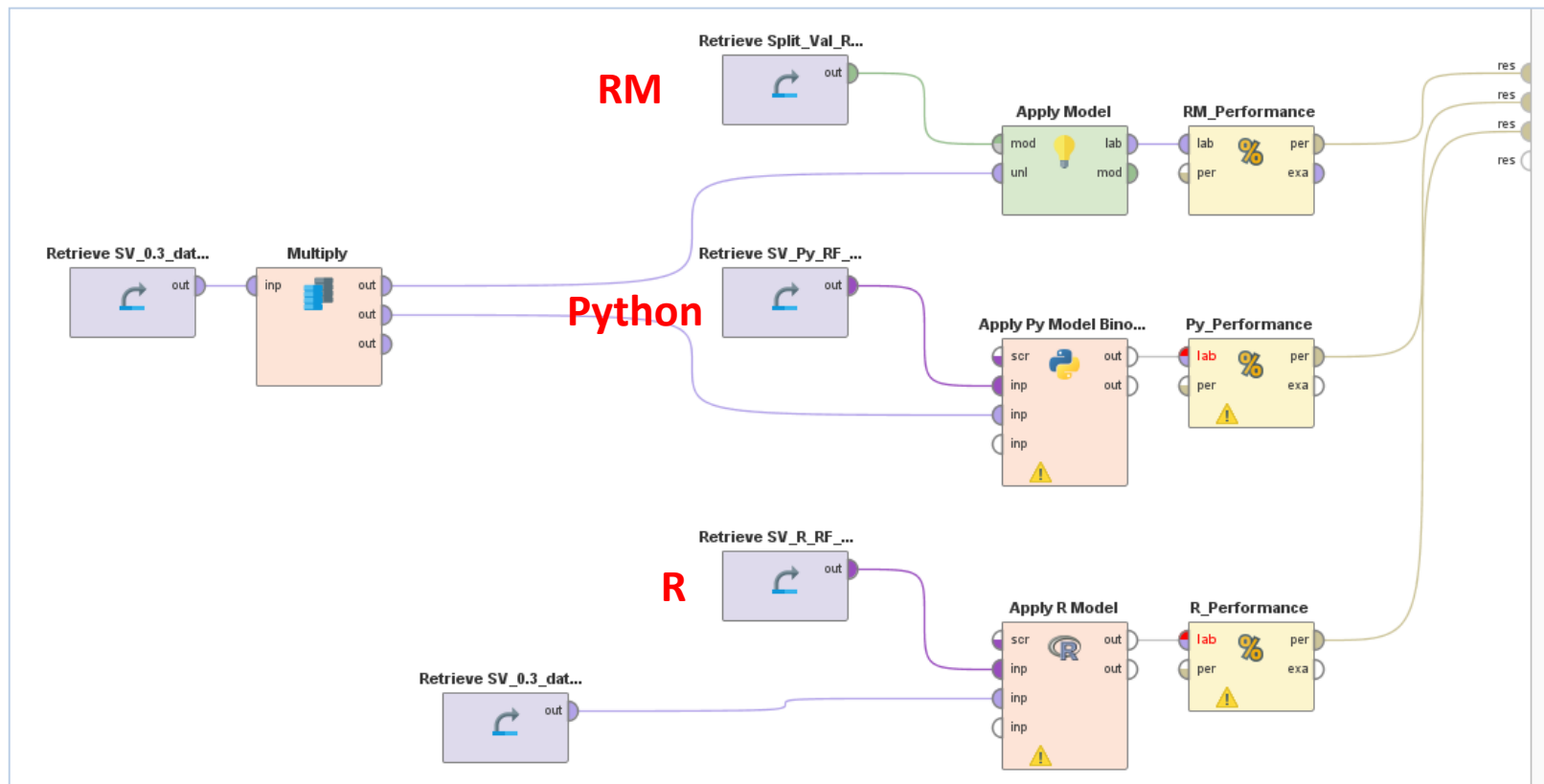
# 機械学習デモ＋実習



学習したモデル (RM, Python, R script) はメタデータとして保存できる

## 6. Modelの再利用

# 機械学習デモ＋実習



メタデータで保存したモデルでの予想は高速で行える