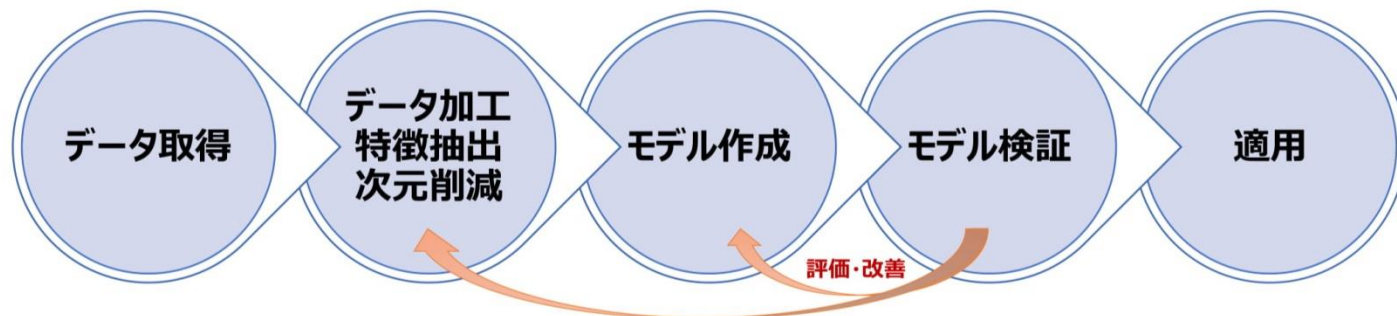


機械学習概要

データ分析の進め方と留意点



1. やりたいこと（目的変数）が明確であること
正常と異常の定義、判別したいものの定義、求めたい数値
2. データの質と量（説明変数）が十分であること
関係ないデータや間違い、欠損が少なく、統計処理に量も十分
3. 予測や判別ができたとして、対策が具体化できること
対策の内容、タイミング、体制、現場の希望を考慮

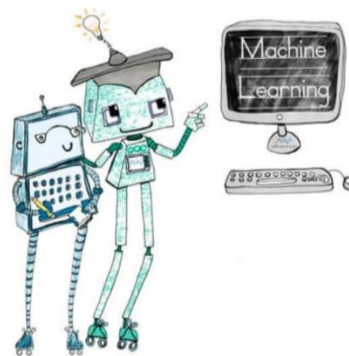
機械学習プロセス



機械学習概要

What Machine Learning ?

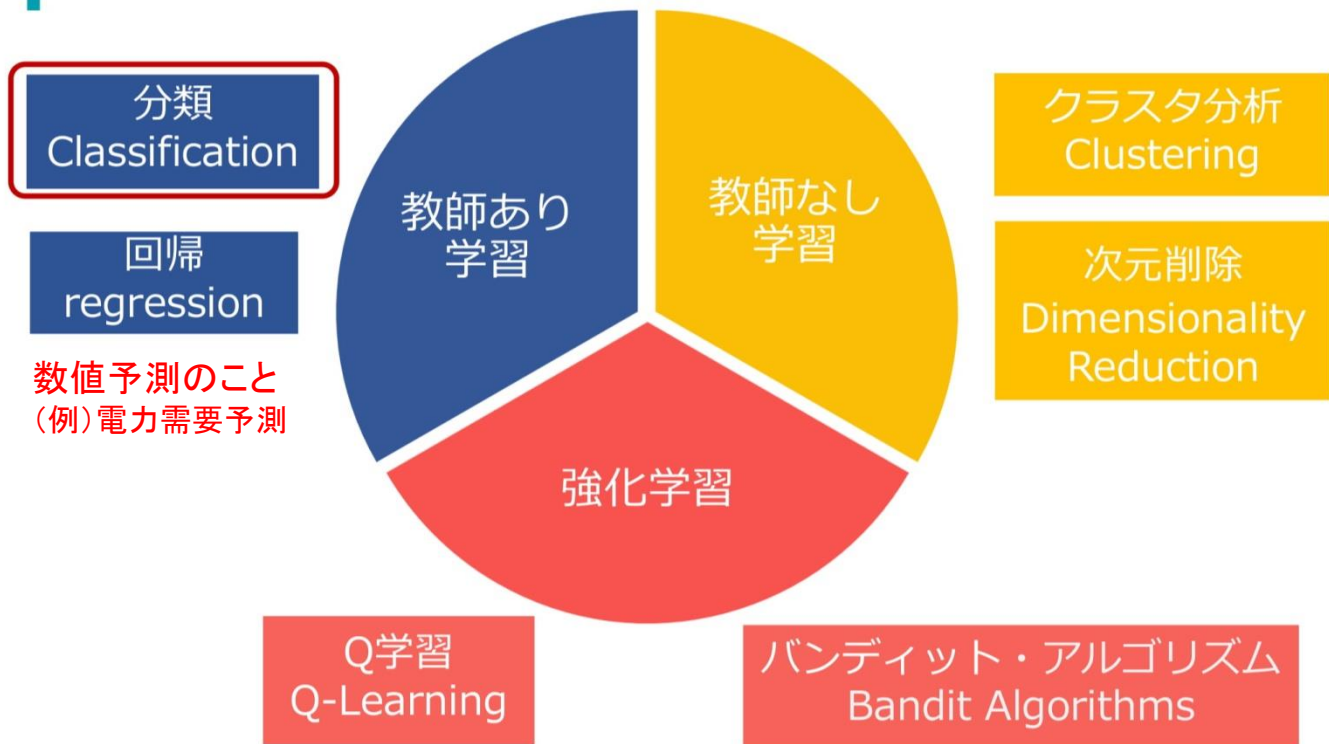
- ・ AI:人口知能の中核技術であり、コンピュータに人間のような問題解決能力を獲得させるための技術の総称
- ・ 大量のデータ（ビッグデータ）から
有用な知識を数式として掘り出す
データマイニングを目的として使用される
- ・ 扱うデータの種類によって、
「教師つき学習」「教師なし学習」「強化学習」
の3種類に分類できる



機械学習概要

What Machine Learning ?

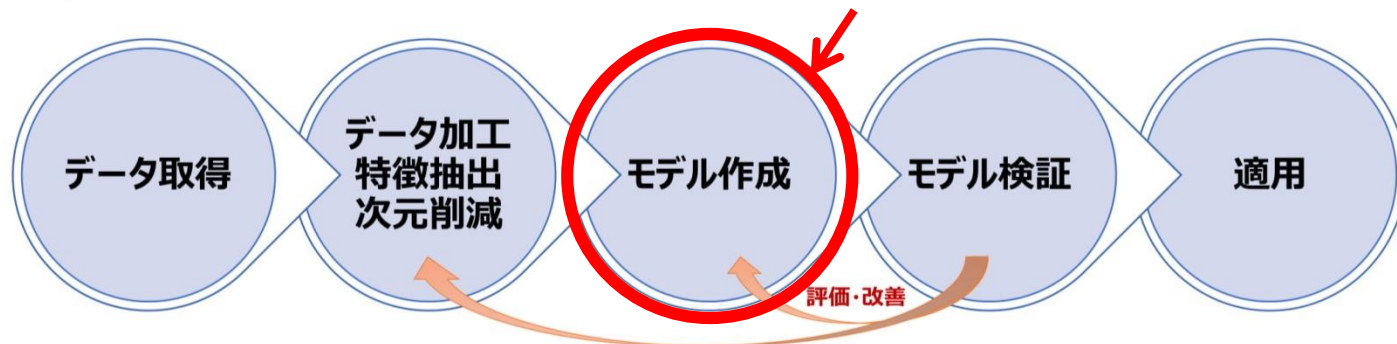
勉強会ではまず
分類をしっかり
やります



機械学習概要

データ分析の進め方と留意点

ここで各種の機械学習アルゴリズムを使う



1. やりたいこと（目的変数）が明確であること
正常と異常の定義、判別したいものの定義、求めたい数値
2. データの質と量（説明変数）が十分であること
関係ないデータや間違い、欠損が少なく、統計処理に量も十分
3. 予測や判別ができたとして、対策が具体化できること
対策の内容、タイミング、体制、現場の希望を考慮

機械学習のアルゴリズム

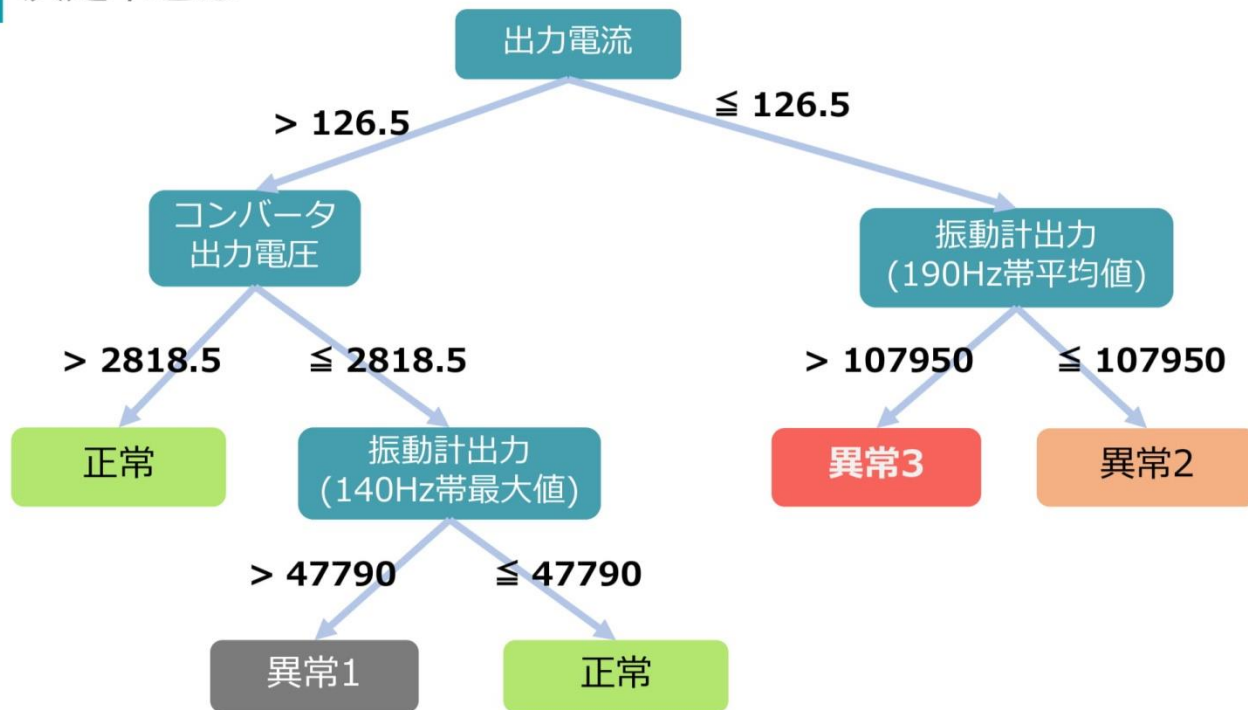
- 線形回帰
- K近傍法
- 決定木
- Gradient Boosting Machine
- ロジスティック回帰
- Support Vector Machine
- Neural Network
- Deep Learning etc...

本日はこのアルゴリズム
だけ使います。



決定木について

決定木とは



決定木について

決定木とは

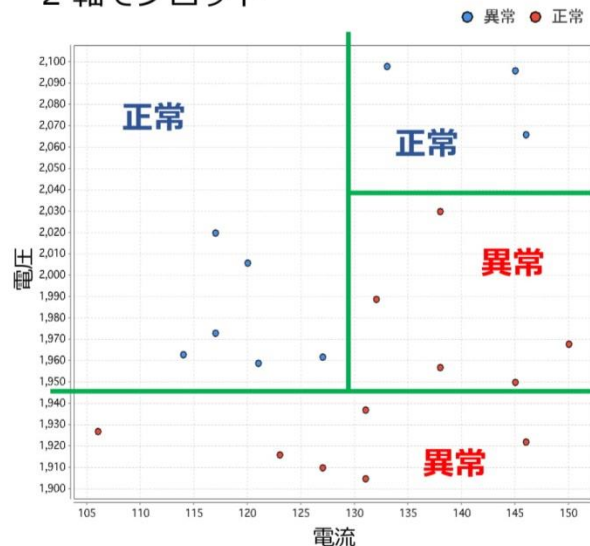
説明変数が 2 つだとすると、

電流	電圧	異常
117	2020	0
138	2030	1
146	2066	0
150	1968	1
117	1973	0
132	1989	1
133	2098	0
106	1927	1
131	1937	1
120	2006	0
127	1962	0
138	1957	1
114	1963	0
145	1950	1
⋮	⋮	⋮

説明変数

目的変数
0: 正常
1: 異常

2 軸でプロット



グラフに水平もしくは垂直の
直線を引く作業に等しい

機械学習デモ＋実習

今回使用するデータ

- 統計学者ロナルド・フィッシャーが測定したアヤメの花のデータセット。3種の花にどのような違いがあるだろうか？その違いから、3種の花を分類することはできるだろうか？



Iris-Versicolor
(バージカラー)



Iris-Setosa
(サトサ)



Iris-Virginica
(バージニカ)

変数A1 : sepal length がく片の長さ(cm)

変数A2 : sepal width がく片の幅(cm)

変数A3 : petal length 花弁の長さ(cm)

変数A4 : petal width 花弁の幅(cm)

設問： 変数A1～A4(説明変数)を用いて、アヤメの種類(目的変数)を分類する

機械学習デモ＋実習

今回準備したデータファイル

Iris_original.csv → オリジナルデータセット
iris_train.csv → モデル学習用データ
Iris_test.csv → 予測用データ

1. データ読み込み

機械学習デモ＋実習

The screenshot displays the RapidMiner Studio interface. The main window shows a process design with two operators: "Read CSV_Train" and "Read CSV_Test". The "Read CSV_Train" operator is connected to the "Read CSV_Test" operator via a data flow line. The "Read CSV_Train" operator has an input port labeled "inp" and an output port labeled "out". The "Read CSV_Test" operator has an input port labeled "inp" and an output port labeled "out". The "Read CSV_Train" operator is connected to the "Read CSV_Test" operator via a data flow line. The "Read CSV_Train" operator is connected to the "Read CSV_Test" operator via a data flow line. The "Read CSV_Train" operator is connected to the "Read CSV_Test" operator via a data flow line.

The interface includes a menu bar (File, Edit, Process, View, Connections, Settings, Extensions, Help) and a toolbar with icons for file operations and execution. A red circle highlights the execution button (a blue play icon) in the toolbar. The "Repository" panel on the left shows a tree view of data sources, including "processes", "results", and "勉強会用" (Study Use), with "第1回" (First Session) selected. The "Operators" panel on the left shows a search for "store" and lists operators under "Data Access" and "Utility". The "Parameters" panel on the right shows settings for the "Process" operator, including "logverbosity" (init) and "logfile". The "Help" panel on the right shows the "Process" operator's synopsis and description.

機械学習デモ＋実習

1. データ読み込み

Local Repository/勉強会用/第1回/processes/1_iris_read_Train_Test_data - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators... etc All Studio

Result History ExampleSet (Read CSV_Test) ExampleSet (Read CSV_Train)

Open in Turbo Prep Auto Model Filter (45 / 45 examples): all

Row No.	id	a1	a2	a3	a4	label
1	id_2	4.900	3	1.400	0.200	Iris-setosa
2	id_6	5.400	3.900	1.700	0.400	Iris-setosa
3	id_7	4.600	3.400	1.400	0.300	Iris-setosa
4	id_8	5	3.400	1.500	0.200	Iris-setosa
5	id_12	4.800	3.400	1.600	0.200	Iris-setosa
6	id_15	5.800	4	1.200	0.200	Iris-setosa
7	id_25	4.800	3.400	1.900	0.200	Iris-setosa
8	id_28	5.200	3.500	1.500	0.200	Iris-setosa
9	id_29	5.200	3.400	1.400	0.200	Iris-setosa
10	id_32	5.400	3.400	1.500	0.400	Iris-setosa
11	id_41	5	3.500	1.300	0.300	Iris-setosa
12	id_43	4.400	3.200	1.300	0.200	Iris-setosa
13	id_45	5.100	3.800	1.900	0.400	Iris-setosa
14	id_49	5.300	3.700	1.500	0.200	Iris-setosa
15	id_50	5	3.300	1.400	0.200	Iris-setosa
16	id_53	6.900	3.100	4.900	1.500	Iris-versicolor
17	id_54	5.500	2.300	4	1.300	Iris-versicolor
18	id_55	6.500	2.800	4.600	1.500	Iris-versicolor

ExampleSet (45 examples, 0 special attributes, 6 regular attributes)

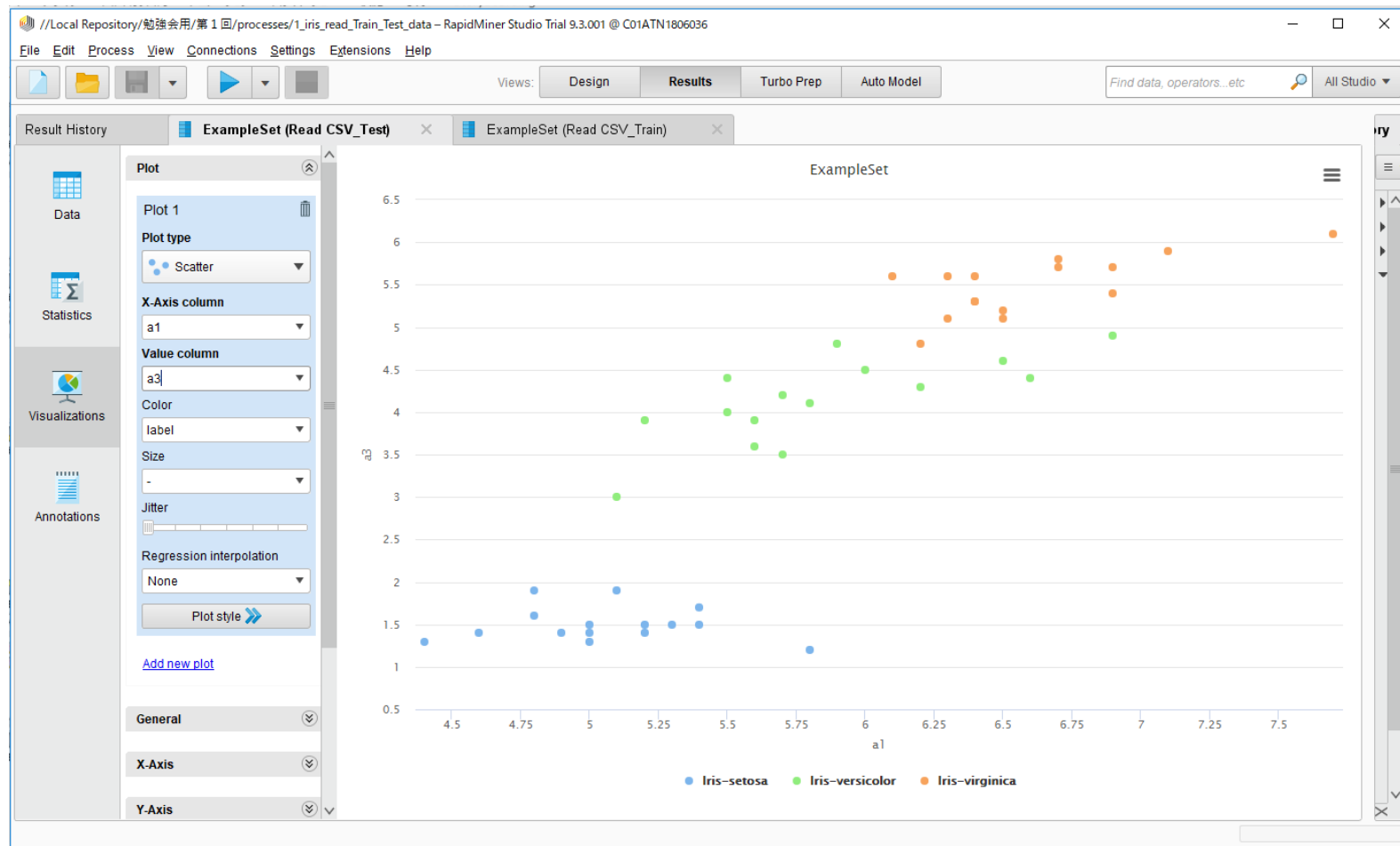
Repository

Import Data

- Samples
- Keras Samples
- DB (Legacy)
- Local Repository (301518)
 - Connections (301518)
 - data (301518)
 - processes (301518)
 - results (301518)
 - 勉強会用 (301518)
 - 第1回 (301518)
 - data (301518)
 - iris_test (301518 - v1, 7/19/19 1:47 PM - 2 kB)
 - iris_train (301518 - v1, 7/19/19 1:47 PM - 5 kB)
 - processes (301518)
 - 0_iris_data_to_CSV (301518 - v1, 7/19/19 1:08 PM - 1 kB)
 - 0_iris_read_CSV_split (301518 - v1, 7/19/19 1:11 PM - 1 kB)
 - 1_iris_read_Train_Test_data (301518 - v1, 7/19/19 1:47 PM - 1 kB)
 - 2_iris_change_set_roles (301518 - v1, 7/19/19 1:47 PM - 1 kB)
 - 3_iris_model (301518 - v1, 7/19/19 1:49 PM - 1 kB)
 - 4_iris_applymodel_performance (301518 - v1, 7/19/19 1:49 PM - 1 kB)
 - results (301518)
 - 第2回 (301518)
 - 第3回 (301518)
 - 第4回 (301518)
 - 第5回 (301518)

機械学習デモ＋実習

1. データ読み込み



2. Set Roles

機械学習デモ＋実習

//Local Repository/勉強会用/第1回/processes/2_iris_change_set_roles - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Repository

- Import Data
- 勉強会用 (301518)
 - 第1回 (301518)
 - data (301518)
 - processes (301518)
 - 0_iris_data_to_CSV
 - 0_iris_read_CSV_s
 - 1_iris_read_Train_
 - 2_iris_change_set
 - 3_iris_model (30151)
 - 4_iris_applymodel_

Process

Process

100%

inp

Read CSV_Train

Set Role

Store

Read CSV_Test

Set Role (2)

Store (2)

res

res

res

Parameters

Store (2) (Store)

repository... test

Show advanced parameters

Operators

store

- Data Access (1)
 - Store
- Utility (1)
 - Process Control (1)
 - Remember

Get more operators from the Marketplace

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Help

Store

RapidMiner Studio Core

Tags: Save, Export, Write, Datasets, Repository, Data Access

Synopsis

This operator stores an IO Object in the data repository

2. Set Roles

機械学習デモ＋実習

Local Repository/勉強会用/第1回/processes/2_iris_change_set_roles - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History ExampleSet (Set Role (2)) ExampleSet (Set Role)

Open in Turbo Prep Auto Model

Filter (45 / 45 examples): all

Row No.	id	label	a1	a2	a3	a4
1	id_2	Iris-setosa	4.900	3	1.400	0.200
2	id_6	Iris-setosa	5.400	3.900	1.700	0.400
3	id_7	Iris-setosa	4.600	3.400	1.400	0.300
4	id_8	Iris-setosa	5	3.400	1.500	0.200
5	id_12	Iris-setosa	4.800	3.400	1.600	0.200
6	id_15	Iris-setosa	5.800	4	1.200	0.200
7	id_25	Iris-setosa	4.800	3.400	1.900	0.200
8	id_28	Iris-setosa	5.200	3.500	1.500	0.200
9	id_29	Iris-setosa	5.200	3.400	1.400	0.200
10	id_32	Iris-setosa	5.400	3.400	1.500	0.400
11	id_41	Iris-setosa	5	3.500	1.300	0.300
12	id_43	Iris-setosa	4.400	3.200	1.300	0.200
13	id_45	Iris-setosa	5.100	3.800	1.900	0.400
14	id_49	Iris-setosa	5.300	3.700	1.500	0.200
15	id_50	Iris-setosa	5	3.300	1.400	0.200
16	id_53	Iris-versicolor	6.900	3.100	4.900	1.500
17	id_54	Iris-versicolor	5.500	2.300	4	1.300
18	id_55	Iris-versicolor	6.500	2.800	4.600	1.500

ExampleSet (45 examples, 2 special attributes, 4 regular attributes)

3. モデル学習

機械学習デモ＋実習

The screenshot displays the RapidMiner Studio interface with a workflow designed for training and testing a Decision Tree model on the Iris dataset.

Repository: The left sidebar shows the project structure. The **iris_train** dataset is highlighted with a red box, indicating it is the selected training data.

Process: The central workspace shows a workflow with two main operators:

- Retrieve iris_train**: The first operator in the process, which loads the training data.
- Decision Tree**: The second operator, which builds a decision tree model from the training data. It is highlighted with a red box.

The workflow is connected by a line, indicating the flow of data from the training data retrieval to the model building process.

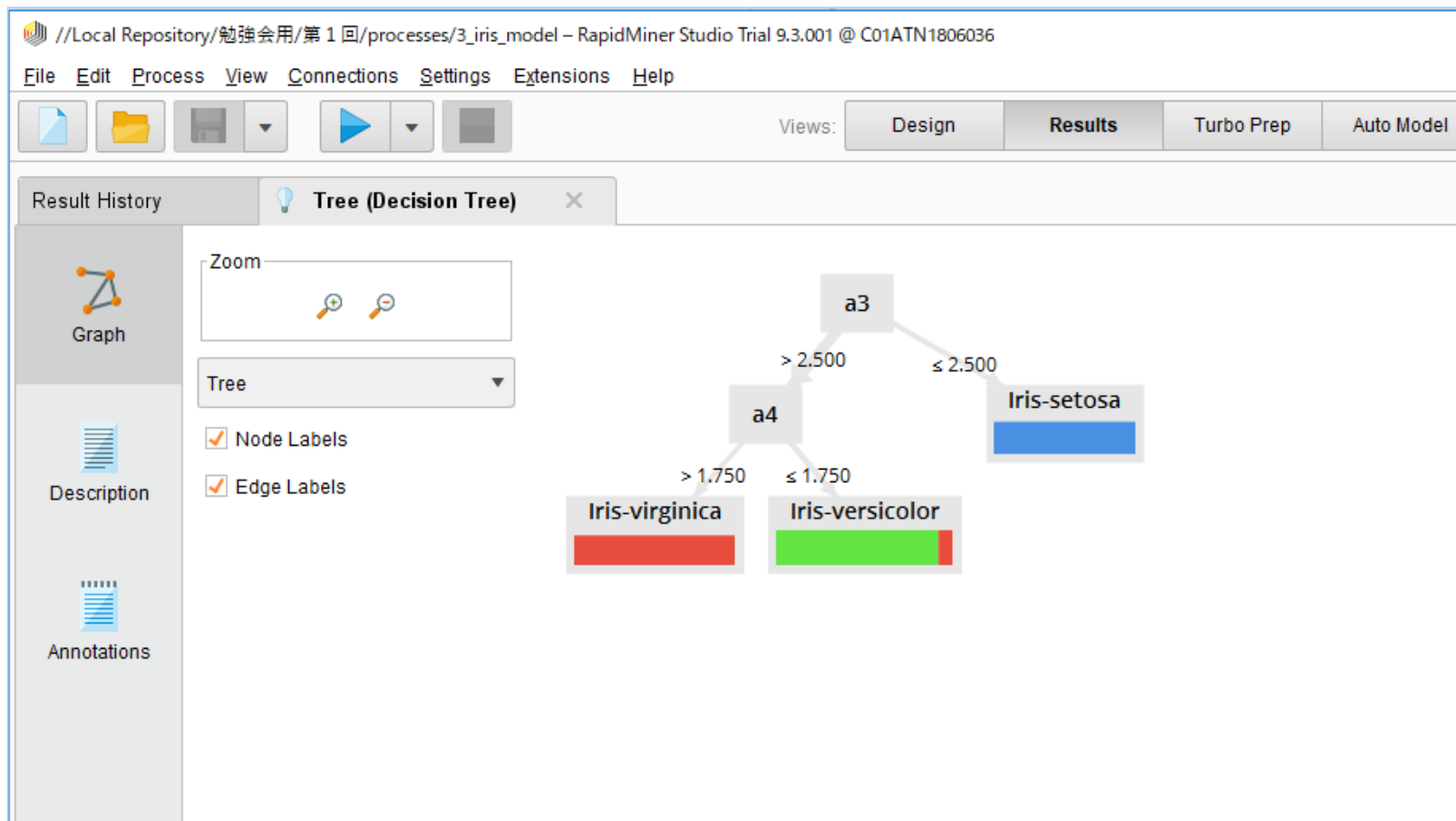
Parameters: The right sidebar shows the parameters for the selected operator (Decision Tree). The **logverbosity** is set to **init**, and the **logfile** is set to an empty field.

Operators: The bottom left sidebar shows the available operators. The **Store** operator is highlighted, suggesting it might be used to save the trained model.

Help: The bottom right sidebar shows the help documentation for the **Process** operator, describing it as the root operator which the outer most operator of every process.

3. モデル学習

機械学習デモ＋実習



4. 分類予測

機械学習デモ＋実習

///Local Repository/勉強会用/第1回/processes/4_iris_applymodel_performance* - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators... etc All Studio

Repository

- Import Data
- iris_test (301518 - v)
- iris_train (301518 - v)
- processes (301518)
 - 0_iris_data_to_CSV
 - 0_iris_read_CSV_s
 - 1_iris_read_Train
 - 2_iris_change_set
 - 3_iris_model (301518)
 - 4_iris_applymodel
- results (301518)

Operators

- store
- Data Access (1)
 - Store
- Utility (1)
 - Process Control (1)
 - Remember

[Get more operators from the Marketplace](#)

Process

Process

100%

Retrieve iris_train

Decision Tree

Retrieve iris_test

Apply Model

Performance

Store

Store (2)

Store (3)

Parameters

Process

logverbosity init

logfile .log

[Show advanced parameters](#)

[Change compatibility \(9.3.001\)](#)

Help

Process

RapidMiner Studio Core

Synopsis

The root operator which the outer most operator of every process.

Description

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

機械学習デモ＋実習

4. 分類予測

人 デザイン 画面切り替え アニメーション スライドショー 校閲 表示 Storyboarding

///Local Repository/勉強会用/第1回/processes/4_iris_applymodel_performance - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators...etc

Result History ExampleSet (Apply Model) PerformanceVector (Performance) Tree (Decision Tree)

Criterion
accuracy

Table View Plot View

accuracy: 93.33%

	true Iris-setosa	true Iris-versicolor	true Iris-virginica	class precision
pred. Iris-setosa	15	0	0	100.00%
pred. Iris-versicolor	0	14	2	87.50%
pred. Iris-virginica	0	1	13	92.86%
class recall	100.00%	93.33%	86.67%	

Performance

Description

Annotations

4. 分類予測

機械学習デモ＋実習

//Local Repository/勉強会用/第1回/processes/4_iris_applymodel_performance - RapidMiner Studio Trial 9.3.001 @ C01ATN1806036

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators...etc All Studio

Result History ExampleSet (Apply Model) PerformanceVector (Performance) Tree (Decision Tree)

Open in Turbo Prep Auto Model Filter (45 / 45 examples): all

Row No.	id	label	prediction(la...	confidence(l...	confidence(l...	confidence(l...	a1	a2	a3	a4
14	id_49	Iris-setosa	Iris-setosa	1	0	0	5.300	3.700	1.500	0.200
15	id_50	Iris-setosa	Iris-setosa	1	0	0	5	3.300	1.400	0.200
16	id_53	Iris-versicolor	Iris-versicolor	0	0.921	0.079	6.900	3.100	4.900	1.500
17	id_54	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.500	2.300	4	1.300
18	id_55	Iris-versicolor	Iris-versicolor	0	0.921	0.079	6.500	2.800	4.600	1.500
19	id_60	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.200	2.700	3.900	1.400
20	id_65	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.600	2.900	3.600	1.300
21	id_68	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.800	2.700	4.100	1
22	id_70	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.600	2.500	3.900	1.100
23	id_71	Iris-versicolor	Iris-virginica	0	0	1	5.900	3.200	4.800	1.800
24	id_76	Iris-versicolor	Iris-versicolor	0	0.921	0.079	6.600	3	4.400	1.400
25	id_79	Iris-versicolor	Iris-versicolor	0	0.921	0.079	6	2.900	4.500	1.500
26	id_80	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.700	2.600	3.500	1
27	id_91	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.500	2.600	4.400	1.200
28	id_96	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.700	3	4.200	1.200
29	id_98	Iris-versicolor	Iris-versicolor	0	0.921	0.079	6.200	2.900	4.300	1.300
30	id_99	Iris-versicolor	Iris-versicolor	0	0.921	0.079	5.100	2.500	3	1.100
31	id_103	Iris-virginica	Iris-virginica	0	0	1	7.100	3	5.900	2.100

ExampleSet (45 examples, 6 special attributes, 4 regular attributes)

(参考情報) 決定木は今最も注目されているアルゴリズムの原型

今最も注目されているアルゴリズムに「LightGBM = light gradient boosting machine」といって、Microsoftがスポンサーになって開発されているアルゴリズムがあります。

Kaggleという機械学習の腕を競うサイトのコンペで、このアルゴリズムは上位入賞者の約半数に使われています。人工知能＝ディープラーニングと思っている人は、人工知能はブラックボックスなので気持ちが悪いか、使いたくないとか言う人が多いのですが、Tree系アルゴリズム(決定木を基にしたアルゴリズム)は数学的に明快で、決してブラックボックスではありません。(しかし、性能を上げるために複雑化していて、決定木ほど視覚的にシンプルに表示することは出来ませんが...))

