

応用プログラミング 3

第 4 回 GUI によるデータ分析

ハンズオン用手順書

専修大学ネットワーク情報学部

田中健太

2023-10-17

はじめに

今回は、GUI (Radiant) の操作が主になるので、ハンズオン操作については、手順書として示します。

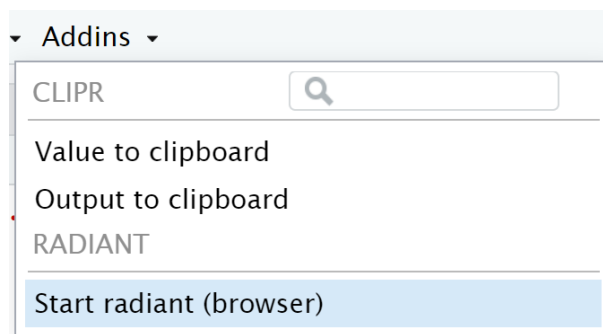
Radiant の起動

今回使用する RStudio Cloud のプロジェクトには、すでに Radiant (radiant パッケージ) をインストールしています。Radiant の起動は、コンソールまたは RStudio のプラグインから行います。

コンソールから起動する場合、以下のように入力します。

```
radiant::radiant_window()
```

また、RStudio には Radiant を起動するためのプラグインが登録されているので、以下の図のようにボタンをクリックします。



すると、Radiant のウィンドウが表示されます。

The screenshot shows the Radiant web interface in a browser window. The address bar shows the URL 127.0.0.1. The interface has a top navigation bar with 'Radiant' and 'Data' tabs. The 'Data' tab is active, showing a sidebar with 'Datasets:' (containing 'diamonds'), 'Display:' (with 'preview' selected), and 'Load data of type:' (with 'rds | rda | rdata' selected). The main area displays a 'Data preview' table with columns: price, carat, clarity, cut, color, depth, table, x, y, z, and date. The table contains 10 rows of data.

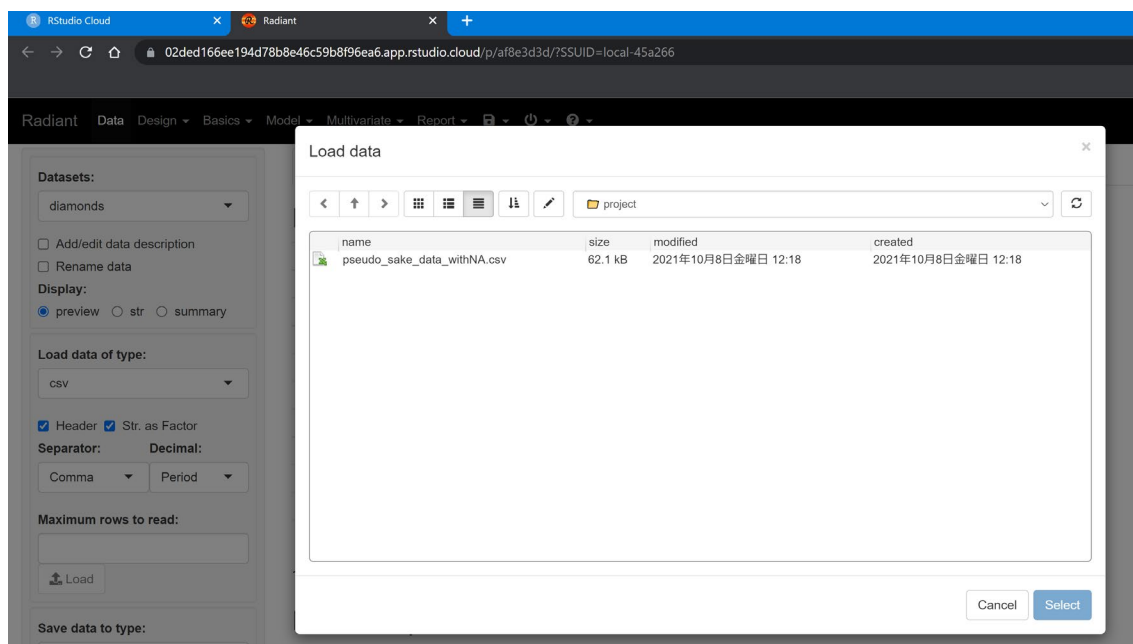
price	carat	clarity	cut	color	depth	table	x	y	z	date
580	0.32	VS1	Ideal	H	61.00	56.00	4.43	4.45	2.71	2012-02-26
650	0.34	SI1	Very Good	G	63.40	57.00	4.45	4.42	2.81	2012-02-26
630	0.30	VS2	Very Good	G	63.10	58.00	4.27	4.23	2.68	2012-02-26
706	0.35	VVS2	Ideal	H	59.20	56.00	4.60	4.65	2.74	2012-02-26
1080	0.40	VS2	Premium	F	62.60	58.00	4.72	4.68	2.94	2012-02-26
3082	0.60	VVS1	Ideal	E	62.50	53.70	5.35	5.43	3.38	2012-02-26
3328	0.88	SI1	Ideal	I	61.70	56.00	6.14	6.18	3.80	2012-02-26

データの読み込み

今回は、RStudio Cloud 上に配置している `pseudo_sake_data_withNA.csv` を読み込みます。Radiant の画面左側の "Load data of type:" メニューのドロップダウンリストから "csv" を選択します。

The screenshot shows the 'Load data of type:' form. It has a dropdown menu set to 'csv'. Below it are two checked checkboxes: 'Header' and 'Str. as Factor'. There are two more dropdown menus: 'Separator:' set to 'Comma' and 'Decimal:' set to 'Period'. At the bottom is a text input field for 'Maximum rows to read:' and a 'Load' button with an upload icon.

そして、"Load" ボタンを押します。すると、ファイル選択ダイアログが開きます。



pseudo_sake_data_withNA.csv を選択し、"Select" ボタンを押します。

すると、ファイルが読み込まれ、データの一部が表示されます。

Datasets:

pseudo_sake_data_withNA

☐ Add/edit data description
☐ Rename data

Display:

☒ preview ☐ str ☐ summary

Load data of type:

csv

☒ Header ☒ Str. as Factor

Separator: Comma **Decimal:** Period

Maximum rows to read:

Manage View Visualize Pivot Explore Transform Combine

Data preview

アルコール分	日本酒度	エキス分	総酸	アミノ酸度	甘辛度	濃淡度	精米歩合	容器	タイプ
15.90	5.16	4.88	1.13	1.05	-0.13	-1.13	73.00	茶色瓶	一般酒
16.30	3.63	5.22	1.03	1.60	0.13	-0.92	71.00	茶色瓶	一般酒
15.40	-1.09	4.80	1.39	1.29	0.11	-1.01	68.00	茶色瓶	一般酒
15.40	-1.38	4.68	1.49	1.04	-0.10	-1.00	69.00	茶色瓶	一般酒
14.10	3.14	2.82	1.08	0.97	-0.13		69.00	茶色瓶	一般酒
15.40	1.79	4.01	1.32	1.89	0.08	-0.91	70.00	茶色瓶	一般酒
14.20	6.50	4.30		0.25		-1.04	73.00	茶色瓶	一般酒
14.20	3.17	5.38	1.33	1.51	0.06	-0.91	69.00	緑色瓶	一般酒
15.00	2.78	4.72	1.26	1.40	-0.04	-1.04	68.00	茶色瓶	一般酒
16.40	5.92	4.91	0.81	0.79	-0.47	-1.07	69.00	茶色瓶	一般酒

10 of 998 rows shown. See View-tab for details.

レポートへの反映

Radiant では、GUI での操作結果を R Markdown によるレポートとして保存、出力できます。さまざまな画面で操作を行い、意図した結果が得られるたびに、左側のメニュー項目下端にあるアイコンを押し、R Markdown 形式で出力する必要があります。

Save data to type:


rds ▼

Save

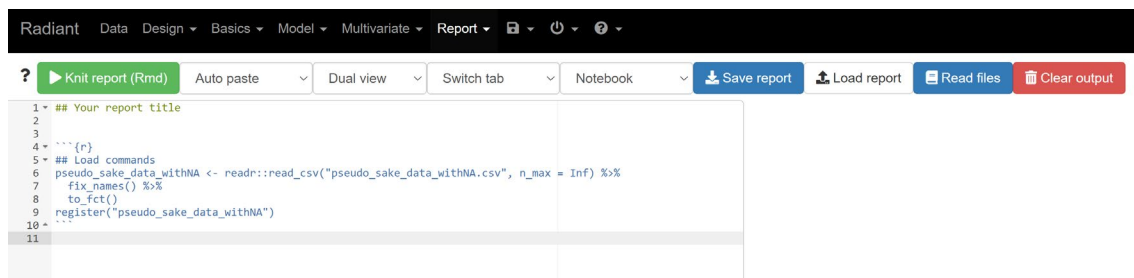
☐ Show R-code

☐ Remove data from memory

?



ボタンを押すと、"Report" メニューが表示されます。

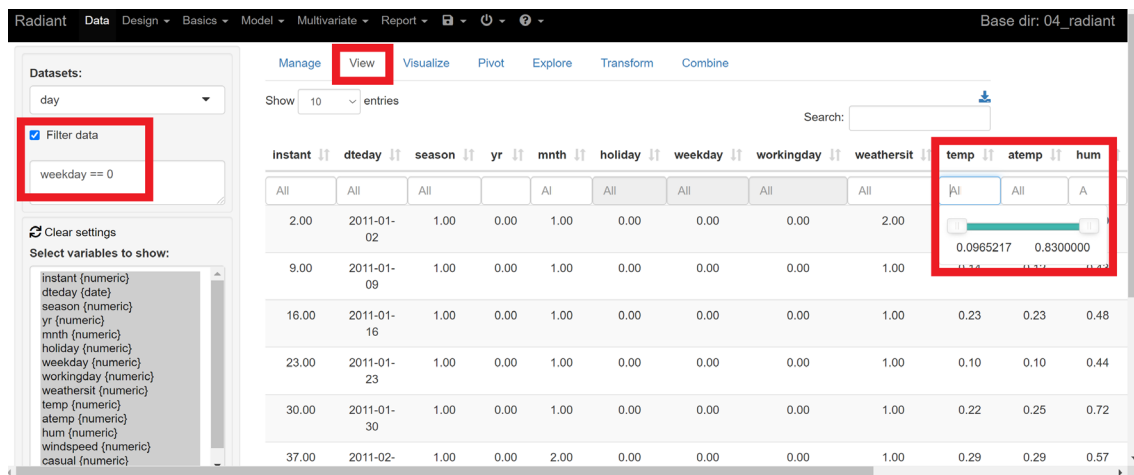


とりあえずこの段階では、GUI の操作が反映されたことを確認したら、ふたたび "Data" メニューをクリックし、元の画面に戻ります。

データの観察

読み込んだデータをさまざまに観察してみましょう。はじめに、"Data" - "View" をクリックします。

すると、データとフィルターや並べ替えのためのメニューが表示されます。



データの抽出

👉 やってみよう

フィルターを使い、(1) アルコール度数が 17%以上、(2) 容器が緑色瓶および青色瓶であるデータを抽出してください。

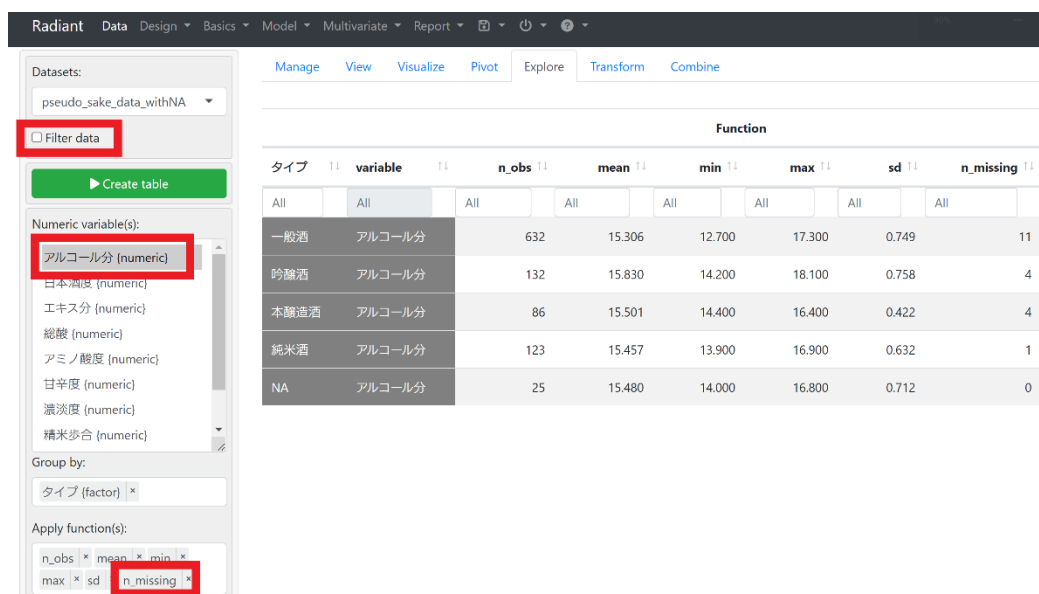
10 件のデータが抽出されるはずです。

レポートへの反映

先ほどと同様、メニュー下端のアイコンをクリックし、R Markdown として出力してください。

データの集計

次に、"Data" - "Explore" を選択してください。すると、対象列やグループ化変数を指定し、任意の基本統計量などを算出できるメニューが表示されます。まず、はじめに "Filter data" のチェックを外し、データすべてを対象にしてください。



👉 やってみよう

1. 左側の "Numeric variables (s)" で アルコール分 を選択し、"Group by" に タイプ を指定してください
2. "Apply function (s)" に欠損値の数をカウントする `n_missing` を追加してください
3. "Create table" ボタンを押してみましょう

基礎集計が簡単にできます。

レポートへの反映

先ほどと同様、メニュー下端のアイコンをクリックし、R Markdown として出力してください。

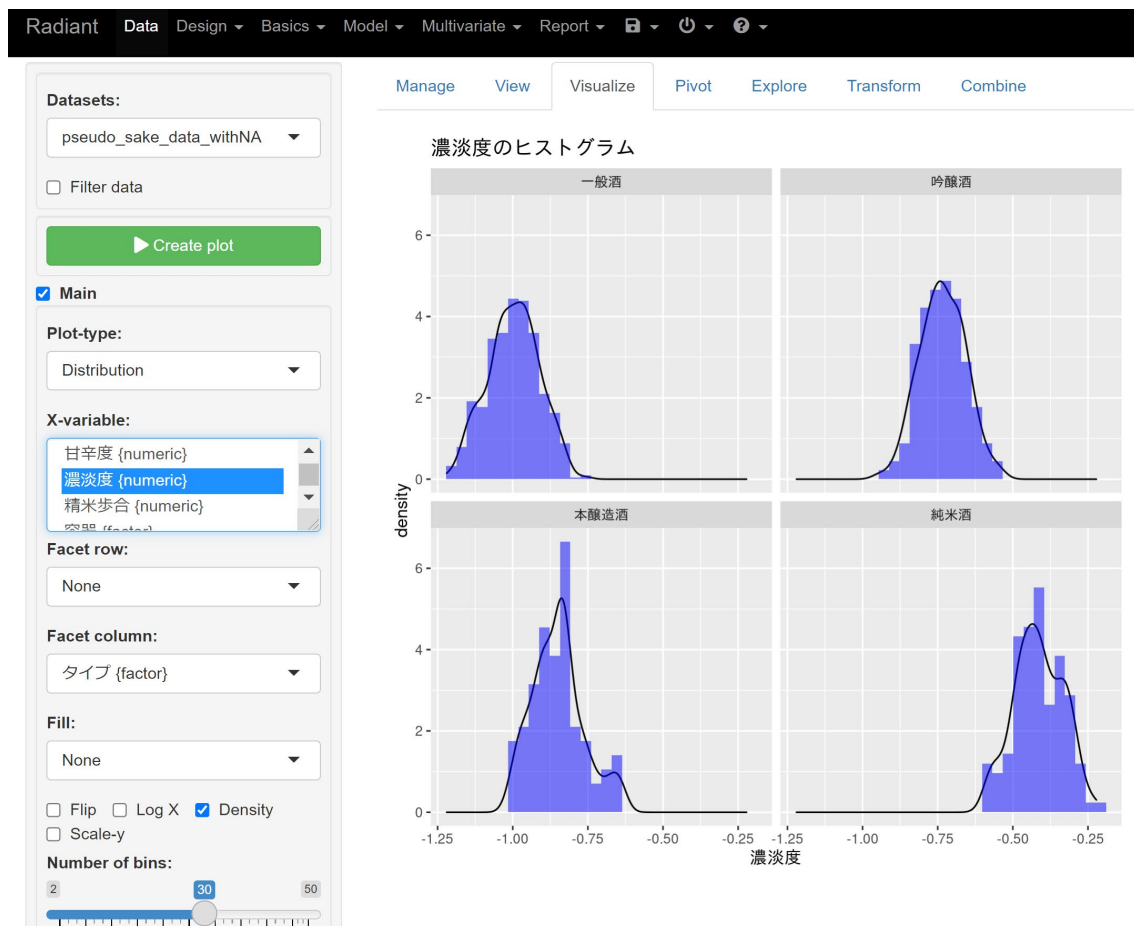
データの可視化

次に、"Data" - "Visualize" を選択してください。さまざまなグラフィックスで可視化できます。

👉 やってみよう

1. "X-variable" に 濃淡度 を指定してください
2. "Facet column" に タイプ を指定してください
3. "Fill" の下のチェックボックスで "Density" にチェックを入れてください
4. "Number of bins" のスライダーを動かし、30 にしてください
5. "Labels" にチェックを入れると設定項目が開くので、"Title" に「濃淡度のヒストグラム」と指定してください
6. 上部の "Create plot" ボタンを押し、グラフを表示してください

すると、以下のような出力が得られます。



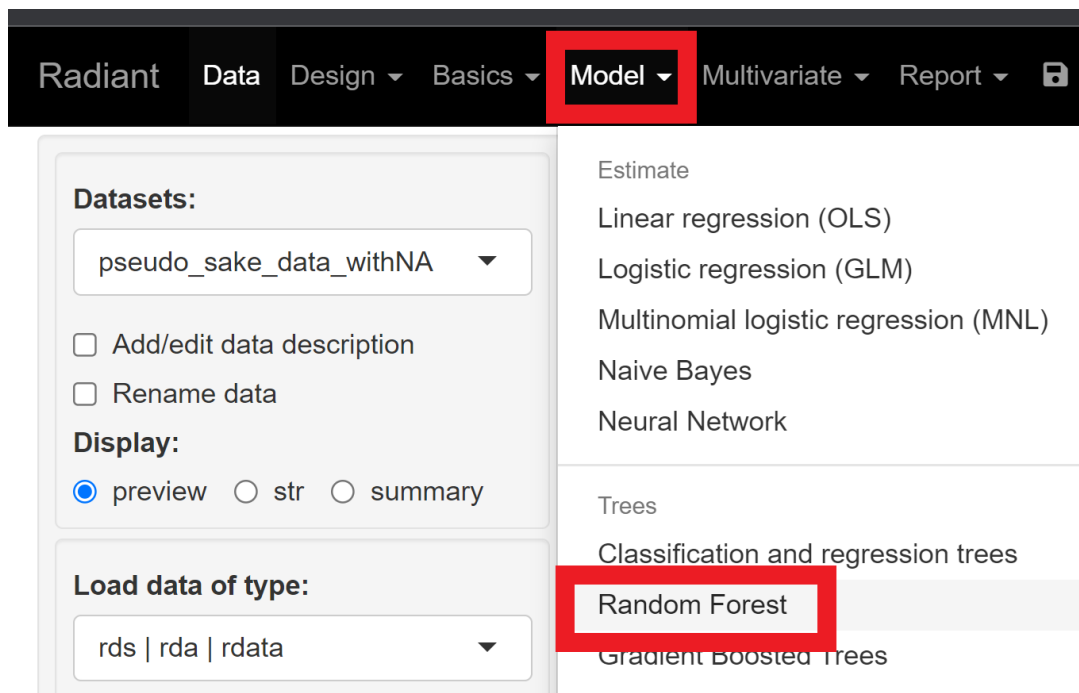
レポートへの反映

先ほどと同様、メニュー下端のアイコンをクリックし、R Markdown として出力してください。

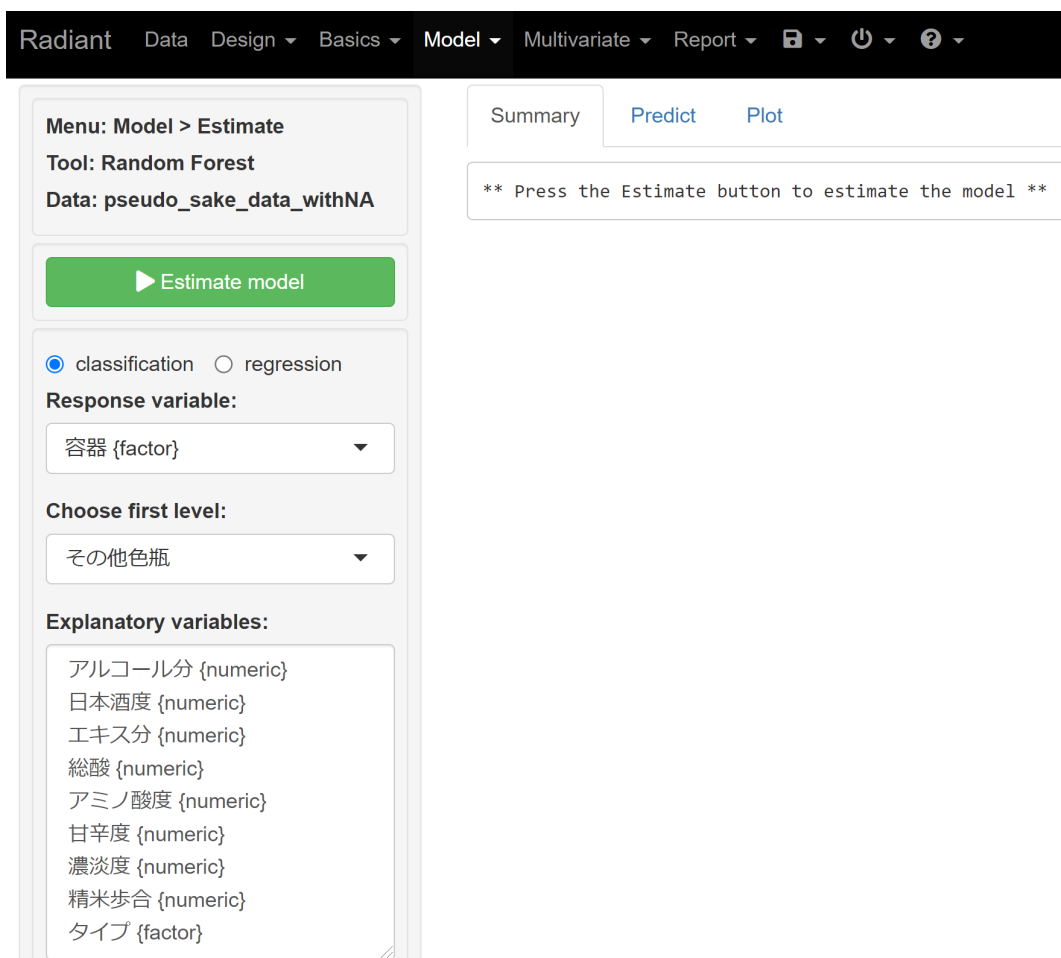
機械学習

Radiant では、機械学習をはじめとした高度なデータ分析も GUI で実行できます。ここでは、機械学習の手法の 1 つであるランダムフォレストを使い、タイプ（一般酒、吟醸酒など）を他の変数で説明するモデルを作成してみましょう。

まず、"Model" - "Random Forest" メニューを選択してください。



すると、ランダムフォレストにおける目的変数（予測したい値）や説明変数（予測に使う値）を指定する画面が表示されます。



モデルの設定

ここから、ランダムフォレストで「タイプ」をその他の変数で説明するモデルを作成していきます。

👉 やってみよう

1. 左側のメニューの "Response variable" について、ドロップダウンリストから タイプ を選択してください
2. "Choose first level" を 吟醸酒 に設定してください
3. "Explanatory variables" について、すべての変数を選択してください
4. "mtry" メニューの値を 4 に変更してください
5. "# trees" メニューの値を 500 に変更してください

RadiantDataDesignBasicsModelMultivariateReport

Menu: Model > Estimate
Tool: Random Forest
Data: pseudo_sake_data_withNA

Estimate model

☒ classification ☐ regression
Response variable:
タイプ {factor}

Choose first level:
吟醸酒

Explanatory variables:

アルコール分 {numeric}
日本酒度 {numeric}
エキス分 {numeric}
総酸 {numeric}
アミノ酸度 {numeric}
甘辛度 {numeric}
濃淡度 {numeric}
精米歩合 {numeric}
容器 {factor}

Weights:
None

mtry: 4 # trees: 500

SummaryPredictPlot

** Press the Estimate button to estimate the model **

9

モデルの作成

それではモデルを作成しましょう。

👉 やってみよう

"Estimate model" ボタンを押してください。

すると、すぐにどのような設定でモデルを作成したか、"Summary" が表示されます。

Summary	Predict	Plot
Random Forest (Ranger)		
Type	: Classification	
Data	: pseudo_sake_data_withNA	
Response variable	: タイプ	
Level	: 吟醸酒 in タイプ	
Explanatory variables	: アルコール分, 日本酒度, エキス分, 総酸, アミノ酸度, 甘辛度, 濃淡度, 精米歩合, 容器	
Mtry	: 4	
Number of trees	: 500	
Min node size	: 1	
Sample fraction	: 1	
Number of threads	: 12	
Number of nodes	: 815	
OOB prediction error	: 0.036	

すでにこの時点でモデルが完成しており、OOB prediction error (Out-of-bag 誤り率) が出力されています。OOB 誤り率について詳しくは触れませんが、[英語版 Wikipedia の記事](#)などを参考にしてください。細かい数字は多少変わる場合がありますが、誤り率が約 3.6%と、非常に精度が高いモデルであることがわかります。¹

レポートへの反映

先ほどと同様、メニュー下端のアイコンをクリックし、R Markdown として出力してください。

精度評価

作成したモデルの評価をしましょう。

👉 やってみよう

"plot" メニューを選択してください

¹ このデータは人工データなので、ノイズも少ないため、予測しやすくなっています。

Menu: Model > Estimate
Tool: Random Forest
Data: pseudo_sake_data_withNA

Plots:
None

?

Summary
Predict
Plot

Please select a

Radiant では、3 つのグラフィックスでモデルの精度を確認できますが、分類タスクでは “Permutation Importance” の描画でエラーが出てしまいます。

Menu: Model > Estimate
Tool: Random Forest
Data: pseudo_sake_data_withNA

Plots:
None

?

None
Permutation Importance
Prediction plots
Partial Dependence

Partial Dependence の算出

今回は、Partial Dependence のグラフを描いてみましょう。

👉 やってみよう

"Partial Dependence" を選択してください

Menu: Model > Estimate
 Tool: Random Forest
 Data: pseudo_sake_data_withNA

Plots:
 Partial Dependence

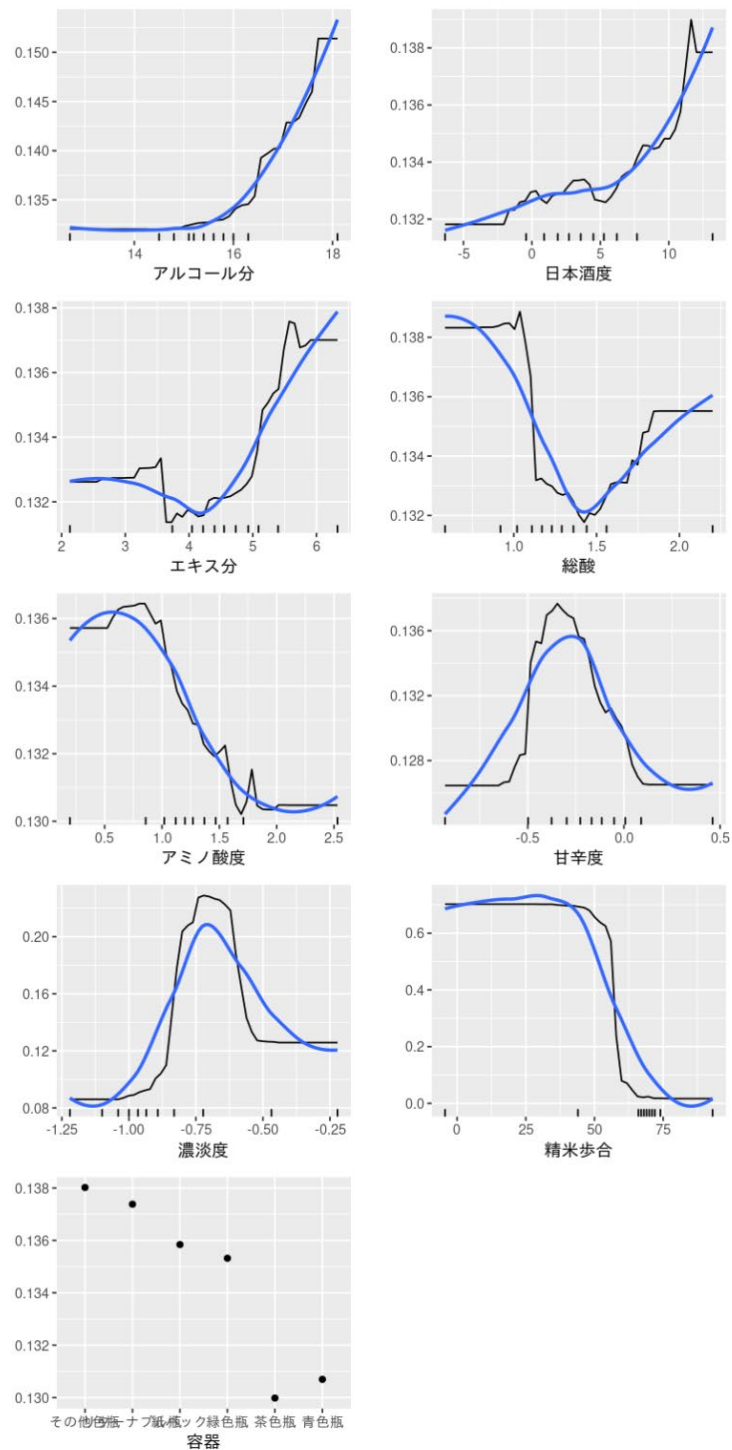
☐ Show quintiles

?

Summary

Predict

Plot



Partial dependence についても詳しくは触れませんが、[5.1 Partial Dependence Plot \(PDP\) | Interpretable Machine Learning](#) に、日本語訳された解説があります。こちらは、それぞれの説明変数が変化すると、ある特定のラベル（ここでは指定した 吟醸酒）になる確率がどのように変化するかをあらわす指標です。

レポートへの反映

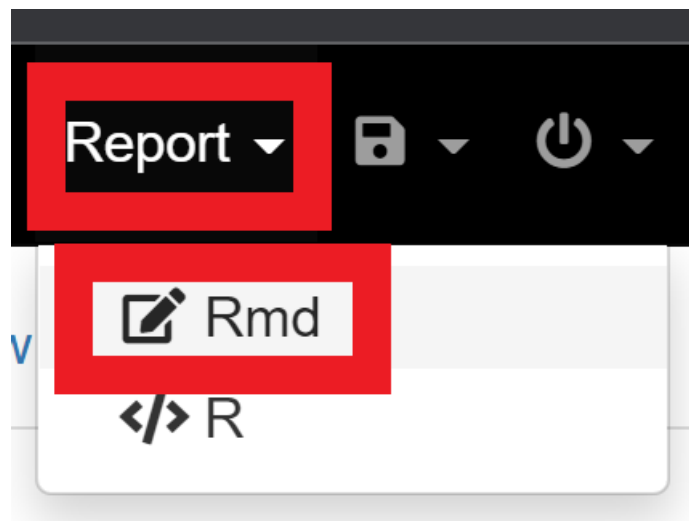
先ほどと同様、メニュー下端のアイコンをクリックし、R Markdown として出力してください。

まとめとレポートの出力

ここまで、Radiant を使い、GUI でデータ分析ができることを体験しました。もちろん、Radiant の機能はもっと豊富なので、ぜひ活用してみてください。

レポートの出力

さて、今週の課題は、ハンズオンとして行った操作を R Markdown のレポートとして記録し、HTML に変換して出力するものです。ここまで、何か操作を行うたびに「レポートへの反映」という指示がありましたが、最終的な結果は "Report" メニューから確認できます。



```
Radiant Data Design Basics Model Multivariate Report ? Knit report (Rmd) Auto paste Dual view Switch tab Notebook Save report Load report Read files Clear output

1 # Your report title
2
3
4 ```{r}
5 ## Load commands
6 pseudo_sake_data_withNA <- readr::read_csv("pseudo_sake_data_withNA.csv", n_max = Inf) %>%
7   fix_names() %>%
8   to_fct()
9 register("pseudo_sake_data_withNA")
10
11
12
13 ```{r}
14 ## filter and sort the dataset
15 pseudo_sake_data_withNA %>%
16   filter(アルコール分 >= 17.0 & アルコール分 <= 18.1 & 容器 %in% c('緑色瓶','青色瓶')) %>%
17   select(アルコール分:タイプ) %>%
18   dtab(dec = 2, nr = 100) %>% render()
19
20
21
22 ```{r}
23 result <- explore(
24   pseudo_sake_data_withNA,
25   vars = "アルコール分",
26   byvar = "タイプ",
27   fun = c("n_obs", "mean", "min", "max", "sd", "n_missing"),
28   nr = Inf
29 )
30 summary(result)
31 # dtab(result) %>% render()
32
33
34
35 ```{r fig.width = 7, fig.height = 7, dpi = 96}
36 visualize(
37   pseudo_sake_data_withNA,
38   xvar = "濃淡度",
39   type = "dist",
40   facet_col = "タイプ",
41   bins = 30,
42   axes = "density",
43   labs = list(title = "濃淡度のヒストグラム"),
44   custom = FALSE

```

ここまで GUI で行った作業がコードとして反映されていることを確認したら、レポートを出力しましょう。

👉 やってみよう

"Knit report (Rmd)" ボタンを押してください

ランダムフォレストの Partial dependence のグラフィックス作成に時間がかかりますが、そのうち右側のペインにレポートが出力されます。

```
1 # Your report title
2
3
4 ```{r}
5 ## Load commands
6 pseudo_sake_data_withNA <- readr::read_csv("pseudo_sake_data_withNA.csv", n_max = Inf) %>%
7   fix_names() %>%
8   to_fct()
9 register("pseudo_sake_data_withNA")
10
11
12
13 ```{r}
14 ## filter and sort the dataset
15 pseudo_sake_data_withNA %>%
16   filter(アルコール分 >= 17.0 & アルコール分 <= 18.1 & 容器 %in% c('緑色瓶','青色瓶')) %>%
17   select(アルコール分:タイプ) %>%
18   dtab(dec = 2, nr = 100) %>% render()
19
20
21
22 ```{r}
23 result <- explore(
24   pseudo_sake_data_withNA,
25   vars = "アルコール分",
26   byvar = "タイプ",
27   fun = c("n_obs", "mean", "min", "max", "sd", "n_missing"),
28   nr = Inf
29 )
30 summary(result)
31 # dtab(result) %>% render()
32
33
34
35 ```{r fig.width = 7, fig.height = 7, dpi = 96}
36 visualize(
37   pseudo_sake_data_withNA,
38   xvar = "濃淡度",
39   type = "dist",
40   facet_col = "タイプ",
41   bins = 30,
42   axes = "density",
43   labs = list(title = "濃淡度のヒストグラム"),
44   custom = FALSE

```

Project: proj

Your report title

アルコール分 日本酒度 エキス分 総酸度 アミノ酸度 甘辛度 濃淡度 精米歩合 容積率 タイプ

Explore

Data	pseudo_sake_data_withNA						
Grouped by	タイプ						
Functions	n_obs, mean, min, max, sd, n_missing						
Top	Function						
タイプ	variable	n_obs	mean	min	max	sd	n_missing
一般酒	アルコール分	632	15.386	12.780	17.380	0.749	11
吟醸酒	アルコール分	132	15.830	14.200	18.100	0.758	4
本醸造酒	アルコール分	86	15.501	14.400	16.400	0.422	4
純米酒	アルコール分	123	15.457	13.900	16.900	0.632	1
<NA>	アルコール分	25	15.480	14.000	16.800	0.712	0

濃淡度のヒストグラム

一般酒

吟醸酒

タイトル、学籍番号、氏名の記入

提出にあたって、これまでのレポート同様、タイトル、学籍番号、氏名を記入します。

現在、左側のエディタペインで ## Your report title と記述されている部分について、以下のように編集してください。

👉 やってみよう

応用プログラミング3第4回 課題レポート

* 学籍番号: xxxx (ご自身の番号)

* 氏名: xxxx (ご自身の名前)

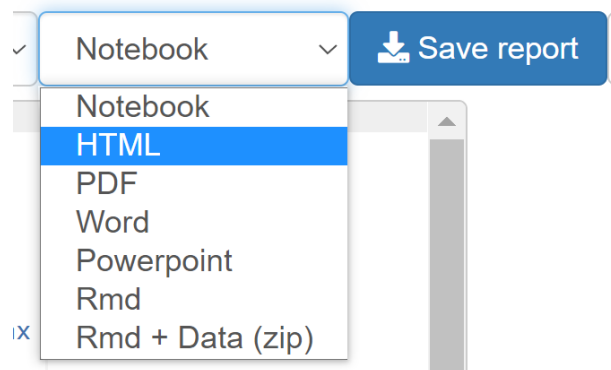
記入したら、ふたたび "Knit report (Rmd)" ボタンを押してください。

レポートの HTML 形式での保存

出力を確認したら、レポートを HTML 形式で保存します。

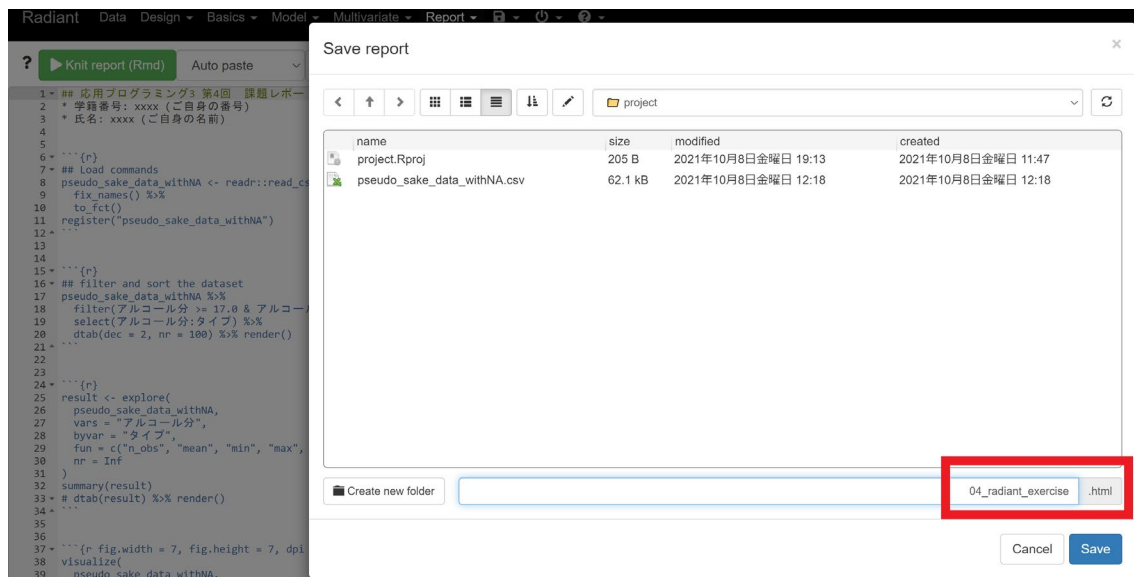
👉 やってみよう

1. 画面中央の "Notebook" と表示されているドロップダウンリストから、"HTML" を選択してください
2. 隣の "Save report" ボタンを押してください



👉 やってみよう

表示されるダイアログでファイル名に "04_radiant_exercise.html" と指定し、"Save" ボタンを押してください



Posit Cloud からのレポートのダウンロード

最後に、これまでの課題と同様、RStudio Cloud から出力結果の 04_radiant_exercise.html をダウンロードし、Classroom に提出してください。

Radiant の終了

作業が完了したら、Radiant の "電源ボタン" メニューから "Stop" を選択し、終了してください。Radiant のタブが閉じ、RStudio Cloud のコンソールにフォーカスが戻ります。

