

RapidMinerGo

モデル評価画面 用語解説

分類/回帰

2022年6月



はじめに

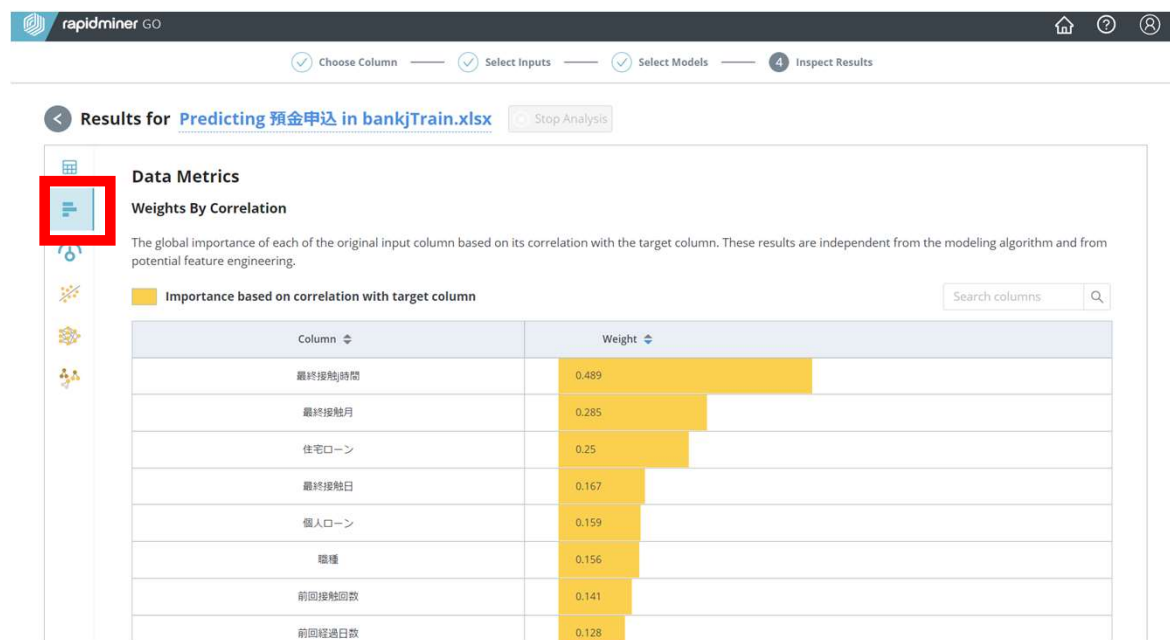
- ❖ 本資料で伝えたい内容
RapidMinerGOで作成したモデルを評価する際に理解しておくべき用語の解説
- ❖ 読み手
 - ・ ビジネスアナリスト(業務分析担当者)、AI活用優先順位付けやビジネスインパクトの見極めをされる方
 - ・ 作成したモデルの評価方法について詳しく知りたい方
- ❖ 以下の用語について解説する。

モデル種類	用語
1.共通	(1)Weights By Correlation
2.分類	(1)モデル比較表 混同行列、正解率、classification error、適合率、再現率、F値 (2)モデル別評価画面 ROC曲線、AUC、column Weights (3)モデルシミュレーター(分類)
3.回帰	(1)モデル比較表 RMSE、MAE、MRE、 R^2 、column Weights (2)モデル別評価画面 Actual vs. Predicted Values、Distribution of prediction errors (3)モデルシミュレーター(回帰)

1.モデル評価（共通）

(1)Weights by Correlation :

- ・ 目的変数との相関が高い項目がランキング形式で確認できる
- ・ 各項目の相関の高さを0.0～1.0の間の値で評価する
- ・ (活用例)顧客情報を使って預金申込するか、しないかの分類モデルを作成した場合
 - ・ ランキングを見ると「最終接触時間」「最終接触月」「住宅ローン」が預金申込に与える影響を与える可能性が高いことが分かる。
 - ・ 「最終接触時間」「最終接触月」「住宅ローン」の値がどのように変化した時に、申込確率が高まるかモデルシミュレータ機能(次ページ紹介)で確認できる



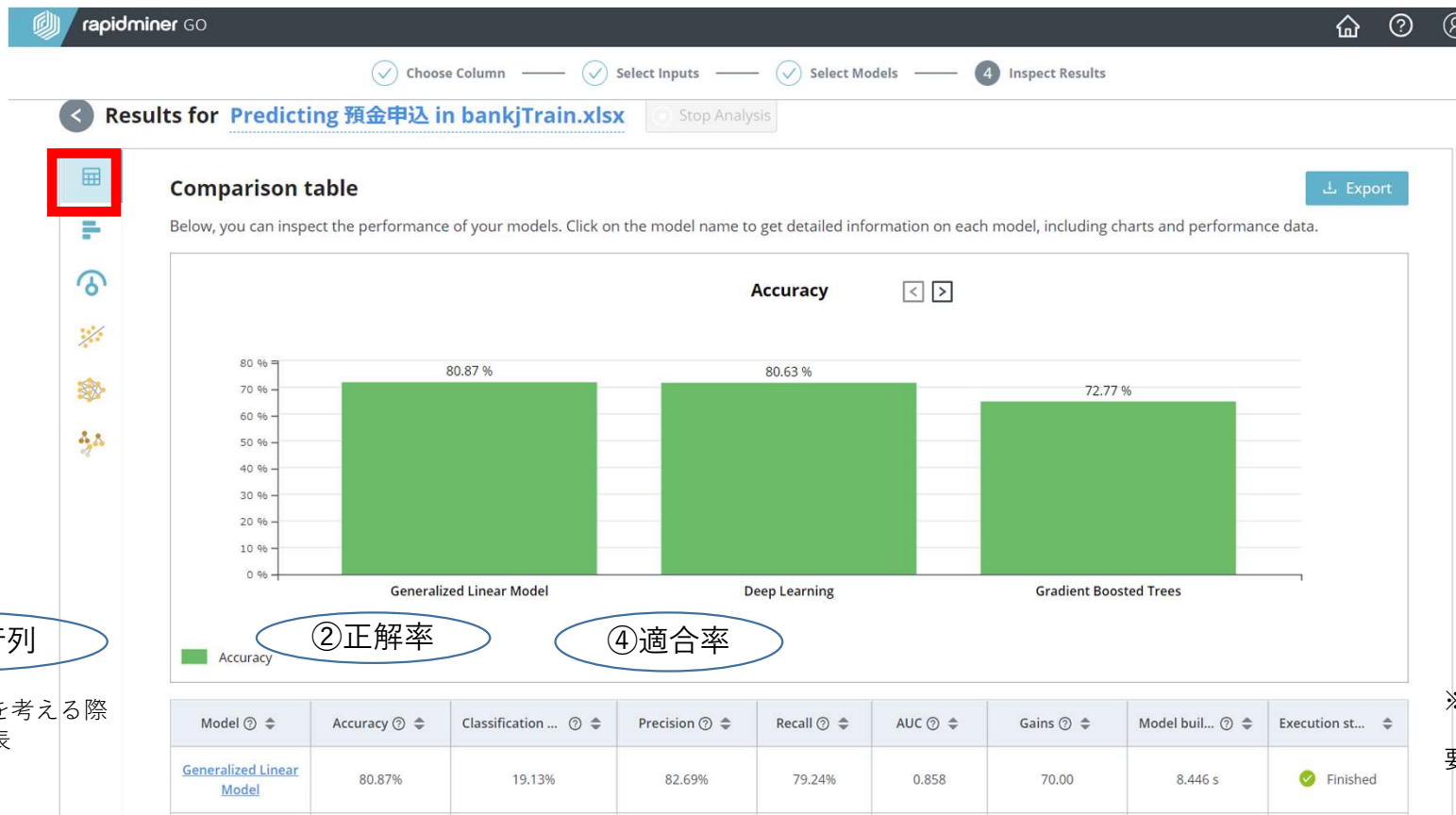
参考文献：②

2.モデル評価 (分類)

Innovation-Leading Company

分類モデル：どのカテゴリ(例：病気が健康か、合格か不合格か など)に属するかを分類するモデルを作成する場合

(1) モデル比較表：各モデルの評価を比較できる画面



①混同行列

※②～⑤評価を考える際の基本となる表

②正解率

④適合率

⑥F値

※④⑤の調和平均をとった指標
GOの画面には表記はないが、重要な指標のため、追加解説

③classification error

⑤再現率

2.モデル評価 (分類)

※2値分類("A"or"B"の二種類に分類する場合)を例に説明

Confusion Matrix

Counts of actual vs. predicted values for column 預金申込 on a test set of 115 rows.

	Actually なし	Actually あり	Class Precision
Predicted なし	● 48	● 12	80.00%
Predicted あり	● 10	● 45	● 81.82%
Class Recall	82.76%	● 78.95%	

※モデル別評価画面の「Model Evaluation」から照会可

① 混同行列

分類モデルの評価を考える際の基本となる行列で、モデルの予測値と実績値の関係を表したもの。
予測結果を正例と負例の2軸で分類し、それぞれに当てはまる予測結果の個数をまとめた表。

- ・ 予測が当たったケース ⇒ TP + TN
- ・ 予測が外れたケース ⇒ FP + FN

	負例 (実績)	正例 (実績)
負例 (予測)	True negative (TN)	False negative (FN)
正例 (予測)	False positive (FP)	True positive (TP)

参考文献：①、④

2.モデル評価 (分類)

(例)癌検診
混同行列が以下の場合の例

	健康 (実績)	癌 (実績)
健康(予測)	3(TN)	1(FN)
癌(予測)	2(FP)	4(TP)

② 正解率(Accuracy)

全ての事象の中で、判定結果が当たっていた数の割合
 $\Rightarrow (TP+TN)/(TP+FN+FP+TN) = 7/10$

③ Classification error

全ての事象の中で、判定結果が外れた数の割合
 $\Rightarrow (FP+FN)/(TP+FN+FP+TN) = 3/10$

④ 適合率(Precision)

予測に対して、実績が当たっていた割合(正例と負例あり)

- ・ (正例) $TP/(TP+FP) = 4/6$
- ・ (負例) $TN/(FN+TN) = 3/4$

⑤ 再現率(Recall)

実績に対して、予測が当たっていた割合(正例と負例あり)

- ・ (正例) $TP/(TP+FN) = 4/5$
- ・ (負例) $TN/(FP+TN) = 3/5$

⑥ F値

適合率と再現率の調和平均、モデルを総合的に評価する
再現率と適合率、どちらか一方に偏らせずに均等に評価できる

$$2 \times \frac{\text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} = 2 \times \frac{\frac{4}{6} \times \frac{4}{5}}{\frac{4}{6} + \frac{4}{5}} = 0.727$$

2.モデル評価 (分類)

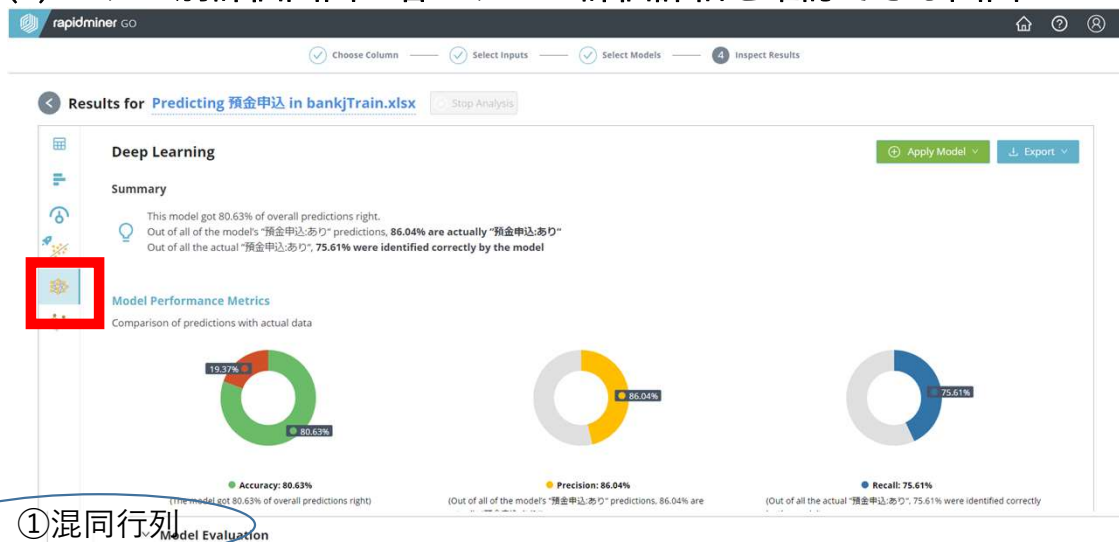
評価指標の活用シーン例

「病気かどうか」という二値分類を予測するモデルを作成した場合

指標	活用シーン
再現率	健康診断の場合、病気である人を見逃したくないため(FNを出したくない)、再現率を重視 ※予測の誤りが多少あっても抜け漏れを少なくしたいという場合に重視される指標
適合率	精密検査の場合、病気でない人を誤診したくないため(FPを出したくない)、適合率を重視 ※見逃しが多少あっても予測可能性が高い対象に対してのみ検知したい場合に重視される指標
F値	一般の外来患者の診察は、“病気を発見すること”、“病気でない人を誤診しないこと”(TPとTNを重視) これら両者のバランスが最もよくなるように診断したいので、F値を重視 ※再現率でも適合率でもなく、これらの両者のバランスが重視される指標

2.モデル評価（分類）

(2) モデル別評価画面：各モデルの評価詳細を確認できる画面



Confusion Matrix

Counts of actual vs. predicted values for column 預金申込 on a test set of 114 rows.

	Actually なし	Actually あり	Class Precision
Predicted なし	49	14	77.78%
Predicted あり	8	43	84.31%
Class Recall	85.96%	75.44%	

ROC Curve

Displays the comparison between True Positives and False Positives for both possible predictions (correct and incorrect).

AUC: 0.846

Sensitivity: 75.61% (True positive rate)

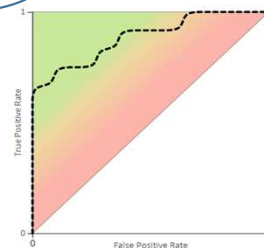
Specificity: 85.91% (False positive rate)

High performance

Low performance

ROC curve

⑧ROC曲線



⑦AUC

Innovation-Leading Company

⑨Column Weights

Column Weights

Influence of columns on the predictions of this specific model

Column	Weight
住宅ローン	0.307
最終接触時間	0.288
年齢	0.108
前回接触回数	0.092
今回接触回数	0.067
最終接触日	0.058
個人ローン	0.057
職業	0.041
結婚歴	0.034
前回経過日数	0.033

2.モデル評価（分類）

⑦ AUC (Area Under the Receiving Operating Curve)

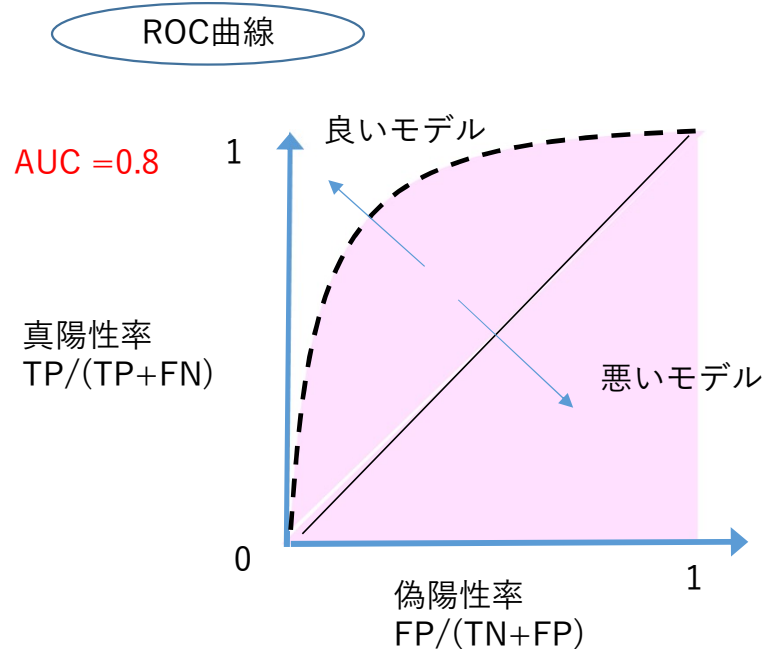
ROC曲線の下領域面積を表す。0～1の値をとるモデルの性能評価指標。

AUC	性能評価
0.8以上	性能が非常に高い 但し1に近すぎると、過学習している可能性があるため要検証
0.7-0.8	性能が良好
0.6-0.7	性能が認められる
0.6未満	性能が認められない

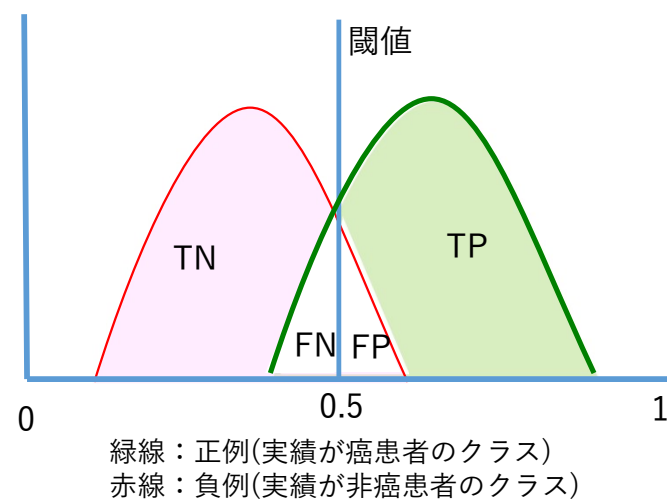
2.モデル評価 (分類)

⑧ ROC曲線(Receiver Operating Curve)

機械学習モデルは、「対象が癌である確率がいくつか?」を算出し予め決定した閾値より大きければ「癌である」、小さければ「癌でない」と判定結果を出す。その閾値を変化させた時の偽陽性率と真陽性率の値をプロットしたものがROC曲線。左上に近づくほど理想的なモデルとなる。凹みが大きいと理想から外れたモデル。



左図ROC曲線の状態に
TP、TN、FP、FNの情報を含めた分布図



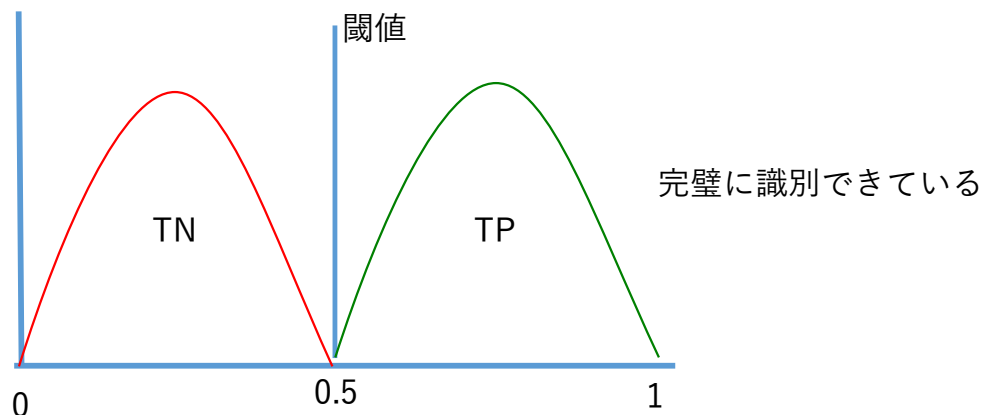
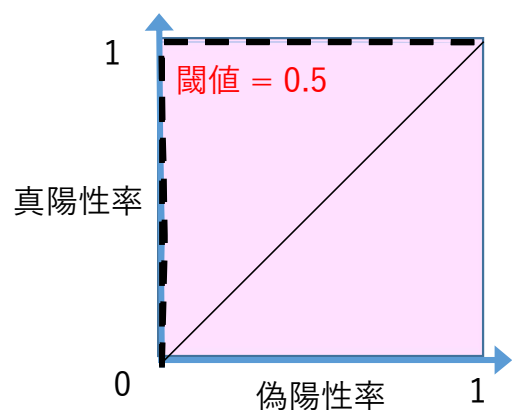
2.モデル評価 (分類)

⑧ ROC曲線 つづき

全ての癌患者を正しく予測し、全ての非癌患者を癌患者ではないと予測できた場合のROC曲線。

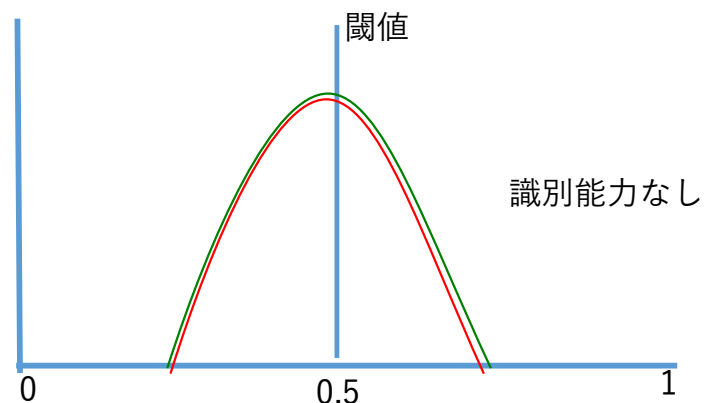
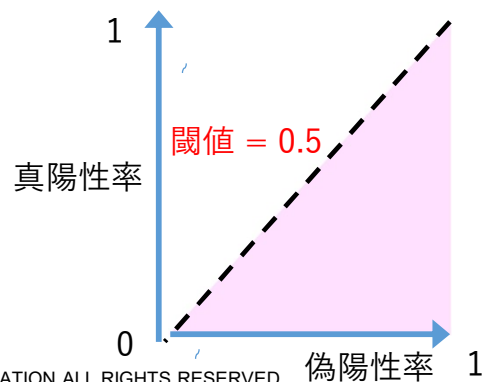
現実的にはありえないため、このようなROC 曲線の場合は、学習データのみしか予測できない過学習な状態となっている可能性が高い。

AUC = 1.0



ROC 曲線が単なる対角線になっている場合、ランダムに推測をしている状態になり機械学習モデルとしては不十分。
一般的には機械学習失敗と判断される。

AUC = 0.5



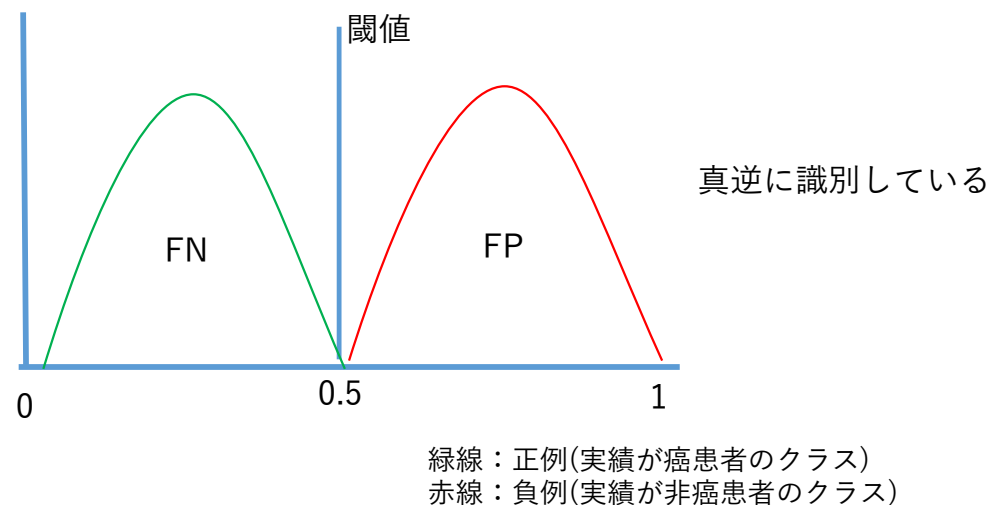
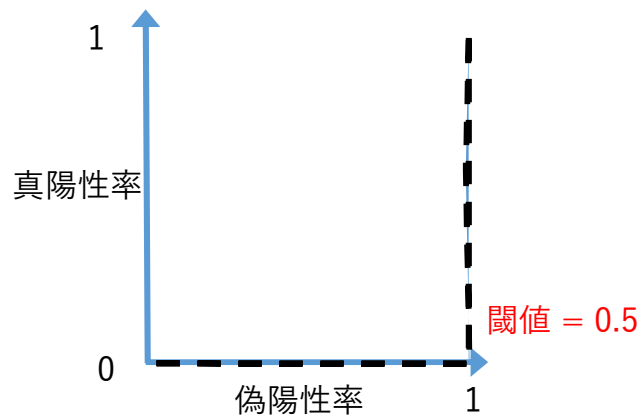
参考文献：③

2.モデル評価 (分類)

⑧ ROC曲線 つづき

ROC 曲線が右と下の辺に沿うような形の場合、全ての癌患者を非癌患者と予測し、全ての非癌患者を癌患者と予測したことを意味する。
正例と負例を真逆に予測してしまったケース。

AUC = 0.0



2.モデル評価（分類）



⑨ column Weights

- ・モデル作成時に重要度が高かった項目がランキング形式で確認できる
- ・各項目の重要度を0.0～1.0の間の値で評価する
- ・ビジネスの現場のユーザが参照し、納得感があるかどうか確認する
- ・アルゴリズム毎に算出している重要度(アルゴリズムに依存する)

Column Weights

Influence of columns on the predictions of this specific model

Search columns

Column 	Weight 
最終接触時間	0.316
住宅ローン	0.252
前回接触回数	0.115
年齢	0.111
今回接触回数	0.059
最終接触日	0.056
最終接触月	0.043
結婚歴	0.042
学歴	0.042
職種	0.029

参考文献：②

2.モデル評価 (分類)

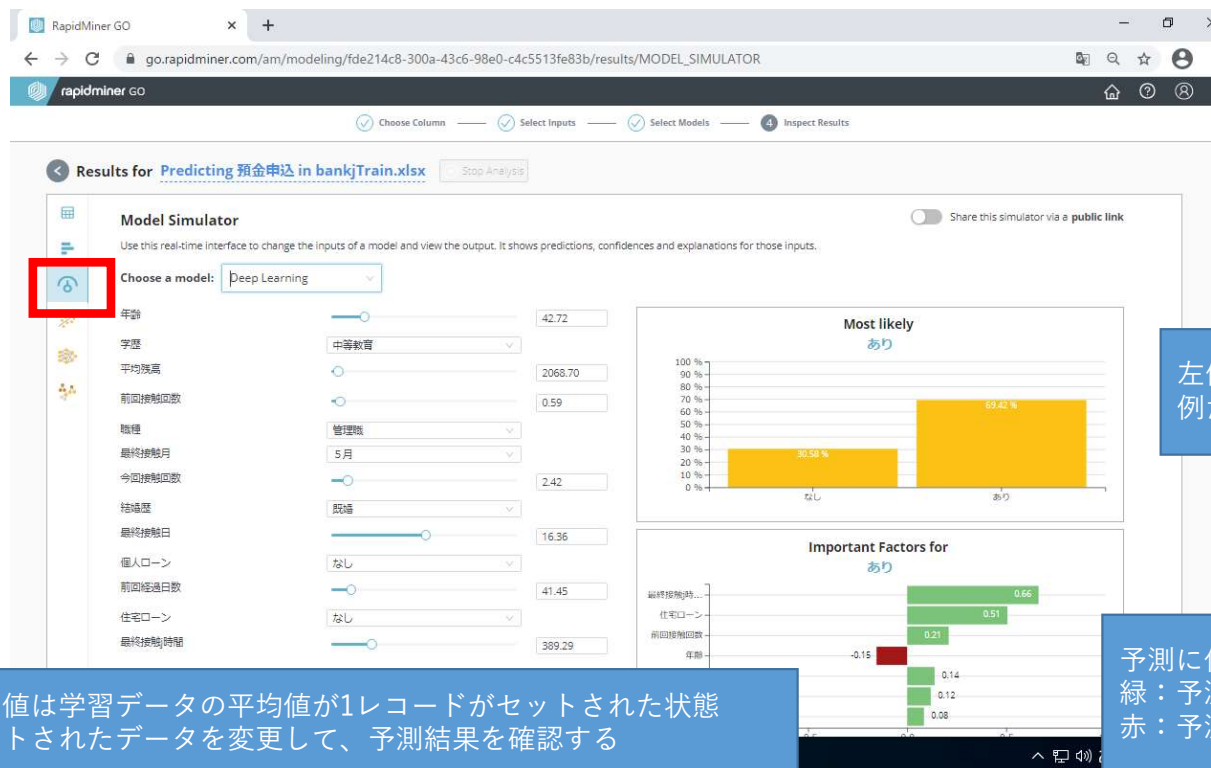
(3) モデルシミュレーター

作成したモデルを用いて、予測結果をシミュレーションすることが可能。

シミュレーション結果を元に納得感があるモデルができたかどうかの判断材料として利用する。

(例)顧客情報を使って預金申し込みするか、しないかの分類モデルを作成した場合

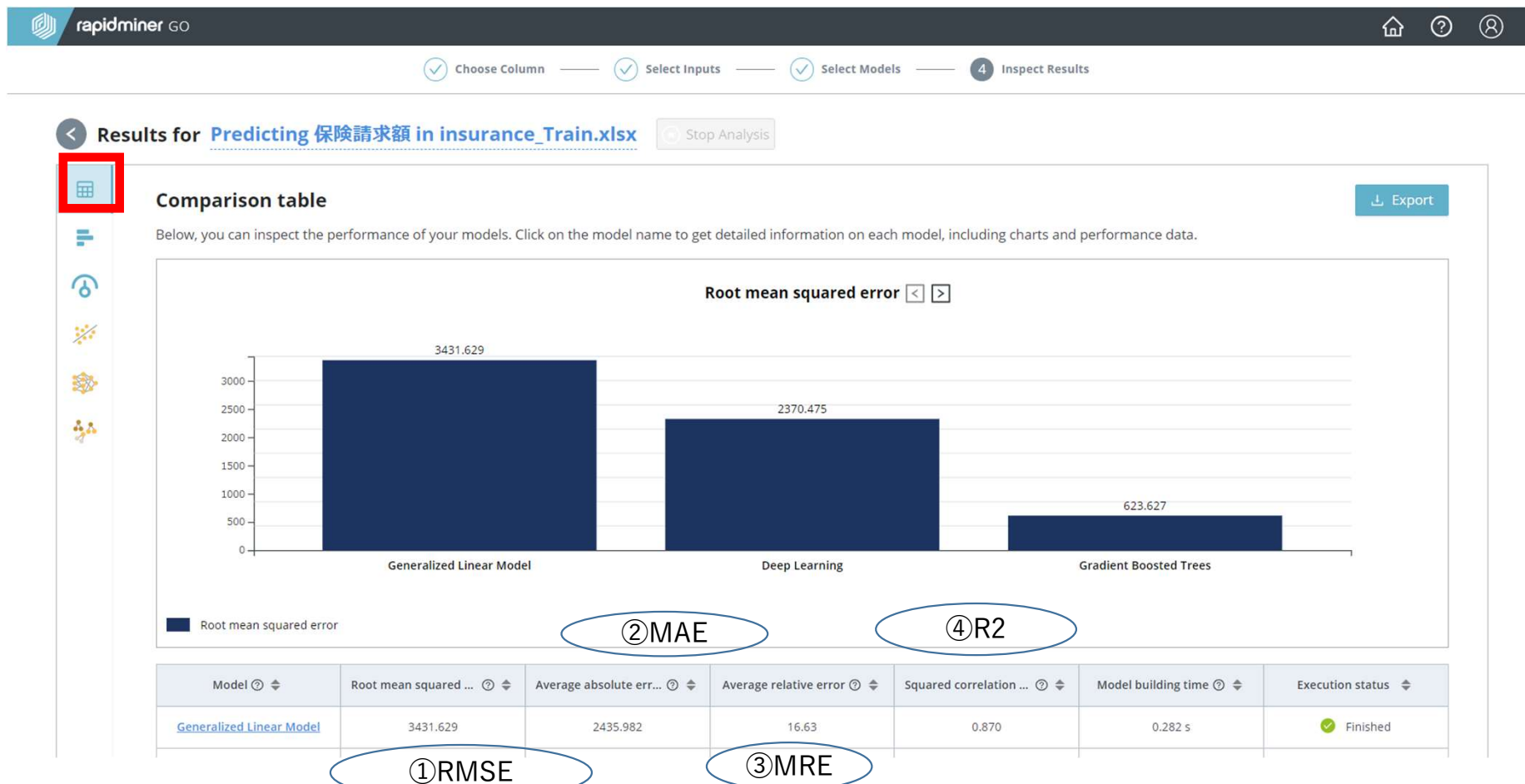
仮にこういう顧客属性の人であれば、どのくらいの確率で預金申込するかシミュレーションすることができる。(=what-if分析)



3.モデル評価 (回帰)

2.回帰モデル：数値(例：売上金額など) を予測するモデルを作成する場合

(1) モデル比較表：各モデルの評価を比較できる画面



3.モデル評価（回帰）

- ① **RMSE(Root Mean Squared Error)** 二乗平均平方根誤差
実測値と予測値のズレを表す指標。0に近いほど良いモデル。
外れ値(極端に大きい値や小さい値)の影響を受けやすい。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- ② **MAE(Mean Absolute Error)** 平均絶対誤差
実測値と予測値のズレを表す指標。0に近いほど良いモデル。
外れ値(極端に大きい値や小さい値)の影響を受けにくい。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- ③ **MRE(Mean Relative Error)** 平均相対誤差
実測値と予測値のズレの割合を表す指標。0に近いほど良いモデル。

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{|y_i|} \right)$$

- ④ **Squared correlation (R^2)** 決定係数
予測値が実測値をどれだけうまく予測しているかを客観的に表す指標。
0～1の値で評価するため、
他のモデルとの性能比較に活用しやすい。

$$\text{Squared correlation} = 1 - \left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right)$$

R^2	性能評価
0.8以上	性能が非常に高い 但し1に近すぎると、過学習している可能性があるため要検証
0.7-0.8	性能が良好
0.6-0.7	性能が認められる
0.6未満	性能が認められない

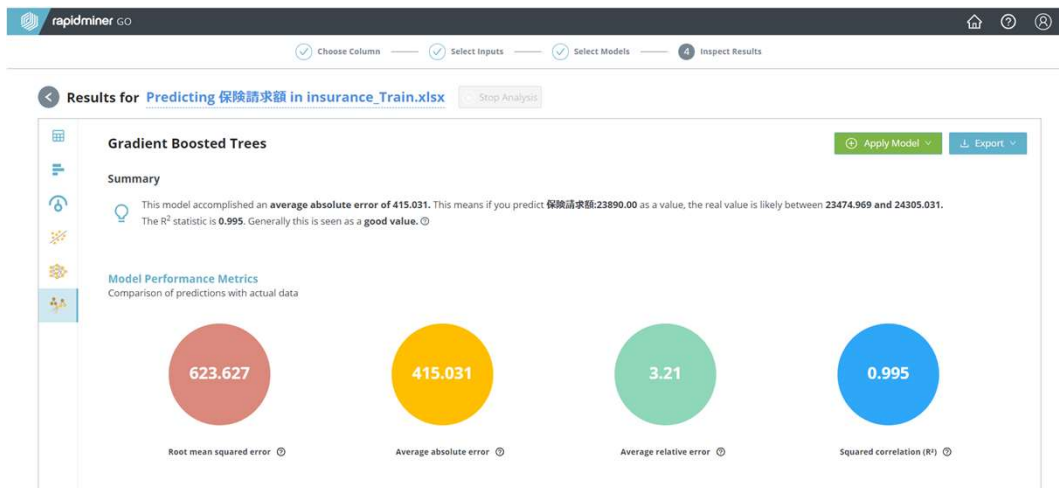
3.モデル評価（回帰）

評価指標の活用シーン

指標	活用シーン
RMSE	大きな誤差をより重視して評価したいケース。 (例) 家賃予測 外れ値(極端に高い金額や低い金額)が存在すると、RMSEの値が大きくなる。高額な物件の家賃の予測誤差を発見したい場合等に活用しやすい。
MAE	外れ値の影響を抑えてモデルを評価したいケース。 (例) 売上予測 MAEはRMSEと比較すると外れ値の影響を受けづらい。極端に売上が高い日の予測誤差の影響を抑えてモデル評価したい場合等に活用しやすい。
MRE	異なるデータセットの間で、モデル性能を比較したいケース。 (例) 東京の家賃予測モデルと大阪の家賃予測モデルの比較 実測値と予測値とのズレ（誤差）の割合を示す指標のため、家賃相場が異なる地域でのモデル性能比較等でも活用しやすい。
Squared correlation (R^2)	他のモデルとの相対的な性能比較

3.モデル評価（回帰）

(2) モデル別評価画面：各モデルの評価詳細を確認できる画面



Innovation-Leading Company

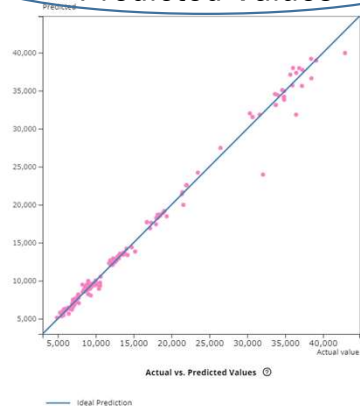
⑦column
Weights

Column Weights

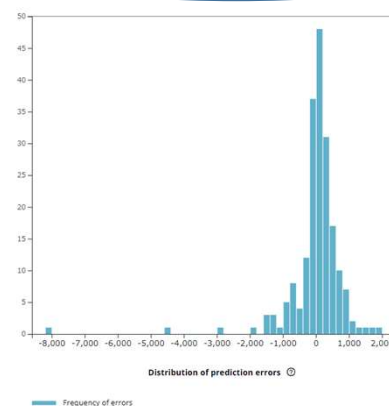
Influence of columns on the predictions of this specific model

Column	Weight
年齢	0.228
BMI	0.089
住居エリア	0.064
期間	0.062
扶養家族	0.051
性別	0.023
年収	0.013

⑤ Actual vs.
Predicted Values



⑥ Distribution of
prediction errors



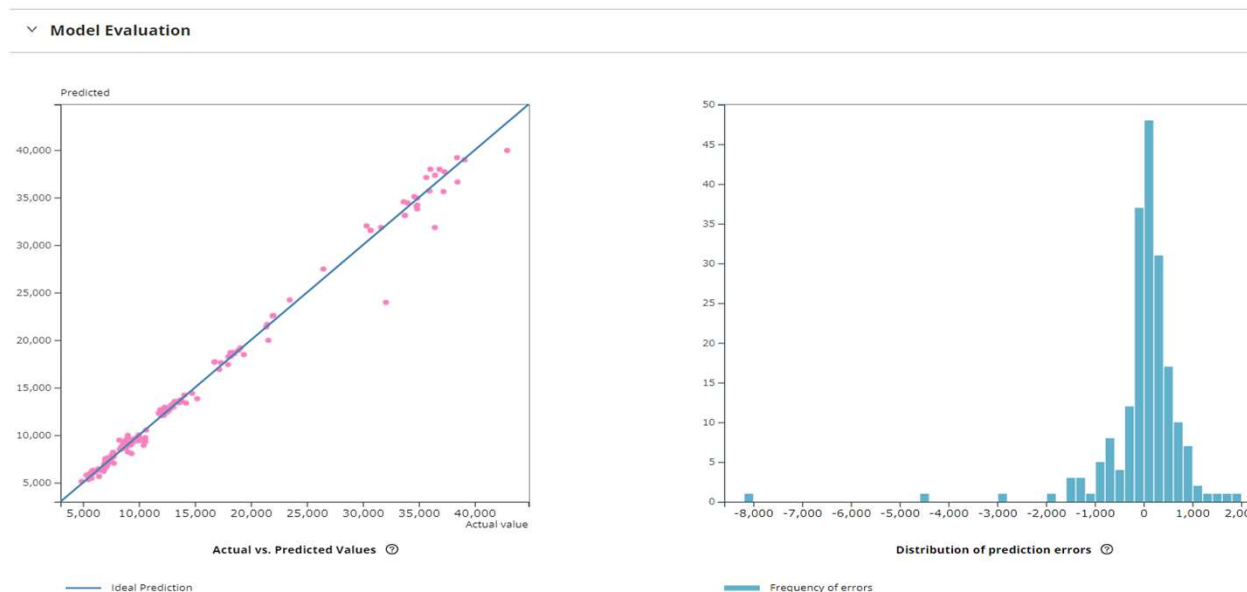
3.モデル評価（回帰）

⑤ Actual vs. Predicted Values

実績値と予測値の対比を確認。ideal Predictionの線上に集まっていれば、正しく予測できている

⑥ Distribution of prediction errors

実績値と予測値の誤差の分布を確認。0周辺に分布が大きいと、正しく予測できている



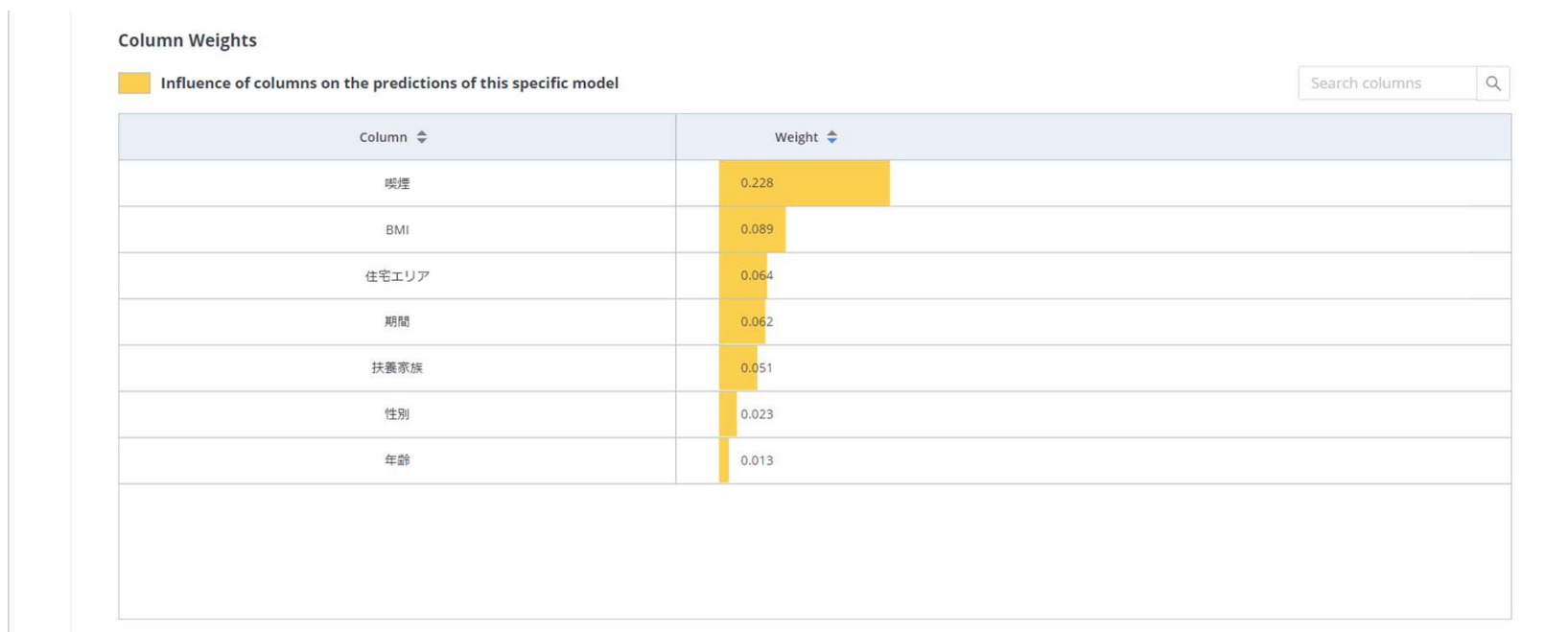
⑤ Actual vs.
Predicted Values

⑥ Distribution of
prediction errors

3.モデル評価（回帰）

⑦ column Weights

- ・モデル作成時に重要度が高かった項目がランキング形式で確認できる
- ・各項目の重要度を0.0～1.0の間の値で評価する
- ・ビジネスの現場のユーザが参照し、納得感があるかどうか確認する
- ・アルゴリズム毎に算出している重要度(アルゴリズムに依存する)



参考文献：②

3.モデル評価 (回帰)

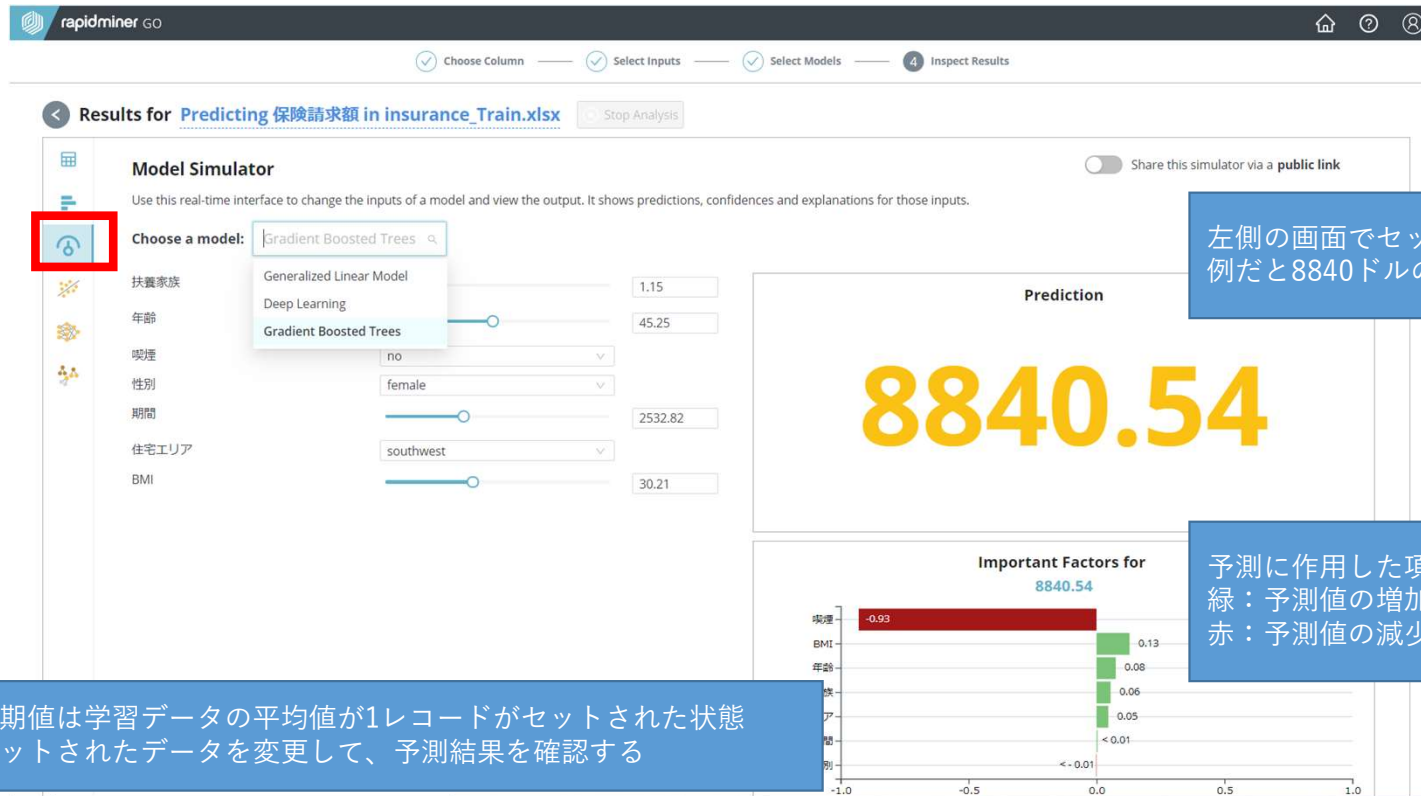
(3) モデルシミュレーター

作成したモデルを用いて、予測結果をシミュレーションすることが可能。

シミュレーション結果を元に納得感があるモデルができたかどうかの判断材料として利用する。

(例)顧客情報を使って保険請求額の回帰モデルを作成した場合

仮にこういう顧客属性の人であれば、保険請求額がいくらになるかシミュレーションすることができる。(=what-if分析)



左側の画面でセットした値に応じて、予測結果をリアルタイム表示
例だと8840ドルの保険請求額と予測されている

予測に作用した項目の影響度が大きい順に表示
緑：予測値の増加に作用した項目
赤：予測値の減少に作用した項目

- ・ 初期値は学習データの平均値が1レコードがセットされた状態
- ・ セットされたデータを変更して、予測結果を確認する

参考資料

- ① 塚本邦尊 「東京大学のデータサイエンティスト育成講座」 マイナビ出版
- ② rapidminerドキュメント(RapidMiner社)
<https://docs.rapidminer.com/latest/go/overview/>
- ③ ROC 曲線とAUC を用いて2値分類機械学習モデルの性能を計測・チューニングする
<https://qiita.com/TsutomuNakamura/items/ef963381e5d2768791d4>
- ④ ゼロからわかる機械学習モデルの性能評価Part1
<https://aizine.ai/confusion-matrix0225>
- ⑤ 回帰モデルの評価関数の種類と計算方法
<https://self-methods.com/mae-rmse-mape-r2>
- ⑥ F値と平均的でない平均の話
<https://www.cresco.co.jp/blog/entry/10325/>
- ⑦ 機械学習の評価関数（二値分類用）の基礎
<https://www.atmarkit.co.jp/ait/articles/2010/27/news020.html#f1score>

ご質問・お問合せはお気軽に
下記までお願い申し上げます。

ap-promo@nissho-ele.co.jp



日商エレクトロニクス株式会社
アプリケーション事業本部

RapidMinerトレーニング情報

