

# 株価を用いた 倒産予知モデルの構築

神奈川大学 工学部 経営工学科

経営システム工学研究室

佐藤 翔

指導教員 片桐英樹

1.研究背景

2.研究目的

3.データの概要

4.分析手法

5.分析結果

6.検討・考察

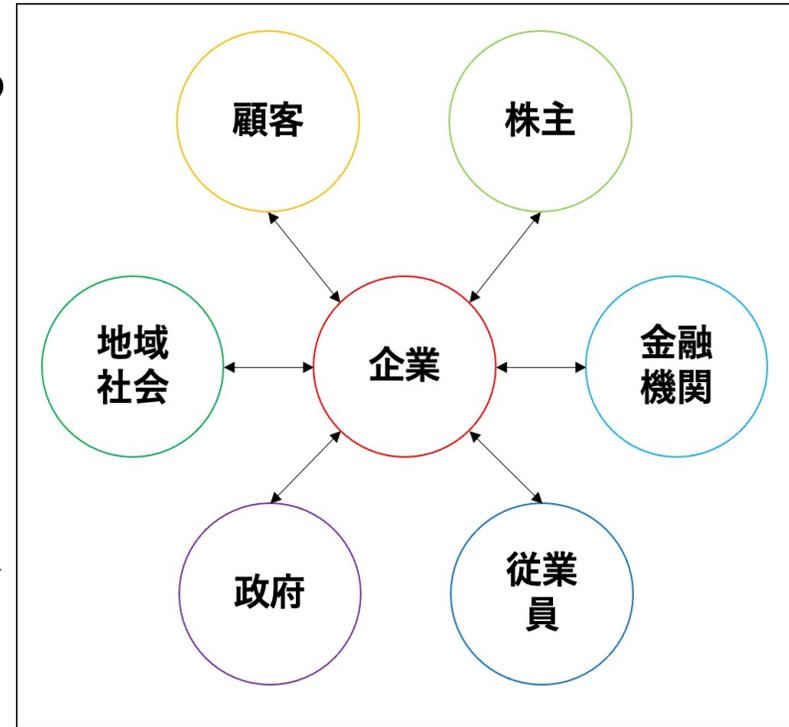
7.結論・今後の課題

参考文献

# 1. 研究背景

**倒産は、企業が経営をするうえで、直接的または間接的に影響を受ける利害関係者であるステークホルダー(stakeholder)に大きな影響を与える。**

具体的には顧客、従業員、株主、地域社会、金融機関などのステークホルダーが、**多大な損失**をうける。したがって倒産は回避をしなければならない。または、早めに**倒産予知**をすることにより、**損失を未然に防いだり、損失を少なく**することができる。



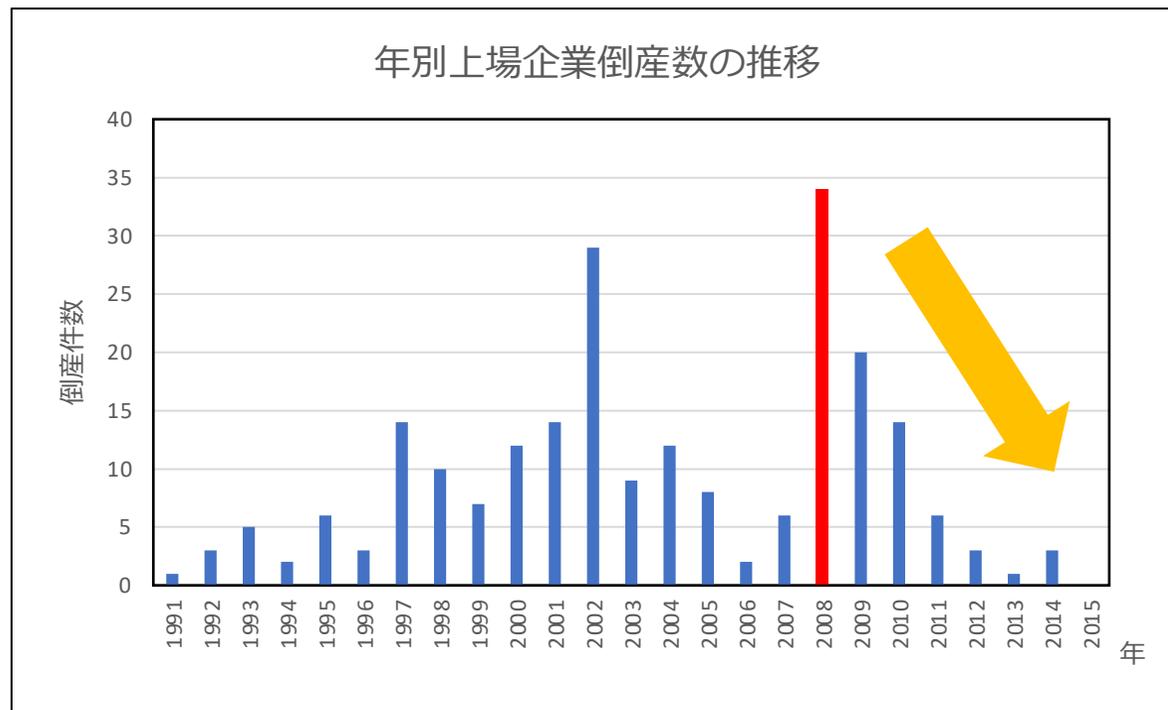
本研究ではそのようなステークホルダーおよび投資家に対して、より精度の良い倒産情報を提供するために、**倒産予知モデルの構築**を行う。

# 1. 研究背景

## 倒産件数の推移

日本の上場企業の倒産数は  
**リーマンショック**の時期を  
ピークに**約90%**減少している。  
しかし、リーマンショックの  
ような事態が今後いつ発生するか  
予測することはできない。

実際、誰も予測し得なかった新型  
コロナウイルス感染症の流行によ  
り多くの日本企業の売り上げが  
減少している。



出所：帝国データバンク『全国企業倒産集計』より作成

# 1. 研究背景

## 本研究での倒産の定義

企業が経営に行き詰まり、破綻し東京証券取引所の  
上場廃止になり、株価などのデータが取得できない状態。

倒産予知モデル構築に当たっては、企業外部に公表されている限られたデータから出来るだけ早く、かつ、精度よく倒産を予知したい立場から研究を行う。データとしては、日経メディアマーケティングが提供するNEEDS-Financial QUESTが提供する資料を用いる。

## 予知手法

倒産予知研究は長く、歴史的に見れば当初の Altman(1968)の Multiple discriminant analysis (MDA)を用いた研究に見られるように、統計的手法が数多く用いられてきた。

近年では機械学習の台頭により、統計的手法よりも判別予知精度が向上している。  
**本研究では機械学習手法を用いてモデル構築を行う。**

## 2. 研究目的

本研究では**建設業**と**不動産業**の2業界のデータを用いて、倒産判別モデルを機械学習の Decision Tree(DT)と Random Forest (RF), Neural Network (NN), Support Vector Machine(SVM)の4手法を用いて倒産予測を行う。

説明変数として、**財務指標のみの23指標**と4つの株価に関する指標を加えた**27指標**を用いることで、**株価に関する指標がモデルに与える影響**を考察する。

また、機械学習の4手法のモデル構築し、これらのモデルの予測精度を比較することで、**機械学習手法による予測精度の違い**についても考える。

財務指標

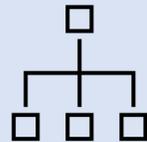
23指標を用いた  
4つのモデル

Decision Tree (DT)

Random Forest (RF)

Neural Network (NN)

Support Vector Machine (SVM)



比較



財務指標

+

株価指標

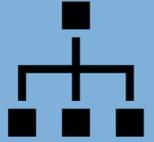
27指標を用いた  
4つのモデル

Decision Tree (DT)

Random Forest (RF)

Neural Network (NN)

Support Vector Machine (SVM)



### 3. データの概要

サンプルのデータは東京証券取引所において1991年から2015年までに上場していた企業のデータをNEEDS-Financial QUESTより収集した。

対象業種：**建設業, 不動産業**

理由：東京証券取引所の33業界において1991年から2015年までに上場していた企業のうち倒産件数が最も多いのが建設業で、次に多いのが不動産業のためである。

具体的には不動産業界においては倒産企業が33社、非倒産企業は445社、建設業界においては倒産企業が44社、非倒産企業は3534社のデータを取集した。

	建設業	不動産業
倒産企業	44	33
非倒産企業	3534	445
合計	3578	478

### 3. データの概要-株価に関する4指標**倒産**企業の場合

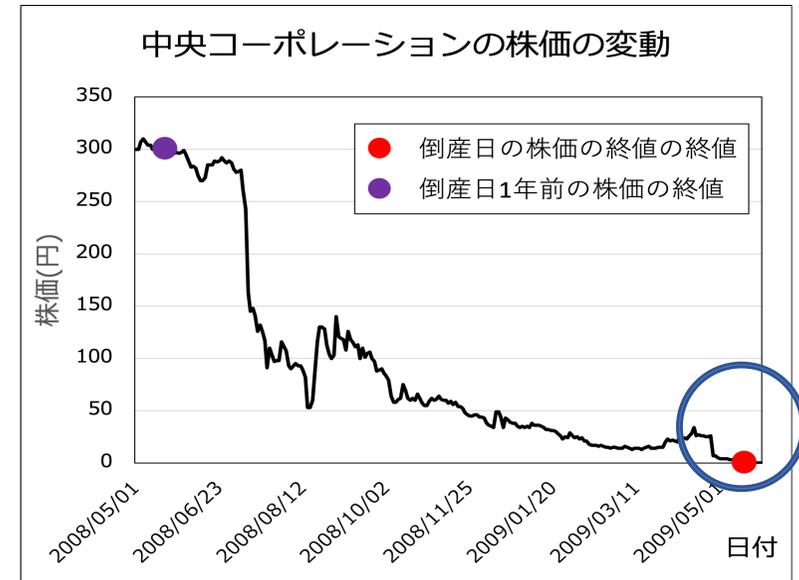
倒産企業の株価に関する4指標は,倒産日の  
 株価の終値とそれぞれ倒産1日前,1週間  
 前,1ヶ月前,1年前の株価の終値の比を取っ  
 ている.

株価日次比率 =  $\frac{\text{倒産日1日前の株価の終値}}{\text{倒産日の株価の終値}}$

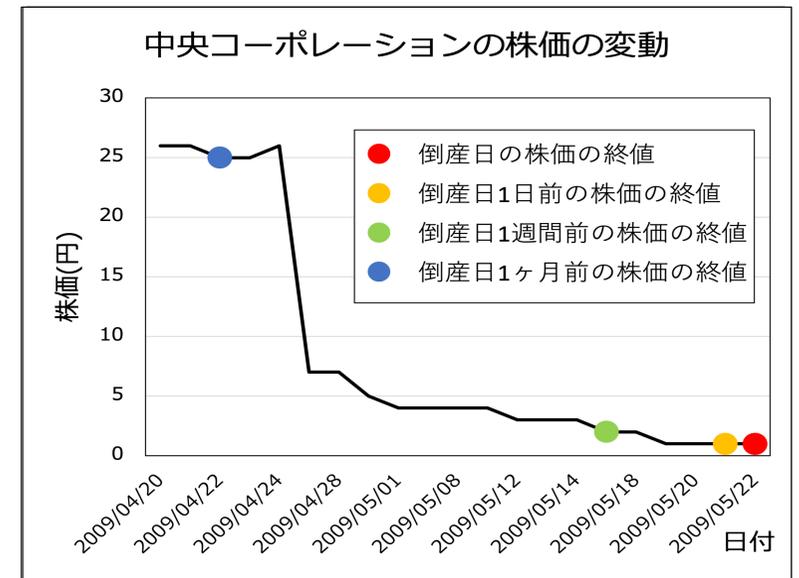
株価週次比率 =  $\frac{\text{倒産日1週間前の株価の終値}}{\text{倒産日の株価の終値}}$

株価月次比率 =  $\frac{\text{倒産日1ヶ月の株価の終値}}{\text{倒産日の株価の終値}}$

株価年次比率 =  $\frac{\text{倒産日1年の株価の終値}}{\text{倒産日の株価の終値}}$



拡大



# 3. データの概要-株価に関する4指標 **非倒産企業**の場合

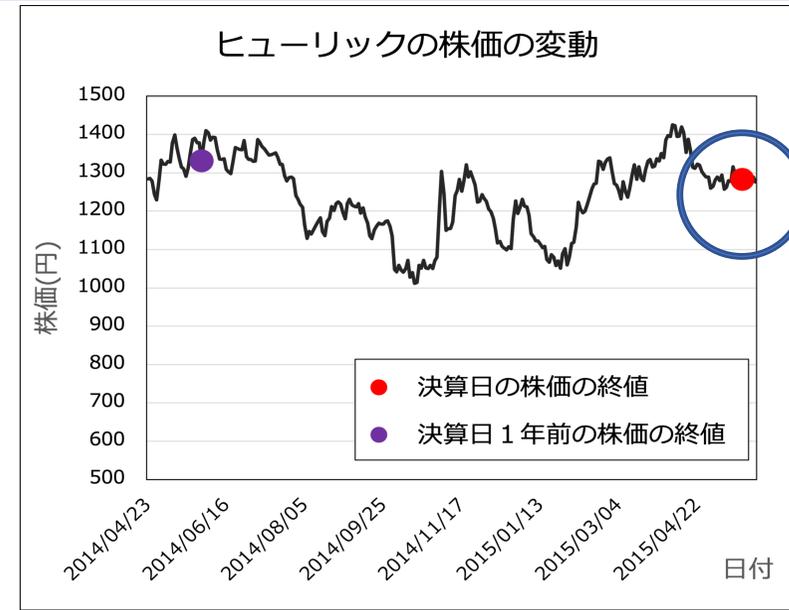
非倒産企業の株価に関する4指標は、決算日の株価の終値とそれぞれ倒産1日前,1週間前,1ヶ月前,1年前の株価の終値の比を取っている。

$$\text{株価日次比率} = \frac{\text{● 決算日1日前の株価の終値}}{\text{● 決算日の株価の終値}}$$

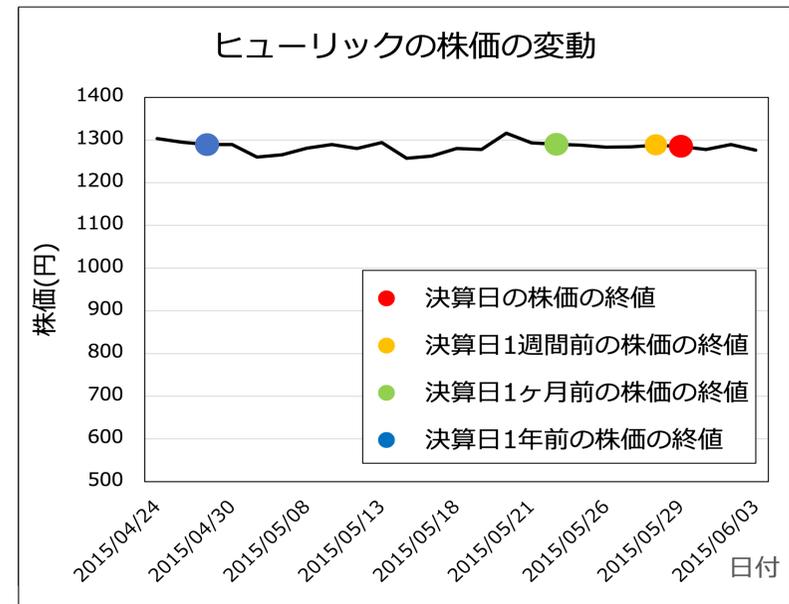
$$\text{株価週次比率} = \frac{\text{● 決算日1週間前の株価の終値}}{\text{● 決算日の株価の終値}}$$

$$\text{株価月次比率} = \frac{\text{● 決算日1ヶ月の株価の終値}}{\text{● 決算日の株価の終値}}$$

$$\text{株価年次比率} = \frac{\text{● 決算日1年の株価の終値}}{\text{● 決算日の株価の終値}}$$



拡大



# 3. データの概要-指標詳細

## 財務指標

### 23指標

- |               |              |
|---------------|--------------|
| ①デットキャパシティレシオ | ⑬借入金月商倍率     |
| ②短期借入金回転期間    | ⑭負債回転期間      |
| ③売上高純利益率      | ⑮売上高経常利益率    |
| ④借入金依存度       | ⑯インタレストカバレッジ |
| ⑤総資産有利子負債比率   | ⑰一人当たり営業利益1  |
| ⑥売上高税引前利益率    | ⑱一人当たり営業利益2  |
| ⑦自己資本分配率      | ⑲売上高金融収支比率   |
| ⑧総資本当期純利益率    | ⑳長期借入金返済年数   |
| ⑨当座比率         | ㉑金融費用比率      |
| ⑩流動負債回転期間     | ㉒経営資本営業利益率   |
| ⑪総資本経常利益率     | ㉓総資本営業利益率    |
| ⑫有利子負債月商比率    |              |

+

## 株価指標

### 4指標

- ①株価日次比率
- ②株価週次比率
- ③株価月次比率
- ④株価年次比率

## 使用する機械学習手法

23指標と27指標のそれぞれで, **Decision Tree (DT)**と **Random Forest (RF)**, **Neural Network (NN)**, **Support Vector Machine (SVM)**の4つの機械学習手法を使用し, 判別モデルを構築する. その際に交差検証をすることで過学習を避け, モデルを汎用的なものにする. モデルの構築には**Visual Mining Studio**を用いる.

## 検討・考察

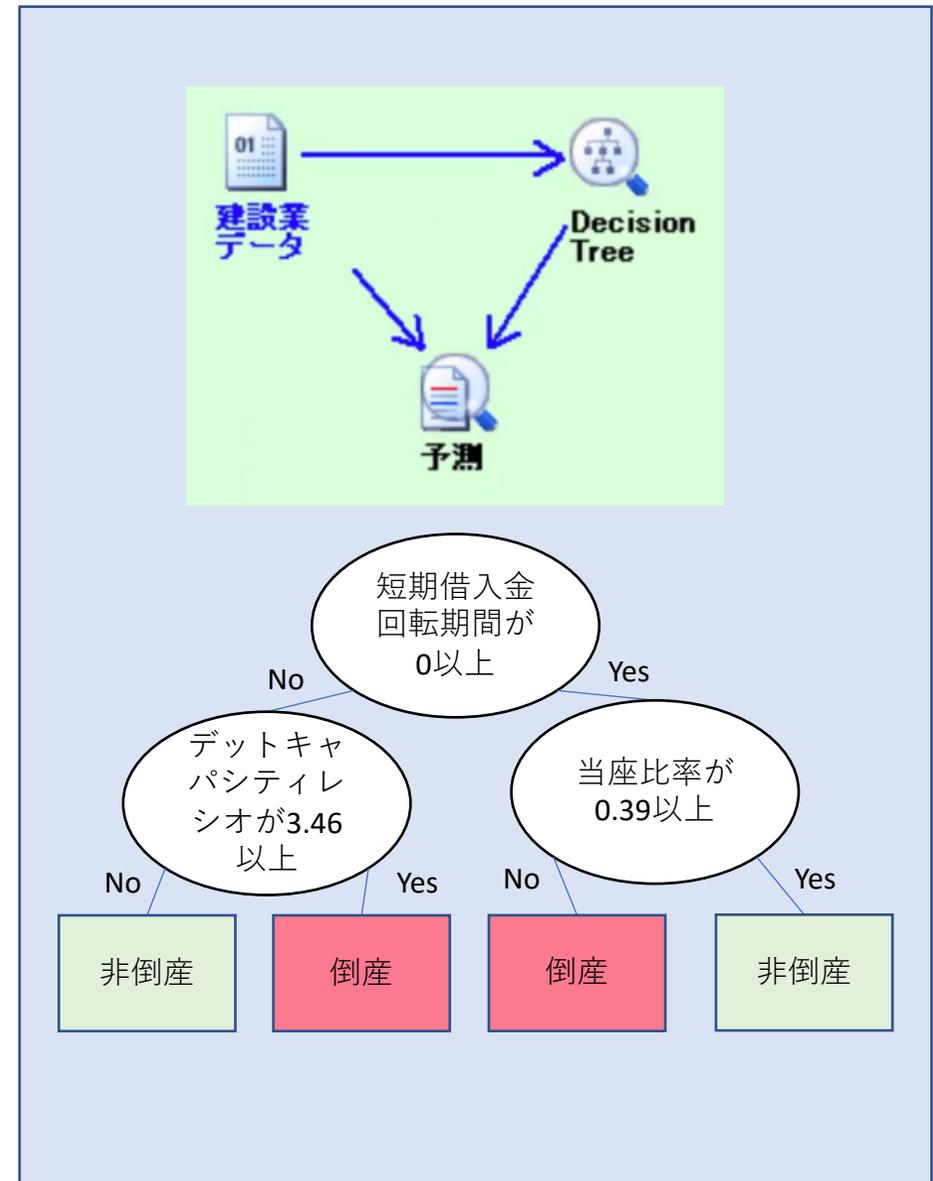
構築したモデルについて指標数や機械学習手法を変えたことによるモデルの精度の違いについて考察する. 精度を評価する指標として, 正確度 (Accuracy), 適合率 (Precision) 再現率 (Recall), 偽陽性率, 偽陰性率を用いる. これらの評価指標のうち倒産を研究対象としているので, 倒産数を多く, かつ正しく予知する**偽陰性率を重視した検討・考察を行う**. また, 説明変数重要度を用いて使用した説明変数の重要度について考察する.

# 4.分析手法- Decision Tree (DT)

学習データを条件分岐によって分割していくことによって回帰や分類を行う手法である。

以下にアルゴリズムを示す。

1. ある領域について全ての特徴量と分割候補について不純度を計算する
2. 分割したときに不純度が最も減る分割で領域を分割する
3. 分割後の領域について,1.2.を繰り返す

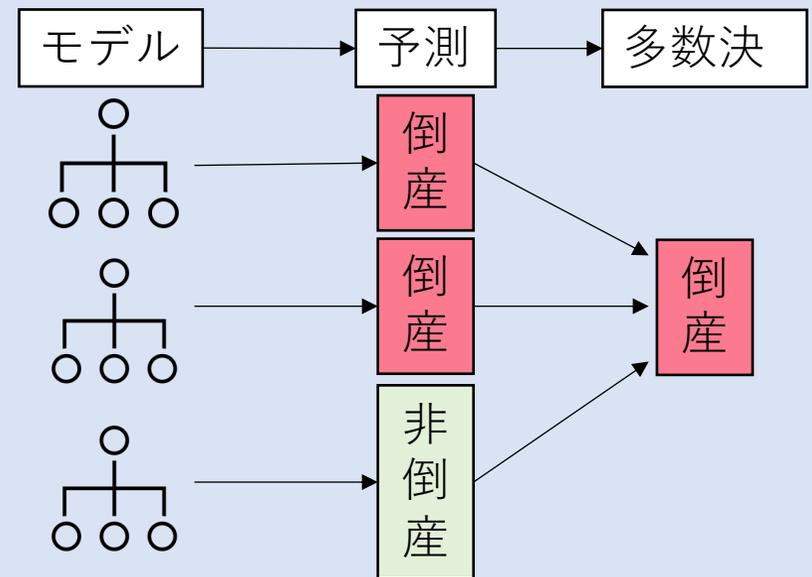
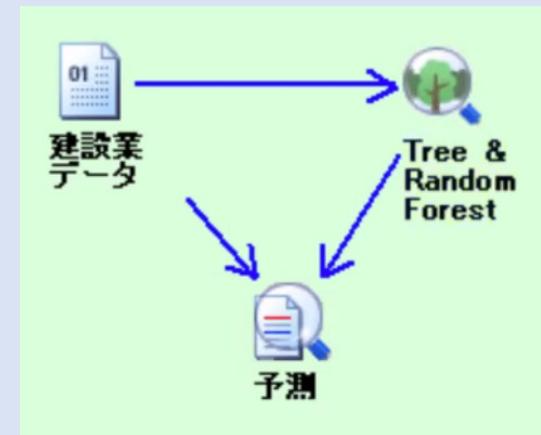


# 4.分析手法-Random Forest (RF)

決定木を複数組み合わせ、各決定木の予測結果を多数決することによって結果を得る。

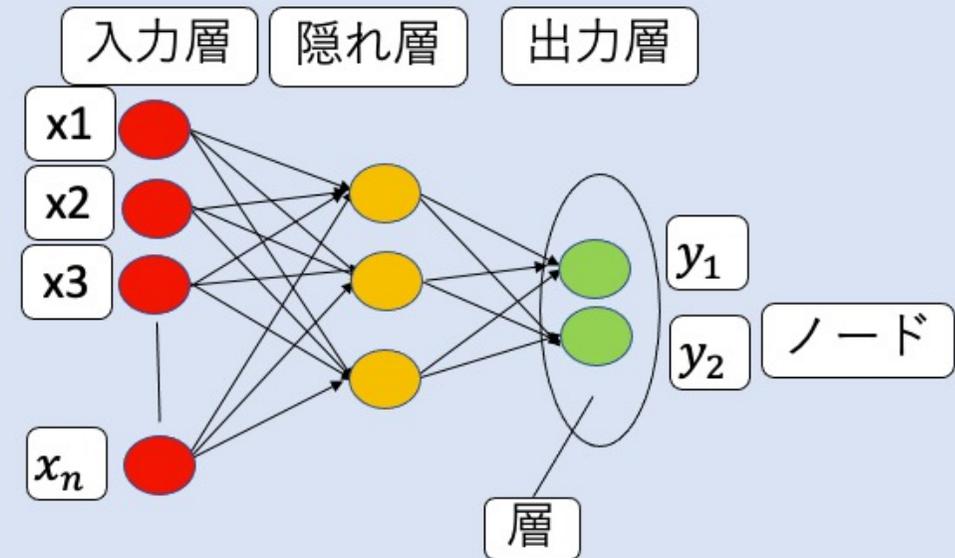
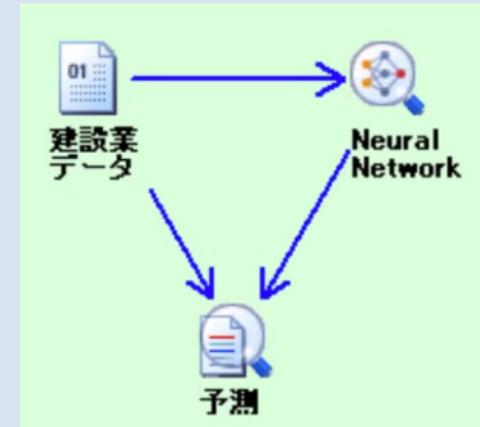
以下にアルゴリズムを示す。

1. ランダムにデータを抽出する
2. 決定木を成長させる
3. 1,2ステップを指定回繰り返す
4. 予測結果を多数決することによって分類閾値(ラベル)を決定する



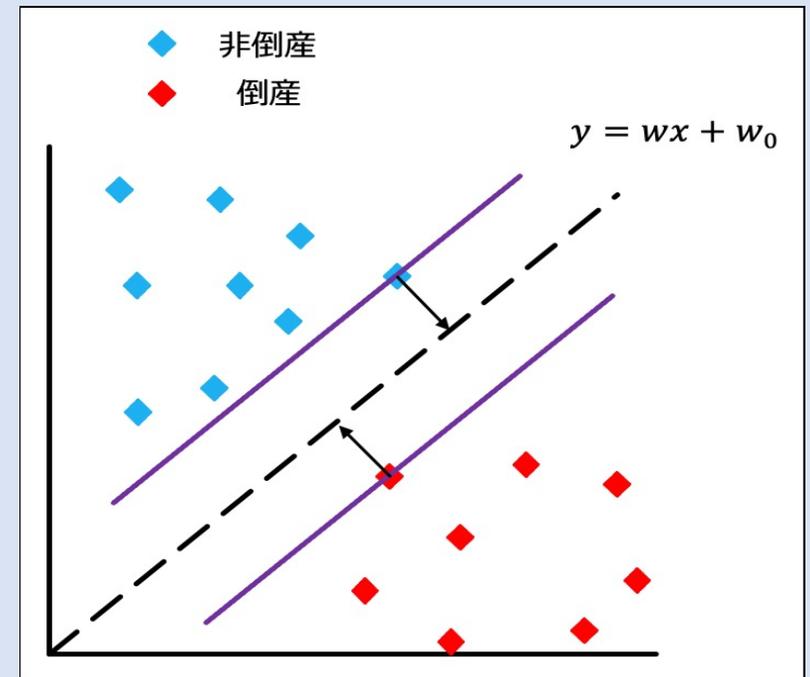
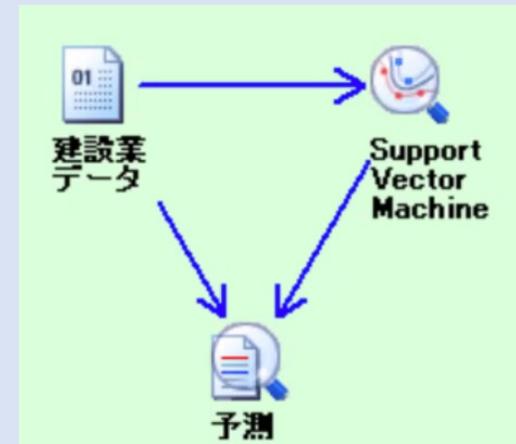
# 4.分析手法-Neural Network (NN)

人間の脳のしくみ（ニューロン間のあらゆる相互接続）から着想を得たもので、脳機能の特性のいくつかをコンピュータ上で表現するために作られた数学モデル。



# 4.分析手法-Support Vector Machine (SVM)

2つの要素を持つ二次元データが複数ある平面に直線を引いて判別を行う。  
直線はマージン最大化を用いる。多次元の要素を持つときは変数を加えてベクトルの内積で求める。



# 4.分析手法-交差検証(cross-validation)

交差検証とはデータをn個に分割し,n-1個のグループのデータを使って学習を行い,残りの一つのグループでテストを行う.これをn回繰り返して,n個の精度の平均を計算して最終的な精度とする.



交差検証

目的変数 企業名

交差検証 オプション

学習 | 重み付け

アルゴリズム  
Support Vector Machir

交差検証方法  
Leave one out

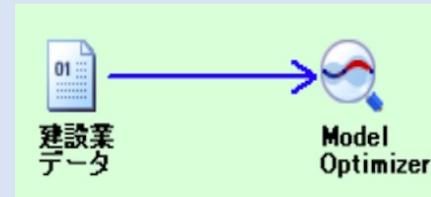
テストブロック数  
100

	Type	説明変数
企業名	カテゴリ	
倒産・非倒	カテゴリ	X
デットキャ	数値	X
短期借入金	数値	X
売上高純利	数値	X
借入金依存	数値	X
総資産有利	数値	X
売上高税引	数値	X
自己資本%	数値	X
総資本当	数値	X
当座比率	数値	X
流動負債	数値	X
総資本経	数値	X
有利子負債	数値	X
借入金月	数値	X
負債回転	数値	X
売上高経	数値	X
インタレ	数値	X
一人当	数値	X
一人当	数値	X

# 4.分析手法-パラメータ最適化

Model Optimizerは学習に適した最適なモデルを自動的に探索し予測精度の高いモデルを効率的に構築する。

探索時に交差検証(cross-validation)を行うため、凡用的に精度の高いモデルを求めることができる。



The screenshot shows the 'Model Optimizer' application window. It has a '基本設定' (Basic Settings) tab selected. The interface is divided into several sections:

- モデル候補 (Model Candidates):** Includes checkboxes for Decision Tree, Random Forest, Neural Network, and Support Vector Machine (checked).
- 最適化オプション (Optimization Options):** Includes a dropdown for '最大探索数' (400), '評価指標' (Recall), a checkbox for '重み付き平均を計算', a dropdown for '評価対象クラス' (倒産), a dropdown for '交差検証のブロック数' (5), and a dropdown for '並列実行数' (5).
- 目的変数 (Target Variable):** A dropdown menu showing '倒産・非倒産'.
- 説明変数 (Explanatory Variables):** A table with 11 rows and 5 columns. The columns are '変数名', '型', and two '採...' columns. The second row, '倒産・非倒産', is highlighted.

	変数名	型	採...	採...
1	企業名	カテゴリ	<input type="checkbox"/>	<input type="checkbox"/>
2	倒産・非倒産	カテゴリ	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3	デットキャパシティレシオ	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	短期借入金回転期間	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5	売上高純利益率	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
6	借入金依存度	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	総資産有利子負債比率 (...)	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	売上高税引前利益率	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9	自己資本分配率	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
10	総資本当期純利益率 (RO...)	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
11	当座比率	実数	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

変数: 29個

Buttons: 実行, 閉じる

# 4.分析手法-5つの評価指標

## 1.正確度 (Accuracy)

正確度とは,全体のデータの中で正しく分類されたものの割合を示す.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

		モデルの予測	
		倒産 (Positive)	非倒産 (Negative)
実際の クラス	倒産 (Positive)	True Positive (TP) 正しく倒産と判別	False Negative (FN) 誤って非倒産と判別
	非倒産 (Negative)	False Positive (FP) 誤って倒産と判別	True Negative (TN) 正しく非倒産と判別

## 2.適合率 (Precision)

適合率はPositiveと判別されたデータの中で実際にPositiveだった割合を示す。

今回の場合は,倒産と判別された中実際に倒産した企業の割合となる。

$$\text{Precision} = \frac{TP}{TP+FP}$$

		モデルの予測	
		倒産 (Positive)	非倒産 (Negative)
実際の クラス	倒産 (Positive)	True Positive (TP) 正しく倒産と判別	False Negative (FN) 誤って非倒産と判別
	非倒産 (Negative)	False Positive (FP) 誤って倒産と判別	True Negative (TN) 正しく非倒産と判別

## 3.再現率 (Recall)

再現率とはPositiveデータの中で実際にpositiveと分類された割合を示す。

今回の場合は,倒産企業の中の実際に倒産と判別された割合となる。

$$\text{Recall} = \frac{TP}{TP+FN}$$

		モデルの予測	
		倒産 (Positive)	非倒産 (Negative)
実際の クラス	倒産 (Positive)	True Positive (TP) 正しく倒産と判別	False Negative (FN) 誤って非倒産と判別
	非倒産 (Negative)	False Positive (FP) 誤って倒産と判別	True Negative (TN) 正しく非倒産と判別

## 4.偽陽性率 (False Positive Rate)

偽陽性率とは、非倒産企業の中の誤って倒産と判別された割合である。

$$\text{偽陽性率} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

		モデルの予測	
		倒産 (Positive)	非倒産 (Negative)
実際の クラス	倒産 (Positive)	True Positive (TP) 正しく倒産と判別	False Negative (FN) 誤って非倒産と判別
	非倒産 (Negative)	False Positive (FP) 誤って倒産と判別	True Negative (TN) 正しく非倒産と判別

## 4.分析手法-5つの評価指標

### 5.偽陰性率 (False Negative Rate)

偽陰性率とは、倒産企業の中の誤って非倒産と判別された割合である。

$$\text{偽陰性率} = \frac{\text{FN}}{\text{TP} + \text{FN}}$$

		モデルの予測	
		倒産 (Positive)	非倒産 (Negative)
実際の クラス	倒産 (Positive)	True Positive (TP) 正しく倒産と判別	False Negative (FN) 誤って非倒産と判別
	非倒産 (Negative)	False Positive (FP) 誤って倒産と判別	True Negative (TN) 正しく非倒産と判別

# 5.分析結果-建設業結果まとめ

全体の結果のまとめを以下に示す.

建設業	手法	正確度	適合率	再現率	偽陽性率	偽陰性率
23指標	DT	98.93	75.00	20.45	0.08	79.55
	RF	99.77	0.00	0.00	0.00	100.00
	NN	98.80	57.14	9.09	0.08	90.91
	SVM	98.85	100.00	6.82	0.00	93.18
27指標	DT	99.63	80.39	93.18	0.28	6.82
	RF	99.63	80.39	93.18	0.28	6.82
	NN	99.86	95.35	93.18	0.06	6.82
	SVM	99.64	96.97	72.73	0.03	27.27

一番精度が  
良い

**偽陰性率**より23指標の倒産判別の精度が悪いことがわかる.

27指標の**偽陰性率**を見るとDT,RF,NNが同じ6.82%を示しているが,**偽陽性率**を見ると0.06%のNNがDTの0.28%とRFの0.28%より低い.

よって**建設業では27指標のNNが最も精度が良かった.**

# 5.分析結果-不動産業の結果まとめ

全体の結果のまとめを以下に示す.

不動産業	手法	正確度	適合度	再現率	偽陽性率	偽陰性率
23指標	DT	97.70	84.38	81.82	1.12	18.18
	RF	99.16	96.77	90.91	0.22	9.09
	NN	99.58	96.97	96.97	0.22	3.03
	SVM	99.37	94.12	96.97	0.45	3.03
27指標	DT	98.32	90.32	84.85	0.67	15.15
	RF	99.60	100.00	93.94	0.00	6.06
	NN	100.00	100.00	100.00	0.00	0.00
	SVM	100.00	100.00	100.00	0.00	0.00

一番精度が良い

**偽陰性率**より23指標の倒産判別の精度が悪いことがわかる.  
 27指標のを見るとNN,SVMの**偽陽性率**と**偽陰性率**が0.00%を示している.  
**不動産業では27指標のNNとSVMが最も精度が良かった.**

# 5. 分析結果- 建設業の混同行列(1)

建設業のDTでは27指標モデルの正しく倒産と判別された企業は**41**社であった。23指標モデルの9社と比べると**32**社多く判別できているため株価指標の影響があったといえる。

建設業のDecision Tree (DT)の混同行列

23指標		モデルの予測		27指標		モデルの予測	
		倒産	非倒産			倒産	非倒産
実際のクラス	倒産	9	35	実際のクラス	倒産	<b>41</b>	3
	非倒産	3	3531		非倒産	10	3534

建設業のDTでは27指標モデルの正しく倒産と判別された企業は**41**社であった。23指標モデルの0社と比べると**41**社多く判別できているため株価指標の影響があったといえる。  
23指標モデルとき全ての企業が非倒産と判別された。

建設業のRandom Forest (RF)の混同行列

23指標		モデルの予測		27指標		モデルの予測	
		倒産	非倒産			倒産	非倒産
実際のクラス	倒産	0	44	実際のクラス	倒産	<b>41</b>	3
	非倒産	0	3534		非倒産	10	3534

# 5. 分析結果- 建設業の混同行列(2)

建設業のNNにおいて、23指標モデルでは正しく倒産と判別できた企業が4社だったのに対して、27指標モデルでは41社を判別できている。そのため**株価指標の影響があったといえる。**

建設業のNeural Network (NN)の混同行列

23指標		モデルの予測		27指標		モデルの予測	
		倒産	非倒産			倒産	非倒産
実際のクラス	倒産	4	40	実際のクラス	倒産	41	3
	非倒産	3	3531		非倒産	2	3532

建設業のSVMでは23指標モデルの正しく倒産と判別された企業は3社、27指標モデルではの32社となった。**他のDF,RF,NNが41社判別できていたのに対し精度の低い結果となった。**

建設業のSupport Vector Machine (SVM)の混同行列

23指標		モデルの予測		27指標		モデルの予測	
		倒産	非倒産			倒産	非倒産
実際のクラス	倒産	3	41	実際のクラス	倒産	32	12
	非倒産	0	3534		非倒産	1	3533

NNの27指標モデルにおいて、倒産を正しく倒産と判別した企業数は41社でDTとRFの結果(p.24)と同じだが、非倒産を誤って倒産と判別した企業については**2社**で、DTとRFの10社より少ない。したがって、**NNはDTとRFより優れたモデルといえる。**

# 5. 分析結果- 不動産業の混同行列(1)

不動産業のDTでは27指標モデルで正しく倒産と判別された企業は28社であった. 23指標の27社と比べると1社多く判別できている. そのため**株価指標の影響があったと考えられる.**

不動産業のDecision Tree (DT)の混同行列

		モデルの予測	
		倒産	非倒産
23指標		27	6
実際のクラス	倒産	27	6
	非倒産	5	440

		モデルの予測	
		倒産	非倒産
27指標		28	5
実際のクラス	倒産	28	5
	非倒産	3	442

不動産業のRFでは27指標モデルで正しく倒産と判別された企業は31社であった. 23指標モデルの30社と比べると1社多く判別できている.

27指標モデルにおいて, DTとRFを比べると**RFの方が倒産企業を正しく判別できている数が多いため, 精度が良いといえる.**

不動産業のRandom Forest (RF)の混同行列

		モデルの予測	
		倒産	非倒産
23指標		30	3
実際のクラス	倒産	30	3
	非倒産	1	444

		モデルの予測	
		倒産	非倒産
27指標		31	2
実際のクラス	倒産	31	2
	非倒産	0	445

# 5. 分析結果- 不動産業の混同行列(2)

不動産業のNNでは27指標モデルで正しく倒産と判別された企業は33社であった。23指標モデルの32社と比べると1社多く判別できている。そのため**株価指標の影響があったと考えられる。**

不動産業のNeural Network (NN)の混同行列

		23指標		モデルの予測	
		倒産	非倒産	倒産	非倒産
実際のクラス	倒産	32	1	32	1
	非倒産	1	440	1	440

		27指標		モデルの予測	
		倒産	非倒産	倒産	非倒産
実際のクラス	倒産	32	1	<b>33</b>	0
	非倒産	1	440	0	445

不動産業のSVMでは23指標モデルで正しく倒産と判別された企業は32社であった。23指標モデルにおいて,NNとSVMを比べるとNNが非倒産を倒産と判別した1社に対してSVMでは2社となっているので**NNの方が精度が良い。**

不動産業のSupport Vector Machine (SVM)の混同行列

		23指標		モデルの予測	
		倒産	非倒産	倒産	非倒産
実際のクラス	倒産	32	1	32	1
	非倒産	2	443	2	443

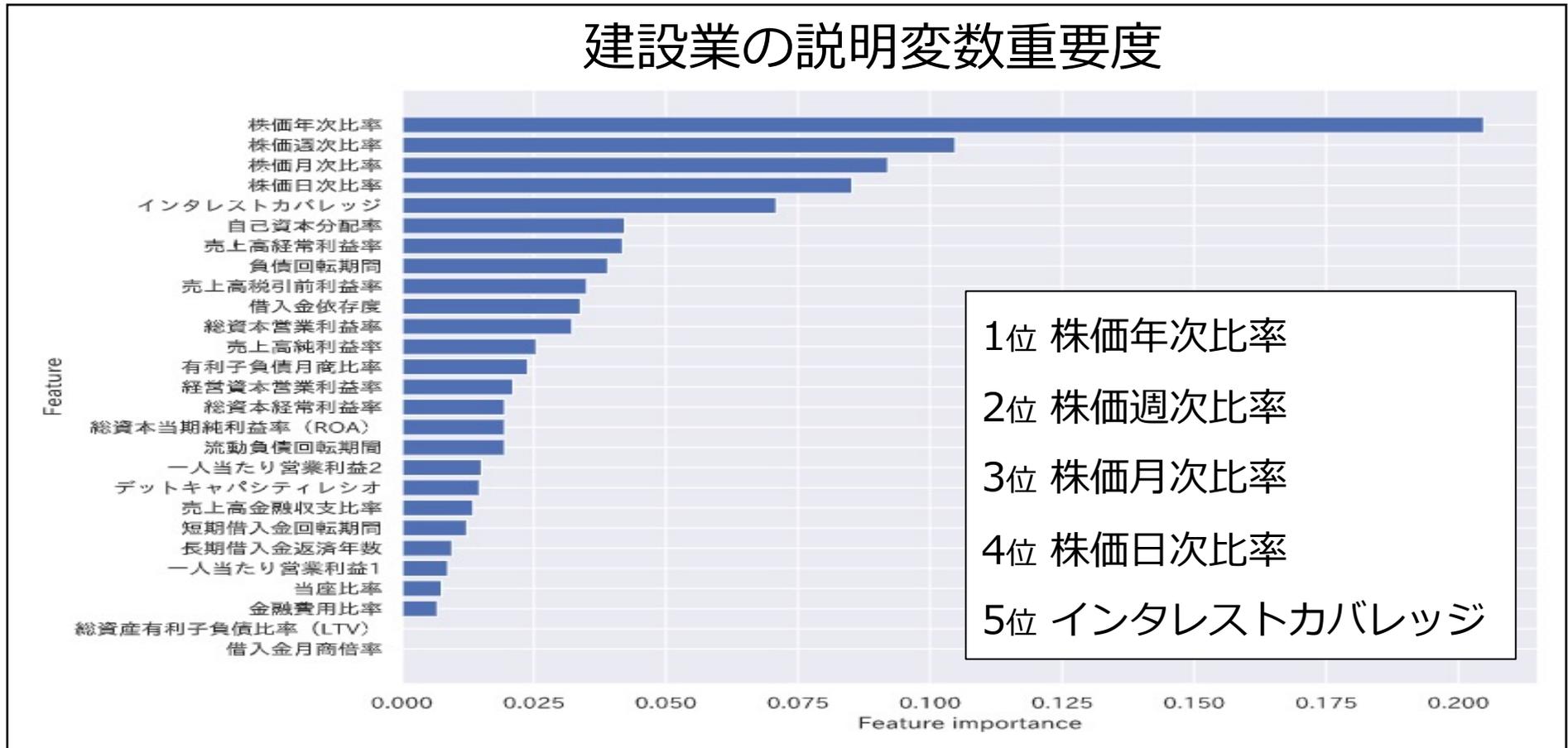
		27指標		モデルの予測	
		倒産	非倒産	倒産	非倒産
実際のクラス	倒産	32	1	<b>33</b>	0
	非倒産	2	443	0	445

**27指標モデルのSVMとNNでは完璧に倒産と非倒産を判別することができた。**

## 建設業の説明変数重要度(Future importance)

説明変数重要度(Future importance)により,モデルが重要視している要因を明確に知ることができる.

### 建設業の説明変数重要度



## 建設業

株価年次比率, 株価週次比率, 株価月次比率および株価日時比率の株価の4指標が, 1位から4位にランクされた。

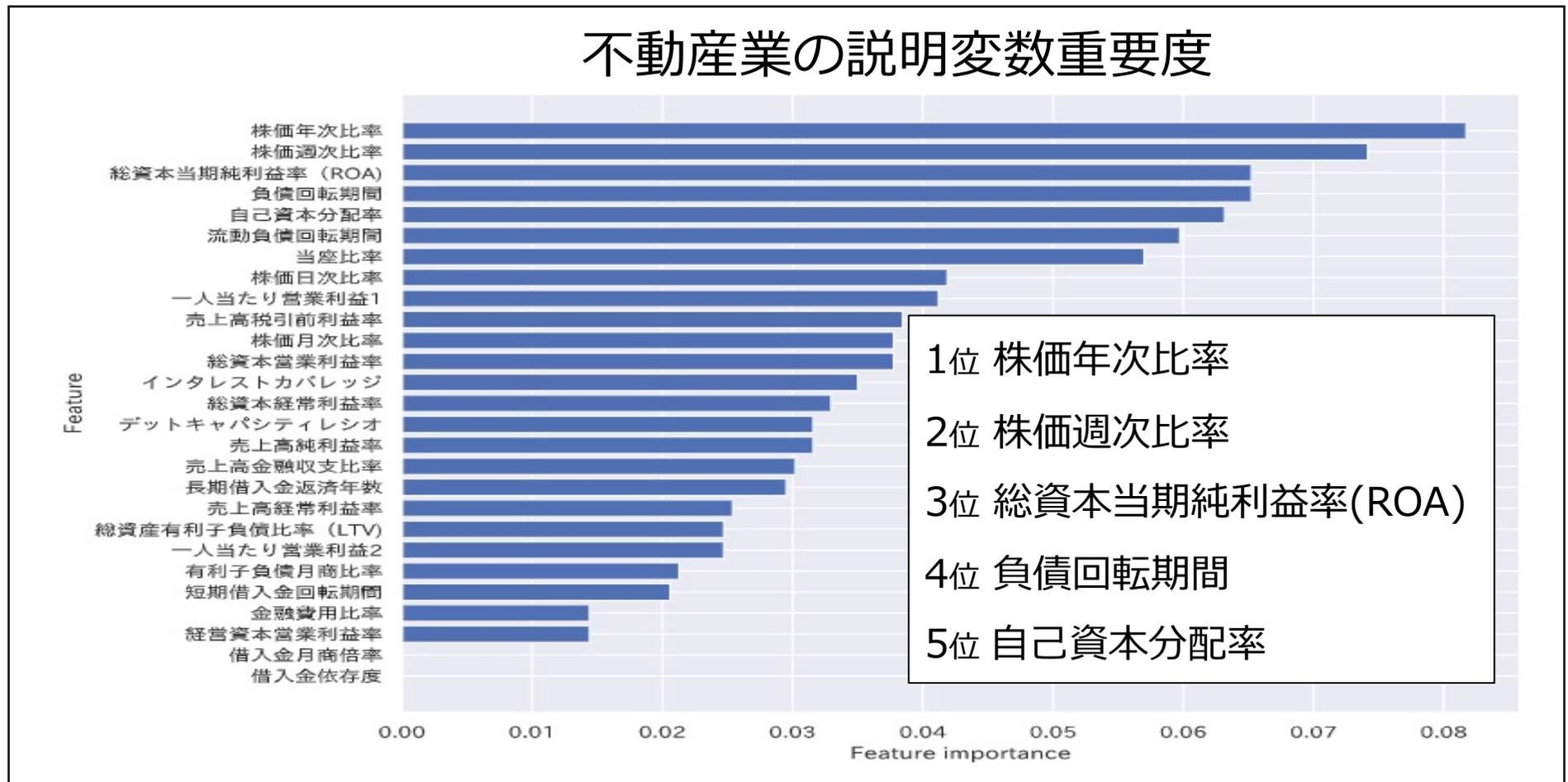
財務指標の23指標では, インタレストインタレストカバレッジが5位にランクされ, 財務指標では一番影響力の強い指標として抽出された。

$$\text{インタレストカバレッジレシオ} = \frac{\text{営業利益} + \text{受取利息} + \text{受取配当金}}{\text{支払利息} + \text{割引料}}$$

インタレストカバレッジレシオは企業の**債務返済能力**を表す指標として用いられる。

企業のデフォルト率予測モデルにもよく用いられる指標で, 本研究でも上位にランクされ, **本研究のモデルは妥当な結果を得たといえる**。企業の信用力を評価するために, 証券アナリストにも企業の安全性分析を行う際にも用いられる。

## 不動産業の説明変数重要度(Future importance)



# 6. 検討・考察 - 不動産業

## 不動産業

株価年次比率,および株価週時比率の株価の2指標が,1位と2位にランクされた。財務指標の23指標では,総資本当期純利益率(ROA)が3位に負債回転期間が4位にランクされた。

$$\begin{aligned} \text{総資本当期純利益率(ROA)} &= \frac{\text{当期純利益}}{\text{総資産}} = \frac{\text{当期純利益}}{\text{売上高}} \times \frac{\text{売上高}}{\text{総資産}} \\ &= \text{売上高利益率} \times \text{資産回転率} \end{aligned}$$

総資本当期純利益率(ROA)は,会社の事業に対して投資された資産について,どれだけ効率よく収益を得ているかを示す指標であり,良く用いられる。

$$\text{負債回転期間} = \frac{\text{流動負債} + \text{固定負債}}{\text{売上高}} \times 12$$

負債回転期間は,負債総額が売上高(月商)の何か月分に相当するかを見る指標である。

所有の不動産の物件が売却できなければ,倒産に陥ってしまう。**不動産業の特性を表す指標が上位にランクされた。**

## 結論

本研究では, Decision Treeと Random Forest, Neural Network , Support Vector Machine 4つの機械学習モデルを用いて建設業3578社と不動産業の478社の23の財務指標と株価に関する4指標を追加した27指標を用いて倒産予知モデルを構築した.

その結果建設業では**Neural Network** , 不動産業では**Neural Network** と**Support Vector Machine**の精度が良かった.

また, **株価指標を追加した27指標の方が**23指標よりも**良い倒産予知の精度を確認することができた.**

## 今後の課題

本研究の結果より,さらなる倒産予知モデルの精度向上を目指す.

- 倒産研究では,非倒産企業数に対して,倒産企業数が少ないという不均衡データを扱わざるを得ない. この解決のためにSMOTO(Synthetic Minority Over-sampling Technic)を用いて不均衡データの影響を回避することを予定している.
- 本研究の結果より, 倒産予知には株価が重要な役割を果たしていることが明確になった. **Visual Mining Studio** には**生存時間解析**も用意されているので,**株価を用いた生存時間解析により倒産日を予測することを予定している.**
- 倒産日が予測できれば,将来の倒産の可能性について,より精度の良い精度の情報を企業に提供することができる.

# 参考文献

- [1] 片桐英樹, 松丸正延, **XGBoost** と多期間データセットを用いた業界別倒産予測モデル構築, 経営システム学会誌, 第38巻, 第1号, pp.1-10, 2021
- [2] Altman, E.I., Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, The Journal of Finance, Vol. 23, No. 4, pp.589-609, 1968
- [3] Friedman, J.H., Greedy function approximation A gradient boosting machine, The Annals of Statistics, Vol.29, No.5, pp.1189-1232, 2001
- [4] Cortes, C., and Vapnik, V., Support-vector networks, Machine Learning, vol. 20, no. 3, pp. 273–297, 1995
- [5] Vapnik, V., in Haykin, S. (Ed.), Statistical Learning Theory, Adaptive and Learning Systems, Vol. 736, John Wiley & Sons Inc, 1998
- [6] Weston, J., Gammernan, A., Stitson, M., Vapnik, V., Vovk, V., and Watkins, C., Support Vector Density Estimation, In B. Scholkopf, C. Burges and A. Smola, Advances in Kernel Methods: Support Vector Machines, MIT Press, Cambridge, MA, pp. 293–306, 1999

- [7] Crammer, K., and Singer, Y., On the algorithmic implementation of multiclass kernel-based vector machines, *Journal of Machine Learning Research*, vol. 2, pp. 265–292, 2001
- [8] Hsu, C. W., and Lin, C. J., A simple decomposition method for support vector machines, *Machine Learning*, vol. 46, no. 1, pp. 291–314, 2002
- [9] Kim, K. J., Financial time series forecasting using support vector machines, *Neurocomputing*, vol. 55, no. 1, pp. 307–319, 2003
- [10] Min, J. H., and Lee, Y. C., Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, *Expert Systems with Applications*, vol. 28, no. 4, pp.603–614, 2005
- [11] Shin, K. S., Lee, T. S., and Kim, H. J., An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, vol. 28, no. 1, pp. 127–135, 2005
- [12] Min, S. H., Lee, J., and Han, I., Hybrid genetic algorithms and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, vol. 31, no. 3, pp. 652–660, 2006

- [13] Boyacioglu, M.A., Kara, Y. and O.K. Baykan, Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey, *Expert Systems with Applications*, vol.36, pp.3355–3366, 2009
- [14] Yang, Z., You, W., and Ji, G., Using partial least squares and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, vol. 38, no. 7, pp. 8336–8342, 2011
- [15] Kim, S. Y., Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis, *The Service Industries Journal*, vol. 31, no. 3, pp. 441–468, 2011
- [16] Chen, M.Y., A hybrid model for business failure prediction-utilization of particle swarm optimization and support vector machines, *Neural Network World*, vol. 21, no. 2, pp. 129-152, 2011
- [17] Fatih E., Comparing the bank failure prediction performance of neural networks, support vector machines: The Turkish case, *Ekonomiska Istrazivanja-Economic Research*, vol.26, no.3, pp.81-98, 2013

- [18] DONG, Y., HAO, X., SATO, H., Investigation of the Impact of Data Comparability on Performance of Support Vector Machine Models for Credit Scoring, Innovation and Supply Chain Management, Vol.9, No.1, 31-38, 2015
- [19] Matsumaru, M., Kawanaka, T., Katagiri, H. and Kaneko, S., Prediction of bankruptcy on industry classification, International Journal of Japan Association for Management Systems, vol. 10, no. 1, pp. 1–12, 2018
- [20] Matsumaru, M., Kawanaka, T., Katagiri, H. and Kaneko, S., Bankruptcy prediction for Japanese corporations using support vector machine, artificial neural network, and multivariate discriminant analysis, International Journal of Industrial Engineering and Operations Management (IJIEOM), Vol.1, No.1, pp.78-96, 2019
- [21]董彦文, 経営と信用のデータ解析, 共立版, 2015
- [22]帝国データバンク, 全国企業倒産集計,  
<https://www.tdb.co.jp/tosan/syukei/20nen.html>(参照2021-12-02)