

機械学習を用いた 倒産判別の研究

神奈川大学 工学部 経営工学科
経営システム工学研究室
山崎祐翔, 山下侑一郎
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.モデル構築

6.分析結果

7.結論・今後の課題

参考文献

1. 研究背景

本研究での倒産の定義

企業経営に行き詰まり，日本取引所グループの上場廃止基準（銀行取引の停止，破産手続・再生手続・更生手続，事業活動の停止）に該当し，株価などのデータが取得できない状態

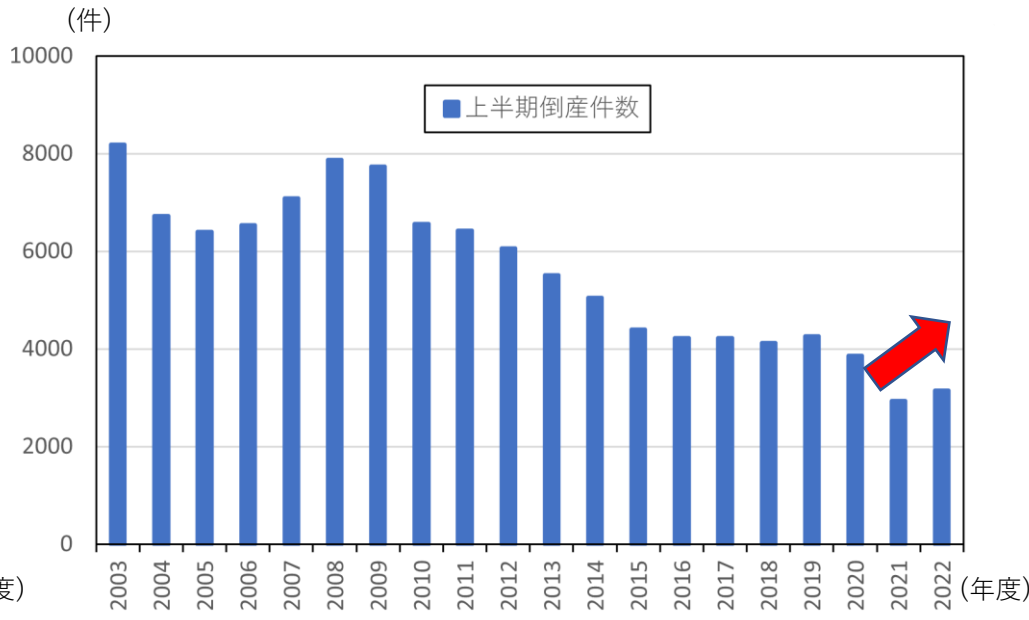
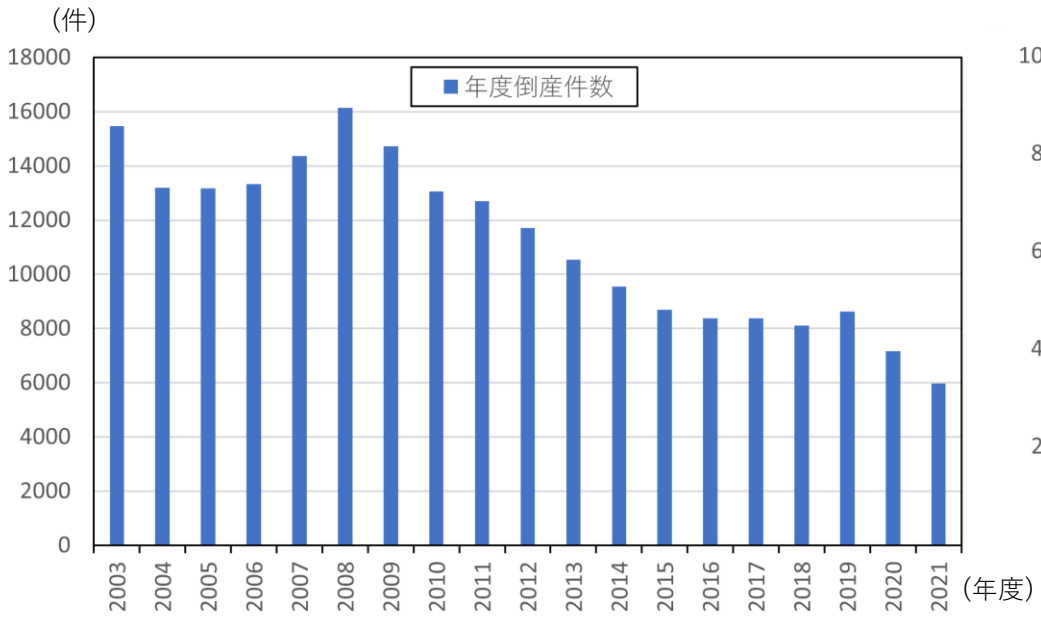
- ・倒産はステークホルダー(株主，金融機関など)に**多大な損失**を与える



1. 研究背景

倒産件数の推移

倒産件数は2008年9月のリーマンショックによる世界同時不況後減少している。しかし、2022年度上半期の倒産件数は上昇している。



出典：東京商工リサーチ, 「全国企業倒産状況」 2022年12月1日時点でのデータ
(データの数値を参考にして資料作成者が独自にグラフを作成)

➡ 倒産を予測して損失を減らす必要がある

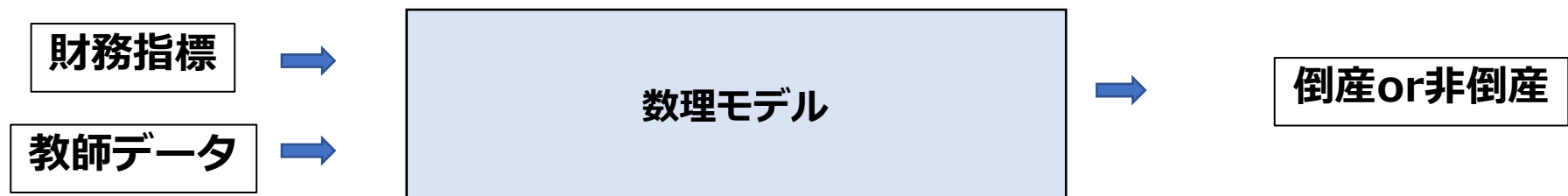
2. 研究目的

研究目的

- ・ 建設業と不動産業の倒産予測モデルを構築する
- ・ ハイパーパラメータの調節や機械学習手法が予測精度に与える影響を考察する
- ・ 各業界の倒産を判別する際に重要となる指標について考察する

倒産予測モデルとは

- ・ 数理モデルを用いて「倒産」or「非倒産」を予測する手法
- ・ 入力は財務指標と教師データ(倒産・非倒産)
- ・ 出力は「倒産」または「非倒産」の予測結果



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1990年から2012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。この時の建設業の指標数は84指標、不動産業の指標は76指標である。

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1990年から2012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。

倒産企業のうち、取得最終年のデータを倒産とし、その前年、前々年のデータは使用せず、それ以外のデータを非倒産として扱った。

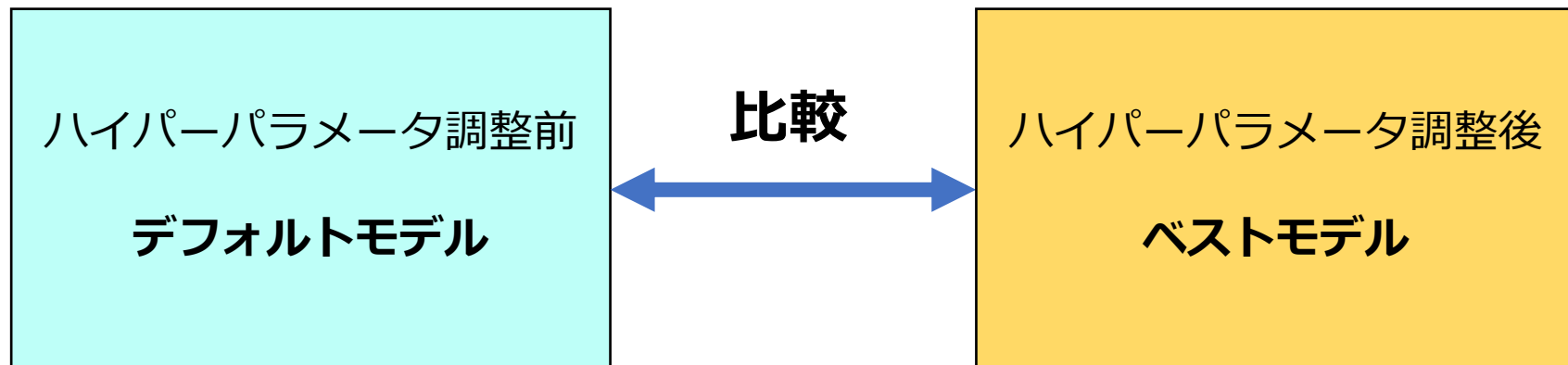
	建設業	不動産業
倒産企業	45	31
非倒産企業	5357	2423
合計	5402	2454

4.分析手法

使用する機械学習手法

勾配ブースティング決定木(GBDT)とSupport Vector Machine (SVM)の2つの機械学習手法を用いて、判別モデルを構築する。その際に、ハイパーパラメータの設定が予測精度にどのような影響を及ぼすのかを確かめる。モデルの構築には**Alkano**を用いる。

本研究では、Alkanoのデフォルトのハイパーパラメータで学習したモデルを**デフォルトモデル**、ハイパーパラメータの探索を行い、精度評価値がベストとなるパラメータで学習したモデルを**ベストモデル**とする。



4.分析手法-勾配ブースティング決定木

「勾配降下法(Gradient)」
「アンサンブル学習(Boosting)」
「決定木(Decision Tree)」
の3つの手法が組み合わされた機械学習手法.



アルゴリズム

- 1.目的変数の平均を求め, 初期予測値とする
- 2.誤差を計算
- 3.決定木を構築
- 4.アンサンブル学習を用いて新たな予測値を算出
- 5.十分な回数2~4を繰り返す
- 6.最終的な予測値を出力する

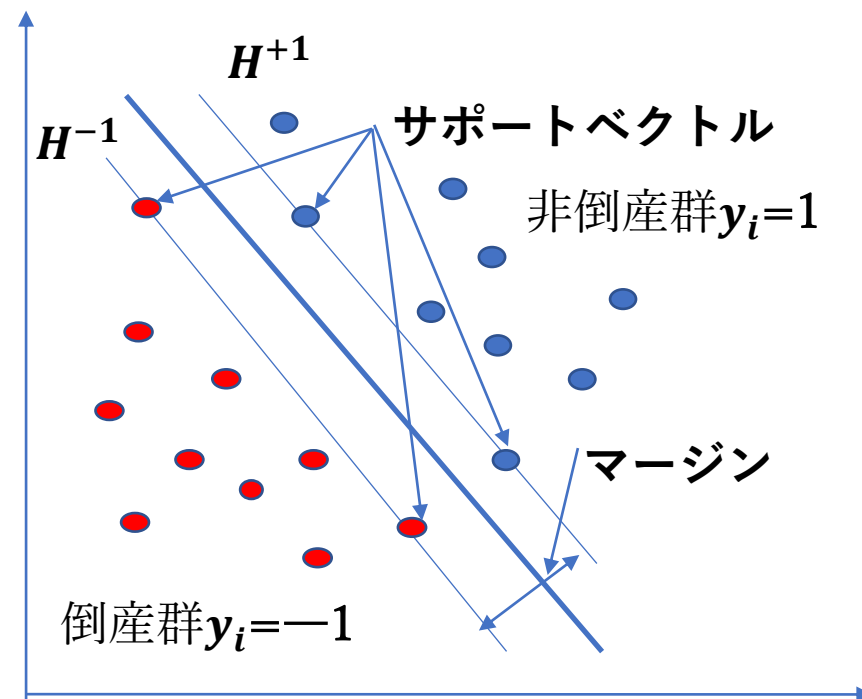
4.分析手法-Support Vector Machine (SVM)

- SVMは線形分離が可能な高次元の仮想空間で、線形的なアプローチで学習を行うシステムである。



- SVMの境界線である $H^{-1}: W^T x + b = -1$ と $H^{+1}: W^T x + b = 1$ の間隔がマージン M である。

- SVMはマージン M を最大化する係数を決定し、判別(予知)を実行する。



4.分析手法-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。

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

2.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。

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

		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

5.モデル構築-デフォルトモデル

5.1 デフォルトモデルの構築 (Alkano)

(1) パラメータの設定

GBDT, SVMいずれも全て
Alkano内の**デフォルト値**を使用する。

The screenshot shows the configuration interface for a Gradient Boosting Decision Tree (GBDT) model. The title is "勾配ブースティング決定木". The interface is divided into several sections:

- 回帰指標:** RMSE (selected)
- イテレーション数 (木の数):** 100
- 学習率:** 0.1
- 正則化:**
 - L1正則化: 0
 - L2正則化: 0
- 決定木:**
 - 葉の枚: 31
 - 葉の最小データ数: 20
 - 最大の深さ: 制限なし
 - 指定する: 1
- 実行:**
 - スレッド数: 自動
 - 指定する: 1
 - 乱数シード: 0 (with a "生成" button)
- サンプリング:**
 - 特徴サンプリング: 1

Buttons for "実行" (Execute) and "保存" (Save) are at the bottom.

The screenshot shows the configuration interface for a Support Vector Machine (SVM) model. The title is "サポートベクターマシン". The interface is divided into several sections:

- SVM:**
 - カーネル関数: ガウシアン (selected)
 - 分散 σ^2 : 自動
 - 指定: 0.1
 - 多項式の次数: 3
 - threshold θ : 0
 - Slack変数の係数: 1
 - 回帰分析の精度: 0.1
- 重み付け:**
 - なし
 - クラス均等化
 - 列指定: 有形固定資産回転率 (selected)

Buttons for "実行" (Execute) and "保存" (Save) are at the bottom.

5.モデル構築-ベストモデル

5.2 分類モデル最適化の利用及び ベストモデルの構築

(1)ハイパーパラメータの設定

GBDT

- ・葉の数($2 \leq x \leq 64$)
- ・特徴サンプリング($0.5 \leq x \leq 1$)
- ・葉の最小データ($20 \leq x \leq 100$)

SVM

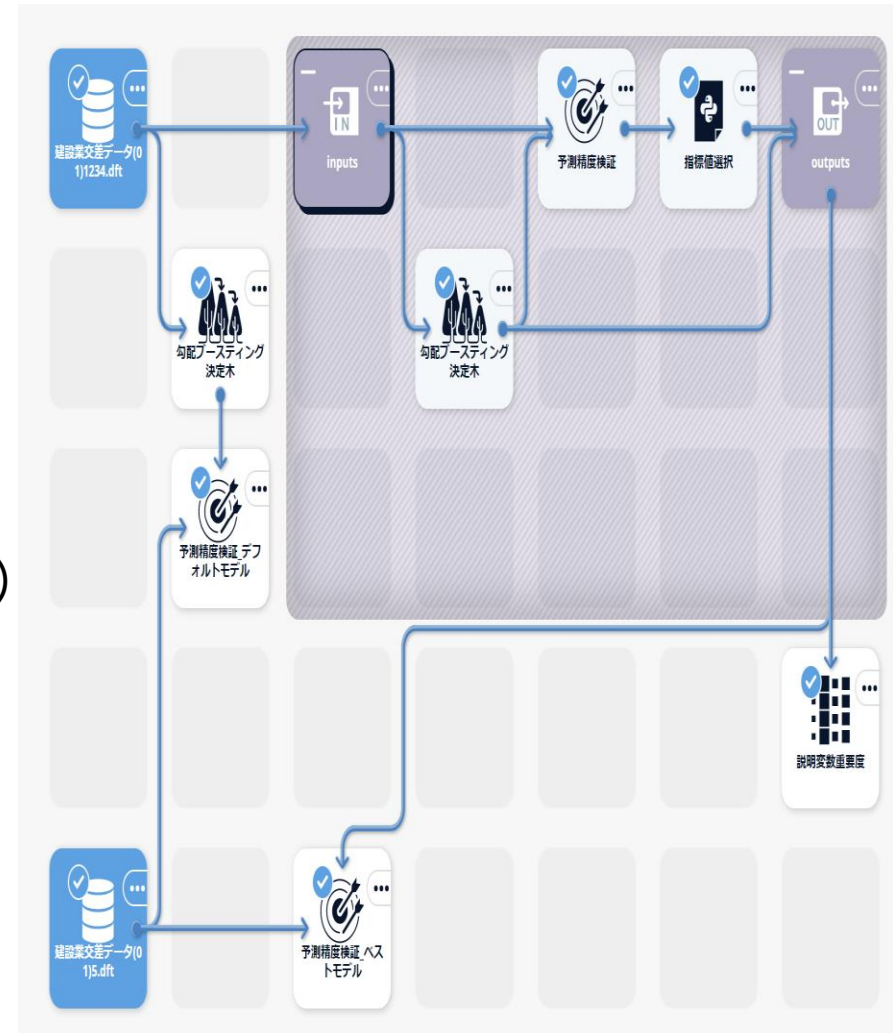
- ・スラック変数の係数($0 < x \leq 3$)
- ・重み付け($x \in ["none", "equalization", "col"]$)

(2)評価値の計算と選択

「分類モデル最適化」モジュール内の
「指標値選択」で、指標値選択を選択する。

最適化対象の評価値：“再現率”

最適化対象ラベル：“倒産”



5.モデル構築-ベストモデル

5.3 最適化の設定と実行

(1)停止条件の設定：7200秒

(2)交差ブロック数：5

(3)層化列：分類

ベイズ最適化により，分類モデルの精度評価値がベストとなるパラメータを探索する。

分類モデル最適化

試行回数
100

タイムアウト [秒]
7200

乱数シード値: 0 シード値生成

詳細設定

ソルバー設定 探索空間設定

検証

検証方法 交差検証

交差対象 table

交差ブロック数 5

グループキー

層化列 分類 × +

実行 保存

5.4 デフォルトモデル及びベストモデルの予測精度の検証

2種類のモデルを用いて，検証データの予測精度を比較する。

6.分析結果-GBDTによる予測精度

「GBDT」を用いて作成したモデルの予測精度の混同行列は以下の通りである。

6.1 建設業

(1)学習データの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	22	0
	非倒産	0	2,678

(2)テストデータに対するデフォルトモデルの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	0	23
	非倒産	3	2,676

(3)テストデータに対するベストモデルの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	1	22
	非倒産	3	2,676

6.分析結果-GBDTによる予測精度

6.2 不動産業

(1)学習データの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	15	0
	非倒産	0	1,211

(2)テストデータに対するデフォルトモデルの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	0	16
	非倒産	1	1,211

(3)テストデータに対するベストモデルの混同行列

84指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	0	16
	非倒産	1	1,211

6. 分析結果-データの形成

6.1 6.2 の結果から, モデルが過学習を起こしており, 汎化性能が低いことが分かった.

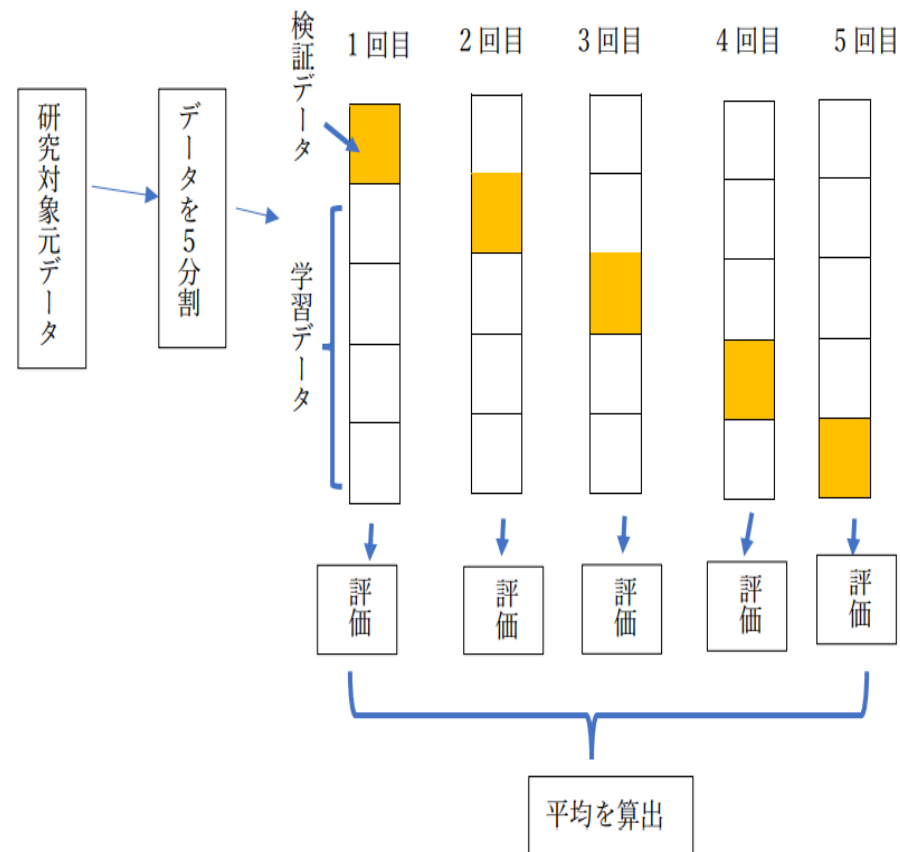
⇒5分割交差検証を行う.

交差検証

「交差検証 (クロスバリデーション)」とは, 機械学習モデルの未知のデータへのモデルの対応能力を評価する一つの指標である.

交差検証では, データの分割を何度も繰り返し行って複数のモデルを訓練し, 複数のモデルの平均値をとって最終的な性能とする.

流れは右の通りである.



6. 分析結果-データの形成

予測精度を高めるため、ハイパーパラメータの探索を多く行わせたい。

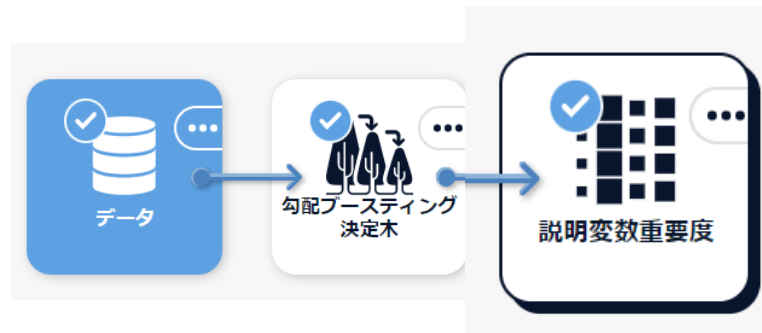
⇒データの比率を「倒産：非倒産 = 1：5」にし、指標数も20指標に絞った。

使用データ

	倒産	非倒産
建設業	45	225
不動産業	31	150

指標数

6.1, 6.2で作成したモデルの**説明変数重要度**から、上位20指標に絞った。



→ハイパーパラメータの探索回数が**28回**から**70回**に増加した。

6. 分析結果-20指標

6.3 建設業 指標詳細

財務指標

- | | |
|---------------------|-------------------|
| ①有形固定資産回転率 | ⑪固定負債対有形固定資産比率(A) |
| ②有形固定資産回転率-取得価格基準 | ⑫固定負債対有形固定資産比率(B) |
| ③売上高販管費率 | ⑬固定比率 |
| ④売上高減価償却費率(B) | ⑭固定長期適合率 |
| ⑤売上高純金利負担率 | ⑮自己資本比率(B) |
| ⑥純利益率 | ⑯負債比率 |
| ⑦売上高事業利益率-利払後 | ⑰有利子負債対自己資本比率 |
| ⑧売上高EBITDA（経常利益ベース） | ⑱純有利子負債対自己資本比率 |
| ⑨労働装備率 | ⑲借入金依存度 |
| ⑩流動比率 | ⑳有形固定資産減価償却累計率 |

6. 分析結果-20指標

6.4 不動産業 指標詳細

財務指標

- | | |
|----------------|--------------------|
| ①有形固定資産減価償却累計率 | ⑪純運転資本額 |
| ②純有利子負債額 | ⑫売上高原価率 |
| ③有利子負債額 | ⑬売上高経常利益率 |
| ④減価償却費 | ⑭固定負債対有形固定資産比率 (B) |
| ⑤自己資本比率 (A) | ⑮有形固定資産回転率 |
| ⑥正味運転資本額 | ⑯売上高販管費率 |
| ⑦借入金依存度 | ⑰流動比率 |
| ⑧支払準備金 | ⑱手元流動性比率 |
| ⑨固定長期適合率 | ⑲純利益率 |
| ⑩自己資本比率 (B) | ⑳負債比率 |

6. 分析結果-形成後データ

6.5 建設業

(1) GBDTの予測精度検証の混同行列

デフォルトモデル

20指標		モデルの予測	
		倒産	非倒産
実際の クラス	倒産	19	26
	非倒産	16	209

ベストモデル

20指標		モデルの予測	
		倒産	非倒産
実際の クラス	倒産	19	26
	非倒産	17	208

(2) SVMの予測精度検証の混同行列

デフォルトモデル

20指標		モデルの予測	
		倒産	非倒産
実際の クラス	倒産	1	44
	非倒産	0	225

ベストモデル

20指標		モデルの予測	
		倒産	非倒産
実際の クラス	倒産	36	9
	非倒産	44	181

6.分析結果-精度評価値

6.6 建設業

全体の結果のまとめを以下に示す.

手法	GDBT	GBDT	SVM	SVM
	(デフォルト)	(ベスト)	(デフォルト)	(ベスト)
正確度	0.8444	0.8407	0.8370	0.8037
再現率	0.4222	0.4222	0.0222	0.8000

GBDTの精度評価値が大きくなる要因

- ・ハイパーパラメータのデフォルト値が適切であること
- ・建設業が行う工事が多様であること（左官, 土木, 電気, 管など）

SVMの精度評価値が大きくなった要因

- ・ハイパーパラメータのデフォルト値が適切でなかったこと

一番精度が
良い

建設業ではベストモデルのSVMの精度が良かった。

6. 分析結果-形成後データ

6.7 不動産業

(1) GBDTの予測精度検証の混同行列

デフォルトモデル

20指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	21	10
	非倒産	2	148

ベストモデル

20指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	23	8
	非倒産	1	149

(2) SVMの予測精度検証の混同行列

デフォルトモデル

20指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	0	31
	非倒産	0	150

ベストモデル

20指標		モデルの予測	
		倒産	非倒産
実際のクラス	倒産	14	17
	非倒産	30	120

6.分析結果-精度評価値

6.8 不動産業

全体の結果のまとめを以下に示す.

手法	GDBT	GDBT	SVM	SVM
	(デフォルト)	(ベスト)	(デフォルト)	(ベスト)
正確度	0.9337	0.9502	0.8287	0.7403
再現率	0.6774	0.7419	0.0000	0.4516

GBDTの精度評価値が大きくなった要因

- ・再現率が高くなるようにハイパーパラメータの値が最適化されたこと

一番精度が
良い

SVMの精度評価値が大きくなった要因

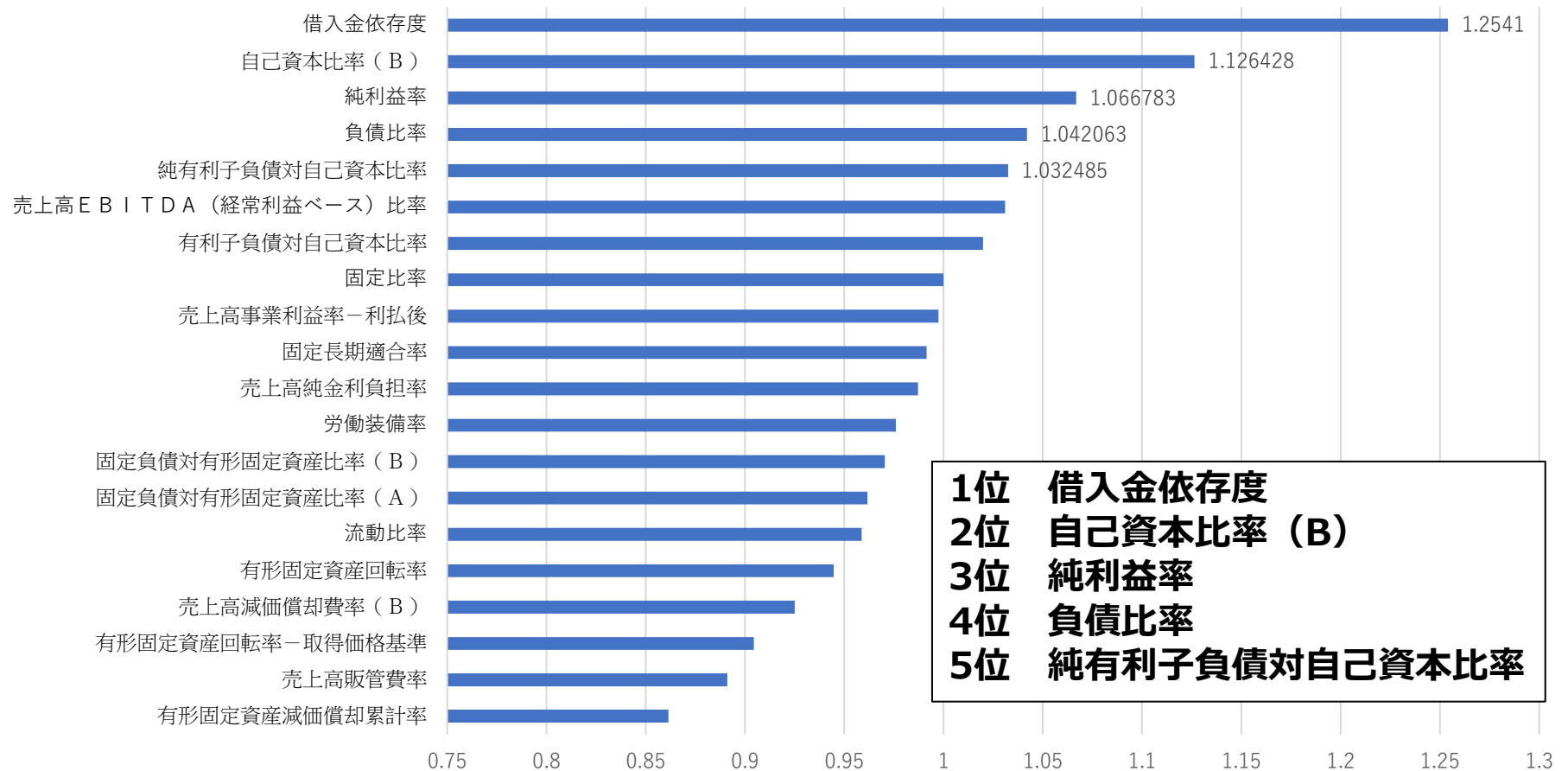
- ・ハイパーパラメータのデフォルト値が適切でなかったこと

不動産業ではベストモデルの勾配ブースティングの精度が良かった.

6.分析結果-説明変数重要度

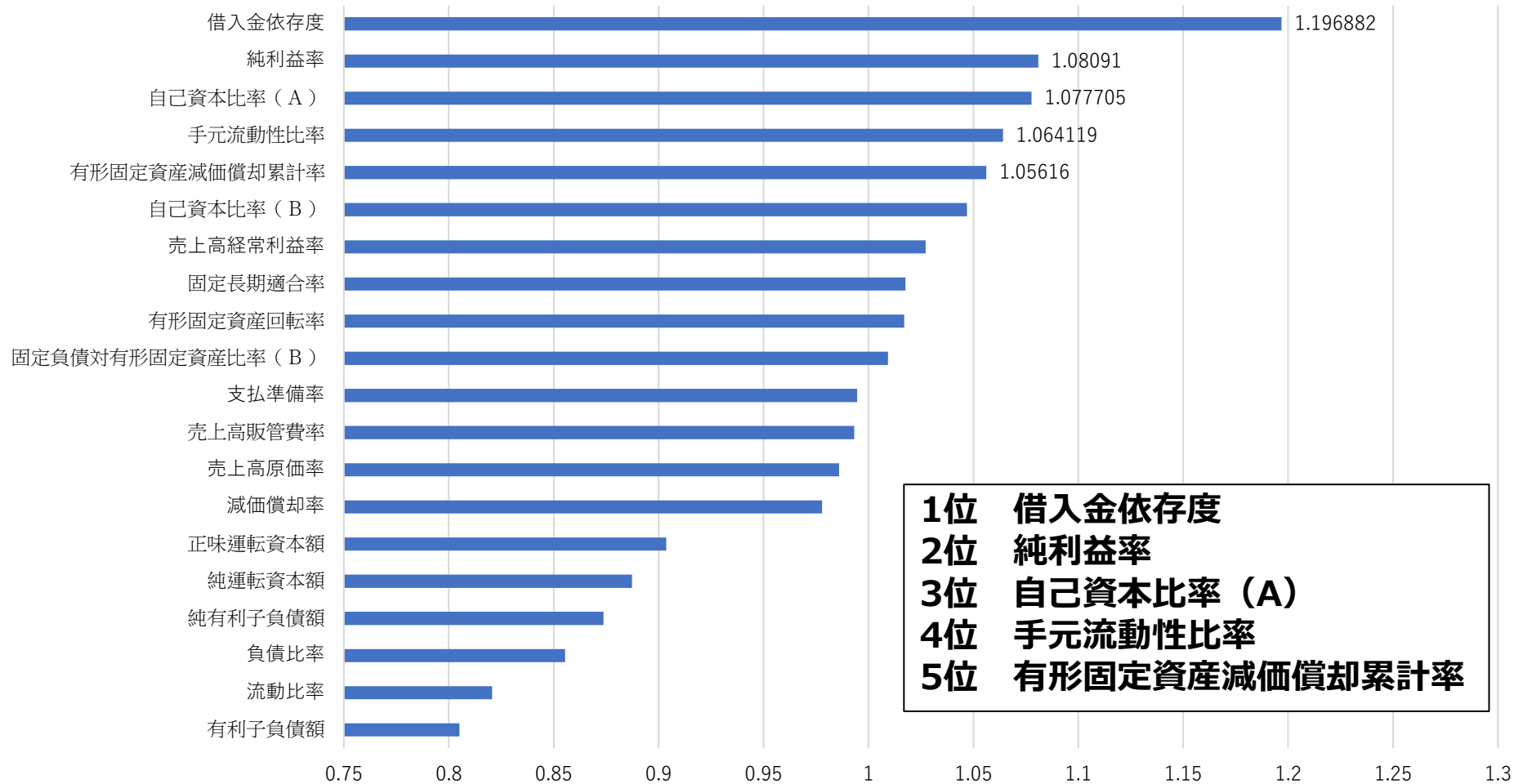
6.9 建設業の説明変数重要度

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。



6.分析結果-説明変数重要度

6.10 不動産業の説明変数重要度



6.分析結果-説明変数重要度上位5指標

6.11 2業界の説明変数重要度上位5指標の比較

2つの業界で重要視している指標に違いがあることが分かる。

建設業

- | | |
|----|---------------|
| 1位 | 借入金依存度 |
| 2位 | 自己資本比率 (B) |
| 3位 | 純利益率 |
| 4位 | 負債比率 |
| 5位 | 純有利子負債對自己資本比率 |

不動産業

- | | |
|----|---------------|
| 1位 | 借入金依存度 |
| 2位 | 純利益率 |
| 3位 | 自己資本比率 (A) |
| 4位 | 手元流動性比率 |
| 5位 | 有形固定資産減価償却累計率 |

共通の指標

借入金依存度, 純利益率, 自己資本比率

業界の特徴指標

建設業 : 負債比率, 純有利子負債對自己資本比率

不動産業 : 手元流動性比率, 有形固定資産減価償却累計額

6.分析結果-説明変数重要度上位5指標

6.12 共通の指標

$$\cdot \text{借入金依存度} = \frac{(\text{純有利子負債額} - \text{従業員預り金})}{\text{負債} \cdot \text{純資産合計}} \times 100$$

総資産に対する借入金の割合を表す指標であり、**返済能力**を示す。

$$\cdot \text{純利益率} = \frac{\text{当期純利益率}}{\text{売上高}} \times 100$$

売上高に対する当期純利の割合を表す指標であり、**収益力**を示す。

$$\cdot \text{自己資本比率} = \frac{\text{自己資本}}{\text{負債} \cdot \text{純資産合計}} \times 100$$

総資産に対する自己資本の割合を表す指標であり、**財務的な健全性**を示す。

→総資本に対する負債が大きく、返済能力もなく、収益力が弱い企業は倒産しやすいため、2業界とも上の3指標が重要となる。

6.分析結果-説明変数重要度上位5指標

6.13 建設業の特徴指標

$$\cdot \text{負債比率} = \frac{\text{負債合計}}{\text{資本合計}} \times 100$$

$$\cdot \text{純有利子負債対自己資本比率} = \frac{\text{純有利子負債額}}{\text{資本合計}} \times 100$$

建設業の特徴

- ・工期が長く、**資金回収**するのに時間がかかる。
- ・資金支出と**資金回収**の時期が異なる。

→資金回収ができ利益が出るまでの**財務耐久力**が重要である。

→負債の**返済余力**を示す負債比率や純有利子負債対自己資本比率が重要となる。

6.分析結果-説明変数重要度上位5指標

6.14 不動産業の特徴指標

- ・ **手元流動性比率** = $\frac{\text{現金預金} + \text{短期有価証券}}{\text{売上高}} \div 12$
- ・ **有形固定資産減価償却累計率** = $\frac{\text{有形固定資産減価償却累計額}}{\text{有形固定資産減価償却累計額} + \text{償却対象資産合計}} \times 100$

不動産業の特徴

- ・ 不動産売買時の現金預金の利用は取引を有利に進めることができる。
 - 不動産売買における他社に対する**競争優位性**を持つことができるため手元流動性比率が重要である。
- ・ 不動産施設が老朽化していると賃貸収益などが低下する。
 - 不動産の**新規性**が重要で絶えず更新していく必要がある。
 - 有形固定資産の老朽度を示す有形固定資産減価償却累計率が重要である。

結論

本研究では、SVM とGBDTの機械学習モデルを用いて建設業270社と不動産業181社と20の財務指標を用いて倒産予測モデルを構築した。

その結果、建設業ではSVMが、不動産業ではGBDTが精度よく倒産予測を行うことができた。

また、倒産予測に強い影響力を与える財務指標は、借入金依存度、純利益率、自己資本比率であった。加えて、建設業界では負債比率と純有利子負債対自己資本比率が影響を与えることが分かった。不動産業界では手元流動性比率、有形固定資産減価償却累計率が影響を与えることが分かった。

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。
- AlkanoにはGBDTやSVM以外にも様々な機械学習モデルが用意されているため、機械学習手法による予知精度の違いについて考察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] 松丸正延, 片桐英樹: 機械学習を用いたサンプルデータ期間の違いによる倒産予知モデル, 経営モデル研究会編, 『経営問題解決のためのモデル構築による「経営モデル」研究』, 経営システム学会誌, Vol.38, No.4 特別号, pp.25-27, 2022
- [2] 株式会社東京商工リサーチ, 全国企業倒産状況,
<https://www.tsr-net.co.jp/news/status/index.html> (参照2022-12-01)
- [3] 董彦文, 経営と信用リスクのデータ科学, 共立出版, 2015
- [4] AIZINE, 機械学習モデルの性能評価入門! 交差検証(クロスバリデーション) とは,
<https://aizine.ai/glossary-crossvalidation/> (参照2022-12-01)
- [5] 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
- [6] Friedman, J.H., Greedy function approximation A gradient boosting machine, The Annals of Statistics, Vol.29, No.5, pp.1189-1232, 2001
- [7] Cortes, C., and Vapnik, V., Support-vector networks, Machine Learning, vol. 20, no. 3, pp. 273-297, 1995

- [8] Vapnik, V., in Haykin, S. (Ed.), *Statistical Learning Theory, Adaptive and Learning Systems*, Vol. 736, John Wiley & Sons Inc, 1998
- [9] 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
- [10] 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
- [11] 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
- [12] 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

- [13] 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
- [14] 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
- [15] 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