

Department of Policy and Planning Sciences

Discussion Paper Series

No.1383

財務諸表データを用いた資金ニーズの

見過ごしチェック AI の開発

(Development of AI to Check for Missed Financial Needs

Using Financial Statement Data)

by

樽 涼稀, 秦 涼太, 今倉 暁, 櫻井鉄也, 岡田幸彦

(Ryoki MOTAI, Ryota SHIN, Akira IMAKURA, Tetsuya SAKURAI and Yukihiro OKADA)

December 29, 2022

UNIVERSITY OF TSUKUBA

Tsukuba, Ibaraki 305-8573
JAPAN

財務諸表データを用いた資金ニーズの見過ごしチェック AI の開発

Development of AI to Check for Missed Financial Needs Using Financial Statement Data

筑波大学大学院サービス工学学位プログラム 罇 涼稀

筑波大学大学院社会工学学位プログラム 秦 涼太

筑波大学システム情報系／人工知能科学センター 今倉 暁

筑波大学システム情報系／人工知能科学センター 櫻井鉄也

筑波大学システム情報系／人工知能科学センター 岡田幸彦

〈論文要旨〉

地域金融機関では、機械学習を活用して、取引先の資金ニーズに適切に対応することが求められている。本研究では、一般的な財務諸表を用いて、長期借入金の増加を判別する方法論を提示し（TRL2）、その技術評価を行った（TRL3）。技術評価は、最も原始的なロジスティック回帰モデルをベースモデルとし、東証一部上場企業の財務諸表を用いて、判別性をテストした。売上高の増減によってデータを分割して、特異値分解（SVD）を行い、目的変数との相関に基づく特徴量選択を行った後に学習する我々の方法論では、どの金融機関でも実行可能な手法として、テストデータにおいて平均的に AUC 0.8 を上回り、80%以上の正解率を実現できることが実験的に示された。今後は、地域金融機関の実データを用いた本格的なラボ実験（TRL4）と、地域金融機関の現場に即した AI サービスの実装実験（TRL5～）が求められる。

〈キーワード〉

地域金融機関, 長期借入金, 機械学習, ビジネスインテリジェンス, アカウンティング・インフォマティクス

1 はじめに

会計学においても機械学習/AIを応用する重要性が指摘されて久しい。会計学分野における機械学習/AIの研究と実践は、不正会計の検知にかかる監査の文脈が主流の1つとなってきた(市原・首藤 2017)。そして、会計学研究においては、機械学習を応用した新たな研究スタイルが、これまでの統計的因果推論を基礎とした実証分析を越えて、証券市場、銀行融資、監査といった会計学の様々な分析対象において、今後ますます登場するであろうことが示唆されている(首藤 2019)。さらに、この分野の最先端を行くアメリカでは、会計業務や監査業務の文脈での機械学習/AIの利活用のあり方が検討されたり(Marshall and Lambert 2018)、そこでの人材育成方法論の提案がなされたりするに至っている(Elliot et al. 2020)。また、株価予想のために会計情報やテキスト情報を用いた機械学習の応用研究(Henry 2006 ほか)や、アニュアルレポートやアナリストレポートなどのテキストデータを用いた機械学習の応用研究(Huang et al. 2014 ほか)などもなされてきた。

会計学分野において、会計データと機械学習を用いた近年の主要な研究として、Chen et al. (2022)が挙げられる。Chen et al. (2022)は、アメリカ上場企業の財務諸表データを用いて、将来の利益変化の方向を予測するタスクに取り組んだ。そして、Chenらは、構築したモデルの予測に基づくヘッジポートフォリオのリターンは、伝統的な統計モデルに基づくもの、およびプロのアナリストの予測に基づくものよりも大きいことを示した。他にも、仕訳データに機械学習技術を応用し、監査業務の効率化を図った研究も報告されている(Liang et al. 2021; Müller et al. 2022)。

管理会計は、ビジネスインテリジェンス(以下、BI)を取り入れることで、その意思決定有用性が向上する(Rikhardsson and Yigitbasioglu 2018)。そして、BIは、企業内外の情報を集約することで組織の意思決定を支援し、競争優位を導く(Božič and Dimovski 2019)。BIの基盤技術研究は豊富であり、仕訳データの可視化によって監査の質を高めることを提案した研究(Guo et al. 2022)や、契約書から原価を抽出する手法を提案した研究(Beaulieu 2020)等が蓄積されている。また、BIシステムの導入から成功までの各過程に関する研究も盛んに行われている(Ain et al. 2019)。

このように、会計学分野における機械学習研究、ならびにデータの利活用に関する研究は活発化しているが、金融機関をユーザーとして想定し、取引先の財務諸表を用いて機械学習を応用する基盤技術に関する研究は、ほとんどなされてこなかった。この点について、首藤(2019, 135)は、「銀行は様々な形でAIを活用していると考えられるが、関連する学術論文はほとんど存在しなかった」と述べている。伝統的にこの分野は、線形回帰モデルを用いることが多い計量経済学のアプローチをとることが多い。あるいは、非線形性を加味して確率モデリングを行うオペレーションズ・リサーチ/金融工学的アプローチによって、信用リス

クを推定したり、クレジット・スコアリングモデルを開発したりしてきた。そしてこれらの研究は、必ずしも財務諸表の会計数値そのものを必要とはせず、機械学習を応用する視点も欠けていた。

一方わが国では、地方銀行や信用金庫を含む地域金融機関の経営悪化の問題が社会課題として認識され、収益力向上や新たな収益源の開拓が強く求められるようになった。この点について、金融庁の「中小・地域金融機関向けの総合的な監督指針」(令和4年8月)では、「中小・地域金融機関は、地方経済を取り巻く厳しい環境の下、中小企業や地域経済から期待される役割を果たすため、取引先や地域への過大なコミットメントコストを負担することにより、かえって収益力や健全性の低下といった状況を招いている場合がある。このように、中小・地域金融機関の実態は、リレーションシップバンキング本来のあり方から乖離している面があり、リレーションシップバンキングの機能強化を図り、地域の金融ニーズへの一層適切な対応や、持続可能性(サステナビリティ)の確保を図る必要があると考えられる」とされた。そして、現実の金融検査の際には、収益力強化と健全性向上についてのコメントが、金融庁から地域金融機関に対してなされることが強化された趨勢がある。

そこで本研究では、どの地域金融機関でも電子的に蓄積している取引先の財務諸表の会計数値を用いて、わが国に現存するどの地域金融機関でも実行可能な機械学習の応用方法論を検討したい。その際の焦点は、取引先の資金ニーズを財務諸表から察知し、取引先の経営を助ける行為を支援することである。本研究で示した結果は、地域金融機関における営業活動、そして、取引先における社内での意思決定に貢献し、ビジネスインテリジェンスの基盤技術となる可能性がある。

以上の事柄を踏まえ、本研究の目的は、上記の支援情報の作成までを、誰もが追試可能な東証一部上場企業の財務諸表データで技術評価すること(TRL2, TRL3)とする。本研究の技術評価から、本研究が想定した「 $t-1$ 年度と t 年度の財務諸表を用いて、 $t+1$ 年度中の長期借入金の増減を分類する」という機械学習上のタスクでは、80%以上の正解率をもたらす方法論を提示することができた。今後は、実用を想定した本格的なラボ実験(TRL4)を進め、実データを用いた現実の文脈での技術評価(TRL5)が求められる。

2 実験デザインとデータ

2.1 本研究の方法論

上で引用した金融庁の指摘のように、地域金融機関は、地方経済を取り巻く厳しい環境の下、中小企業や地域経済から期待される役割を果たすことが求められている。しかしながら、地域金融機関は、取引先や地域への過大なコミットメントコストを負担することにより、収益力や健全性が低下する状況がしばしば生じている。そこで本研究は、リレーションシップ

バンキングの機能強化を図り、地域の金融ニーズへの一層適切な対応や持続可能性の確保を支援することを目指す。具体的には、どの地域金融機関でも有する取引先の財務諸表を活かし、地域の資金ニーズを費用対効果良く察知するための方法論を開発する。

本研究では、水戸信用金庫へのヒアリングと意見交換を基礎として、上記の目的を策定するとともに、機械学習上の目的変数に「取引先の長期借入金の増加」を置くこととした。取引先の長期借入金は、地域金融機関にとっては収益力の源泉である。そして、リレーションシップバンキングとしては、いかに早く正確に取引先の資金ニーズを察知し、地域のために寄り添った将来の事業計画・返済計画の合意形成を行えるかが重要である。ただし、ある取引先は、取引する金融機関が多様である。そして、いつ、どのような目的と条件で、どの金融機関と債務契約をしたのかに関する詳細な情報は、ある地域金融機関からすると即座に正確にはわからない。ある地域金融機関として最も早く入手できる確実な情報は、取引先から地域金融機関に報告される決算情報である。

わが国では一般的に、取引先は決算日の3か月から4か月後に最新年度の財務諸表を金融機関に報告することが慣習となっている。この取引先の財務諸表は、地域金融機関内のデータベースに蓄積され、多くの金融機関では、報告された財務諸表データを用いて取引先の財務分析を行い、その結果を取引先へフィードバックしている。つまり、わが国地域金融機関の実情と慣習を鑑み、取引先の財務諸表データが入手されて即座に、機械学習によって取引先の資金ニーズを分析し、営業支援情報として活かす方法論と組織ルーチンの開発が重要となる。しかしながら、先の首藤（2019）の引用にもあるとおり、この挑戦的な取り組みの基盤となる技術評価研究はほとんどなされてこなかった。

そこで本研究では、そもそも、取引先の財務諸表（貸借対照表、損益計算書、キャッシュフロー計算書）が、地域金融機関における取引先の長期借入金の増加の判別に資するのかを技術評価する必要があると考えた。その際、どの地域金融機関でも応用可能な方法論で技術評価をすることが重要となる。そのため本研究は、水戸信用金庫とのヒアリングや意見交換をふまえ、過去3年分の財務諸表を用いて、「 $t-1$ 年度と t 年度の財務諸表から得られる情報から、 $t+1$ 年度中の長期借入金の増加を判別する」という機械学習上のタスクを設定することとした。したがって、「 $t+1$ 年度の長期借入金/ t 年度の長期借入金」の値が1以上の企業については目的変数を1として、それ以外の企業については目的変数を0とした。

なお、本研究は嚆矢的な技術実験であるため、もっとも単純な機械学習手法の1つであるロジスティック回帰モデルを用い、回帰問題ではなく分類問題として「 $t+1$ 年度中に長期借入金が増加した」企業を判別する実験デザインとした。学習したモデルの性能評価には、分類問題の性能尺度として一般的なACC（Accuracy：正解率）とAUC（Area Under the Curve）を用いることとした。

ここで、ACC と AUC の説明を加える。まず、モデルを学習した後に、1 社以上の財務データをモデルに入力した状況を考える。すると、入力した企業数分の予測値がモデルから出力される。ただし、本研究ではロジスティック回帰モデルを使用するため、予測値は 0 から 1 の小数值を取る。次に、予測値を 0、または 1 の予測結果に変換することを考える。このとき、ある閾値未満の企業の目的変数を 0 とし、それ以外の企業の目的変数を 1 とする。本研究では、この閾値の算出に Youden index (Youden 1950) を用いた。これより、ACC は、「予測結果と目的変数が等しい企業数 / モデルに入力した企業数」で求めることができる。一方、AUC は、ROC 曲線より下部の面積で定義される。これは、閾値の恣意的な設定に影響を受けない指標 (Mandrekar 2010) であり、二値分類問題の評価指標として広く利用されている (Lobo et al. 2010)。ROC 曲線は、可能なすべての閾値に対して、偽陽性率と真陽性率を算出し、それらの点を結ぶことで得られる。

最後に、本項で注意すべき点は、本研究で用いるデータについてである。現実の地域金融機関が保有・蓄積する取引先の財務諸表の電子データは機密情報であるため、営利目的で第三者に渡すためには、地域金融機関内での厳重な審査と承認にもとづく契約が必要となる。そして、そのデータを用いた実験の追試も困難である。そこで本研究では、現実の地域金融機関が保有・蓄積する取引先の財務諸表データを用いたラボ実験 (TRL4) は、現状では不確実性が高いため行わず、誰もが本研究の追試をできる NEEDS 財務データに収録された東証一部上場企業の財務諸表データによって、コンセプトの提示 (TRL2) と実証 (TRL3) にあたる技術実験と評価を行うこととした。

2.2 3 段階の実験デザイン

本研究は先行研究がほとんどない中での嚆矢的な技術実験・評価であるため、3 段階からなる実験デザインを設計した。実験 1 は、あるべき方法論の探索実験である。実験 2 は、実験 1 の結果をふまえた方法論の詳細比較実験である。実験 3 は、実験 1 と実験 2 の結果をふまえた、本研究が提案する方法論にかかる実験である。

実験 1 のあるべき方法論の探索実験は、財務諸表データの取り扱いに関連している。財務諸表上に表示される会計情報は、例えば流動資産合計の金額の中に現金預金のような個別の勘定科目の金額が含まれるように、相関のある入れ子構造のデータとなっている。そのため、流動資産合計と現金預金のそれぞれをそのまま機械学習の入力として学習させてしまってもよいのかの懸念が生じる。むしろ、因子分析 (以下、FA)、主成分分析 (以下、PCA)、独立成分分析 (以下、ICA)、特異値分解 (以下、SVD) といった教師なし学習に属する次元削減手法を用い、同様の変量を束ねて新たな変量を生成し、それらを入力として学習させた方が、判別性能が高まる可能性がある。そしてその場合、どの程度次元削減を行うべきなのかにつ

いても検討が必要となる。

加えて、そもそも財務諸表に表示される会計情報は、企業単位で規模感が異なる。総資産 1 兆円スケールの企業もあれば、10 億円スケールの企業もあるのである。そのため、財務諸表上で表示された会計数値を素朴にそのまま機械学習の入力とするべきなのか、それとも、標準化や正規化によって、事前にデータをスケールした上で機械学習の入力とした方がよいのかを検討する必要がある。これは、上述の次元削減手法を適用する際にも同様の問題となる。

さらに、入力候補となった変数を全て入力すべきなのか、という特徴量選択の問題も検討する必要がある。本研究では「 $t+1$ 年度中の長期借入金増加」を目的変数とすることは先に述べたが、これを機械学習の出力として学習させるにあたり、事前にこの目的変数との相関が強い変数を絞って入力とした方が、判別性能が高まるかについても調べる必要がある。

実験 1 では、以上の懸念点を鑑み、(1) データのスケールリングの方法（標準化、正規化）、(2) 次元削減処理の方法（FA, PCA, ICA, SVD）とその削減度合い、(3) 目的変数との相関による特徴量選択の度合い、という 3 点について順番に場合分けを行い、それら全ての組み合わせについて判別性能の評価を行う。実験 1 では、全ての場合分けを行った後に、結果の頑健性を高める目的の下で層化 20 分割交差検証を行い、ACC と AUC を 20 回の平均値で算出する。そして、テストデータにおいて、ACC と AUC がいずれも高い方法論のパターンを見つけ出すことを目指す。

実験 2 で行う方法論の詳細比較実験は、実験 1 で最も高い AUC を示した方法論を用いて、その方法論の判別性能の善し悪しを決める機械学習上の性質について丁寧に調べる。具体的には、データスケールリング処理の効果、次元削減処理としての SVD の効果、そして目的変数との相関による特徴量選択の度合いの効果について、比較検証を行う。実験 2 についても、層化 20 分割交差検証を行い、ACC と AUC を 20 回の平均値で算出し、ACC と AUC が高まる機械学習上の性質を見つけ出すことを目指す。この実験 2 により、財務諸表の会計数値を機械学習で用いる際の注意点や望ましい方向性が明らかになると考えられる。

実験 3 の本研究が提案する方法論にかかる実験は、実験 1 と実験 2 の結果をふまえ、さらに、地域金融機関での利用を想定して行った水戸信用金庫との意見交換から、 $t-1$ 年度から t 年度にかけて売上が増加した企業と減少した企業とを分け、それぞれのデータで学習する方法論の技術評価を行う。実験 3 についても、層化 20 分割交差検証を行い、ACC と AUC を 20 回の平均値で算出し、我々の提案方法がどの程度、判別性能を高めるのか検証する。

2.3 データ

本研究で使用するデータは、NEEDS 財務データに採録された 2005 年 1 月から 2017 年 9 月

までを決算月とする東証一部上場企業の財務諸表データ（貸借対照表・損益計算書・キャッシュフロー計算書）である。なお、本研究では、決算月が t 年 4 月から $t+1$ 年 3 月までのデータを、 t 年度のデータと括った。取得したデータの総サンプル数は 24,471 であり、財務諸表関連の抽出可能な勘定科目数は 425 であった。

本研究で扱う特徴量の種類とその個数を表 1 に示す。モデルの入力となりうる特徴量候補の変量は、値、値の増減、総資産比、総資産比の増減を基本とした。これらはどの地域金融機関でも容易に作成できるものである。ここで、表 1 に示した「 $t-1$ 年度の値」と「 t 年度の値」とは、各年度末時点における各勘定科目の値であり、それらの値の増減（(t 年度の値) - ($t-1$ 年度の値)）も作成した。次に、「 $t-1$ 年度の総資産比」は、($t-1$ 年度の値 / $t-1$ 年度の総資産額) で算出しており、「 t 年度の総資産比」は、(t 年度の値 / t 年度の総資産) で算出した。そして、それらの総資産比の増減（(t 年度の総資産比) - ($t-1$ 年度の総資産比)）も作成した。最後に、本研究では、 $t-1$ 年度から t 年度の売上高の増加倍率（ t 年度の売上高 / $t-1$ 年度の売上高）、そして年度ダミーの計 12 変量も作成した。これらの特徴量候補数の総計は 2,559 であった。

表 1 予測モデルに使用する特徴量候補数

種類	個数
$t-1$ 年度の値	425
t 年度の値	425
t 年度の値と $t-1$ 年度の値との差	425
t 年度の売上高 / $t-1$ 年度の売上高	1
$t-1$ 年度の総資産比	424
t 年度の総資産比	424
t 年度の総資産比と $t-1$ 年度の総資産比との差	424
年度ダミー	11

こうして生成された基礎データに対して、以下の前処理を行って本研究の技術実験用のデータセットを整形した。これにより、本研究の 3 段階からなる技術実験用の総サンプルは 12,988、特徴量候補は 421（年度ダミー含む）となった。なお、ここでいう総サンプル数の内訳は、基準年を t 年度にした場合、 $t-1$ 年度から $t+1$ 年度までの財務諸表データが存在する企業のみで構成されている。

1. 3 年分のデータを持たない企業を削除
2. 長期借入金の値が欠損している企業を削除
3. 貸借対照表、損益計算書、キャッシュフロー計算書とは関係ない特徴量を削除

4. 分散が 0 の特徴量を削除
5. 欠損率 10% 以上の特徴量を削除
6. 欠損率 10% 以上の企業を削除
7. 「 $t+1$ 年度の長期借入金/ t 年度の長期借入金」が 1 以上の企業を 1, 1 未満の企業を 0 とする目的変数を作成
8. 2017 年 4 月～9 月（2017 年度）のデータはサンプル数が 85 と少数であったため削除

次に、技術実験用データの基準年度ごとのサンプル数、および目的変数が 1 である企業の割合を表 2 に示す。表 2 より、長期借入金が増加した企業の割合は、34.9% から 53.2% までの範囲に分散しており、全期間にわたるその割合は 45% となっている。

表 2 技術実験用データの概要

基準年度 t	企業数	$t+1$ 年度中の長期借入金増加企業の割合
2005	983	0.429
2006	1,151	0.456
2007	1,156	0.532
2008	1,163	0.453
2009	1,175	0.349
2010	1,177	0.425
2011	1,212	0.444
2012	1,224	0.498
2013	1,229	0.483
2014	1,258	0.440
2015	1,260	0.441
全期間	12,988	0.450

3 実験結果

3.1 実験 1：あるべき方法論の探索実験

実験 1 では、まず基準年度を 2015 年度とし、 $t-1$ 年度（つまり、2014 年度）と t 年度（つまり、2015 年度）で作成した変数を入力、そして $t+1$ 年度（つまり、2016 年度）中の長期借入金の増加を出力とする機械学習モデルによって、あるべき方法論を探索する。基準年度を 2015 年度としたとき、年度ダミーは分類に資する情報を持たないため削除する。そのため、特徴量候補は 410 変量となる。

実験 1 のための予備実験として、この 410 変量に対して、何も処理を加えず、ロジスティ

ック回帰モデルで学習させた。その結果得られたテストデータに対する判別性能は、ACC については 0.636, AUC については 0.610 であった。ここで、表 2 に示したとおり、基準年度が 2015 年度のとき、目的変数が 1 である企業の割合は 44.1% である。したがって、すべての企業に対して長期借入金が増加する（目的変数=1）と予想した場合に期待される ACC は 0.559 となる。つまり、この時点で、ACC は期待値よりも 7.7% ポイント高い。

この予備実験の結果を前提として、実験 1 では、(1) データのスケーリングの方法（なし、標準化、正規化）、(2) 次元削減処理の方法（なし、FA、PCA、ICA、SVD）とその圧縮程度（ n 次元に圧縮、 $n=410, 400, 300, 200, 100, 50, 10$ ）、そして (3) 目的変数との相関による特徴量選択の度合い（ m 変量を選択、 $m=410, 400, 300, 200, 100, 50, 10$ ）という 3 点について順番に場合分けを行い、全ての組み合わせについて ACC と AUC の評価を行った。AUC で見た時の判別性能が高い機械学習モデルの上位 10 パターンは、表 3 のとおりである。

表 3 AUC が高い上位 10 パターン ($t=2015$)

スケーリング (1)	次元削減 手法 (2)	特徴量		Train		Test	
		次元削減 (2)	相関 (3)	ACC	AUC	ACC	AUC
正規化	SVD	410	200	0.779	0.861	0.722	0.747
正規化	SVD	400	200	0.775	0.860	0.722	0.749
なし	SVD	410	200	0.780	0.860	0.734	0.750
正規化	SVD	400	100	0.739	0.820	0.738	0.751
なし	SVD	410	100	0.741	0.817	0.728	0.751
正規化	SVD	410	100	0.742	0.820	0.737	0.753
なし	SVD	400	100	0.741	0.817	0.733	0.753
標準化	SVD	400	100	0.737	0.815	0.720	0.754
なし	SVD	400	200	0.781	0.859	0.734	0.754
標準化	SVD	410	100	0.747	0.818	0.722	0.755

表 3 に示された上位 10 パターンを見ると、SVD を用いるが次元をほとんど削減しない場合に、テストデータにおける判別性能 (ACC と AUC) が相対的に高いことがわかる。加えて、SVD によって生成された新たな変量について、それら全てを特徴量として機械学習の入力とするのではなく、目的変数に対して相関がある変量の特徴量選択した後に入力とした方が、判別性能が高まい傾向にあった。一方、データのスケーリング処理については、何もしない場合、標準化、正規化のそれぞれによって結果に大きな差異は確認されなかった。

表 3 のとおり、基準年度を 2015 年度とした場合、AUC が最も高くなる組合せは、データを標準化し、SVD で 410 変量を新たに生成し、それらの中で目的変数との相関が高い 100 変

量を選択したモデルであった。このとき、ACCは0.722、AUCは0.755であり、ACCについては期待値よりも16.3%ポイントのACCの向上が実現されている。ただし、この結果がこの基準年だけの結果の可能性は否定できない。そこで、頑健性テストとして、基準年度を2005年度から2015年度へと推移させ、各基準年度のデータセットで学習と予測を行い、ACCとAUCの計算を行った結果が表4である。

表4 頑健性の検証結果

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.782	0.858	0.755	0.783
2006	0.766	0.845	0.753	0.784
2007	0.737	0.817	0.717	0.747
2008	0.756	0.828	0.727	0.756
2009	0.737	0.808	0.716	0.734
2010	0.760	0.837	0.740	0.783
2011	0.759	0.832	0.743	0.766
2012	0.741	0.821	0.718	0.749
2013	0.728	0.805	0.710	0.732
2014	0.734	0.808	0.733	0.742
2015	0.747	0.818	0.722	0.755
全期間	0.634	0.680	0.637	0.668

表4の実験結果から、実験1で探索して最良であった方法論を適用して得られたモデルが、どの基準年度においても高い判別性能を示したことがわかる。全ての基準年度において、テストデータのACCは71%以上の正解率を示し、テストデータのAUCは0.732以上と、相対的に高い判別性能を示している。ここで注意すべきは、年度を区切らず全期間のデータで学習した結果である。このモデルでは、テストデータのACCが0.637、AUCが0.668であり、基準年度で分割した場合の結果と比較して、判別性能が劣化している。表2の全期間を想定した目的変数の分布より、 $t+1$ 年度に長期借入金が増加した企業は45%である。したがって、全ての企業について、目的変数=0と予測した場合に期待されるACCは0.550であるため、全期間のデータで学習した場合は、8.3%ポイントの精度向上がなされている。その一方で、ある基準年度に特化したモデルの判別性能と比較すると、約10%ポイント劣るのである。このことは、財務諸表上の会計数値が有する情報として、その年前後のマクロ経済的趨勢に従って重要な特徴量に変化することを意味していると考えられる。

3.2 実験 2：方法論の詳細比較実験

実験 1 から、AUC が最も高くなる組合せは、データを標準化し、SVD で 410 変量を新たに生成し、それらの中で目的変数との相関が高い 100 変量を選択したモデルであることがわかった。実験 2 では、これらの設定それぞれの相違によって判別性能にどの程度の影響が出るのかを詳細に比較する。まず比較するのは、SVD の効果である。表 5 は、生データのスケール処理（処理なし、標準化、正規化）のそれぞれについて、変量そのままを入力としたモデルと、SVD によって新たに生成した変量を入力としたモデルとを比較している。それ以外の条件や設定、そして用いたデータは同一である。ただし、SVD を用いているものの、次元削減は行っていない。

表 5 より、スケール処理方法に依存せず、SVD によって同様の変量をまとめて新たに変量を生成した方が、明らかにテストデータに対する ACC と AUC が高まることが確認された。一方、スケール処理の方法の違いが判別性能に及ぼす効果は見られなかった。

表 5 SVD の効果（ $t=2015$ 、目的変数に対する相関上位 100 変量の特徴量選択）

スケール処理	次元削減手法	Train		Test	
		ACC	AUC	ACC	AUC
標準化	SVD	0.747	0.818	0.722	0.755
標準化	なし	0.687	0.749	0.674	0.662
正規化	SVD	0.742	0.820	0.737	0.753
正規化	なし	0.664	0.715	0.678	0.670
なし	SVD	0.741	0.817	0.728	0.751
なし	なし	0.627	0.647	0.643	0.612

表 6 は、次元削減の割合によって判別性能がどのように推移するかを示したものである。ここでは、比較可能性の担保のために、SVD による次元削減ではなく、目的変数との相関が強い変量の特徴量として選択する方法によった比較を行った。表 6 の相関列は、目的変数に対して相関の高い m 個の変量を選択したことを意味する。つまり、表 6 では、相関の高い 100 変量を選択した場合、相関の高い 200 変量を選択した場合、およびすべての 410 変量を使用した場合、という 3 つのパターンの結果を表示している。

表 6 から、SVD の有無によらず、全ての特徴量を用いると、モデルが学習データに過適合してしまい（学習データに対する ACC と AUC が相対的に高くなり）、判別性能が劣化する（テストデータに対する ACC と AUC が相対的に低くなる）という重要な傾向が見られる。一方で、興味深いことに、SVD では、相関の高い 100 変量と 200 変量のそれぞれのモデルで、判別性能に大きな差は見られなかった。これらの傾向は、表 6 に示した 6 つのパターンそれ

それをさらに詳細に各年で比較した参考 1～参考 6（文末に補足資料として掲載）においても同様であり，どの年でも頑健に現れる傾向であった。

表 6 目的変数との相関による次元削減の効果（ $t=2015$ ，標準化）

次元削減手法	相関	Train		Test	
		ACC	AUC	ACC	AUC
SVD	100	0.747	0.818	0.722	0.755
SVD	200	0.782	0.857	0.728	0.746
SVD	410	0.786	0.870	0.615	0.579
なし	100	0.687	0.749	0.674	0.662
なし	200	0.726	0.791	0.653	0.647
なし	410	0.747	0.822	0.661	0.628

そして，参考 1～参考 6 の情報を平均値と標準偏差の観点から集計したものが，表 7 と表 8 である。表 2 の全期間の目的変数の分布から，目的変数となる $t+1$ 年度中に長期借入金が増加した企業は 45% であるため，ACC の期待値は 0.550 程度とみなすことができる。この期待値と比較して，次元削減処理を行わない，つまり財務諸表から得られる会計数値をそのまま入力として用いるモデルだと，10% 程度の判別性能の向上が安定して見られることがわかる。一方で，SVD を用いた場合，全特徴量を用いた場合は判別性能が 6% 程度の向上と相対的に落ちるものの，目的変数との相関上位 100 変数，もしくは 200 変数を選択した場合のモデルでは，判別性能が 18% 程度向上するという相対的に最も良い結果が得られた。

表 7 テストデータに対する方法論ごとの比較結果（元の 410 変数を用いる場合）

	相関上位 100 特徴量		相関上位 200 特徴量		全特徴量	
	ACC	AUC	ACC	AUC	ACC	AUC
平均値	0.650	0.625	0.638	0.609	0.634	0.600
標準偏差	0.013	0.023	0.013	0.022	0.013	0.019

表 8 テストデータに対する方法論ごとの比較結果

（SVD によって生成した 410 変数を用いる場合）

	相関上位 100 特徴量		相関上位 200 特徴量		全特徴量	
	ACC	AUC	ACC	AUC	ACC	AUC
平均値	0.730	0.757	0.735	0.754	0.615	0.573
標準偏差	0.015	0.018	0.017	0.019	0.014	0.020

※平均値はテストデータに対する「(2005 年度の ACC + ... + 2015 年度の ACC)/11」とテストデータに対する「(2005 年度の AUC + ... + 2015 年度の AUC)/11」で算出した。

3.3 実験 3：本研究が提案する方法論にかかる実験

実験 1 と実験 2 の結果をふまえ、東証一部上場企業では、 $t-1$ 年度と t 年度の財務諸表の会計数値から簡単に作成できる変量から、 $t+1$ 年度中の長期借入金の増加を判別するための、あるべき方法論の道筋が見えてきた。具体的には、データのスケーリング処理はあまり重要ではない点、作成した変量をそのまま用いるのではなく、SVD のような教師なし学習としての次元削減手法を用いる点、ただし次元削減手法によって変量を減らすのではなく、目的変数との相関が高い特徴量を選択することが重要となる点である。

一方で、水戸信用金庫との意見交換から、地域金融機関では「財務的に苦しいからお金を借りる」というようなネガティブな理由の融資案件と、「経営はうまく行っており、リニューアルや事業拡大のためにお金を借りる」というようなポジティブな理由の融資案件があるという。そこで本研究では、財務諸表からこれらのネガティブ／ポジティブな状況を区分し、より判別性能を高めるモデルを提案したい。

財務諸表からこれらのネガティブ／ポジティブな状況を明確に区分することは難しいが、少なくとも $t-1$ 年度の売上よりも t 年度の売上が減少している企業は、 $t+1$ 年度においてネガティブな状況の場合が多く、逆に、 $t-1$ 年度の売上よりも t 年度の売上が増加している企業は、 $t+1$ 年度においてポジティブな状況の場合が多いという仮説を、水戸信用金庫との意見交換から導いた。そこで本研究が提案する方法論では、「 t 年度の売上高/ $t-1$ 年度の売上高」が 1 以上となった企業と、1 未満となった企業とでデータを分割し、それぞれの分割後データに対して、実験 1 と実験 2 で得られた最も良さそうな方法論を適用し、判別性能を評価することとした。なお、売上の増減を基準にデータを分割するため、各基準年度のサンプル企業数は減少する。そのために、SVD の計算上の性質から、企業数が生成したい変数の数より少ない場合、SVD 後の特徴量候補数は、企業数と同一の値としたことに注意されたい。

表 9 は、基準年度 t について、 $t-1$ 年度から t 年度にかけて売上が増加した企業と減少した企業のそれぞれの企業数と、それらの中で $t+1$ 年度中に長期借入金が増加した企業の割合を示したものである。また、表 9 では同時に、売上の増減でデータを分割した後に、学習済のモデルにテストデータを入力したときの ACC と AUC も示している。この実験は、実験 1 と実験 2 の結果をふまえ、データを標準化し、SVD で 410 変量を新たに生成し、相関上位 100 変量の特徴量選択したモデルを使用した。

表 9 の結果から、興味深い結果が確認された。売上が増加した企業で学習をした場合、リーマンショックが起きた 2008 年度と翌年 2009 年度以外は、テストデータの ACC が 0.70 程度、AUC が 0.77 程度と、判別性能の大きな向上は見られない。しかしながら、売上が減少した企業で学習した場合には、同じく 2008 年度と 2009 年度以外は、テストデータの ACC も AUC も、0.95 程度の非常に高い判別性能を報告している。一方で、全期間のデータで学習し

た場合は、判別性能が大きく向上するわけではない。

表 9 売上によるデータ分割後の企業数、目的変数の分布、および実験 3 の結果

売上増加企業					売上減少企業				
基準年度 t	企業数	割合	ACC	AUC	基準年度 t	企業数	割合	ACC	AUC
2005	784	0.451	0.761	0.783	2005	199	0.342	0.970	0.987
2006	974	0.461	0.738	0.776	2006	177	0.429	0.949	0.965
2007	891	0.557	0.758	0.779	2007	265	0.449	0.932	0.949
2008	428	0.432	0.900	0.928	2008	735	0.465	0.786	0.821
2009	254	0.350	0.957	0.966	2009	921	0.349	0.742	0.754
2010	844	0.427	0.759	0.786	2010	333	0.420	0.933	0.963
2011	802	0.445	0.767	0.793	2011	410	0.442	0.907	0.930
2012	800	0.494	0.781	0.815	2012	424	0.505	0.889	0.927
2013	1054	0.491	0.738	0.753	2013	175	0.440	0.988	0.987
2014	955	0.440	0.732	0.756	2014	303	0.439	0.950	0.971
2015	873	0.464	0.774	0.807	2015	387	0.390	0.935	0.967
全期間	8659	0.465	0.635	0.664	全期間	4329	0.421	0.660	0.684

これらの結果から、(1) マクロ経済的な状況が悪い時に売上を増加させた企業群、(2) マクロ経済的な状況が悪くない時に売上が減少した企業群、の 2 つには、それぞれその基準年度前後の状況固有の財務諸表上の相関関係が生じ、判別性能を高める可能性が示されたと言える。逆に、ある年度の文脈に依存せず一般的な財務諸表上の相関関係が存在し、どの年でも頑健に $t+1$ 年度中の長期借入金の増加を判別することは十分にできていないことに限界があることも忘れてはならない。なお、表 9 の結果の背後にある正誤状況は、参考 7 と参考 8 (文末に補足資料として掲載) の混同行列のとおりである。

最後に、本研究が実験 3 で提案した、売上増減でデータを分割して学習する方法論を適用した場合の判別性能と、実験 2 の参考 1 で示した、売上増減によりデータを分割しない場合の判別性能とで、平均的な ACC を比較した結果を表 10 に示す。表 10 によると、基準年度を 2006 年度とした場合に見られる 1.8% ポイントの判別性能の向上を最小として、平均的に 6.9% ポイントの判別性能の向上を確認できた。

そして、本研究が提案する方法論は、最低でも 77.1%、平均的に 80.9% の「 $t+1$ 年度中の長期借入金の増加」の判別を正解できるという実験結果となった。このことは、期待値としての正解が 55% 程度であること、そして実験 1 の予備実験で前述したように、会計情報を素朴にそのまま学習させたモデルの ACC が 63.6% であることを鑑みると、財務諸表上の会計

数値は、機械学習を用いて企業の資金ニーズを予測する上で、意味ある情報内容を有していると考えられる。

表 10 テストデータにおける ACC (実験 2 と実験 3 の比較)

基準年度 t	売上増減によりデータを分割した場合の ACC	売上増減によりデータを分割しない場合の ACC	提案した方法論を用いた場合の効果
2005	0.804	0.755	+0.049
2006	0.771	0.753	+0.018
2007	0.798	0.717	+0.081
2008	0.828	0.727	+0.101
2009	0.788	0.716	+0.072
2010	0.809	0.740	+0.069
2011	0.814	0.743	+0.071
2012	0.819	0.718	+0.101
2013	0.774	0.710	+0.064
2014	0.785	0.733	+0.052
2015	0.824	0.722	+0.102
全期間	0.801	0.730	0.069

4 議論と展望

地域金融機関では、機械学習を活用し、取引先企業の資金ニーズに対して適切に対応することが求められている。本研究では、一般的な財務諸表を用いて長期借入金が増加していたであろう企業を判別する方法論を提示し、その技術実験と評価を行った。技術実験は、最も原始的なロジスティック回帰モデルをベースモデルとし、東証一部上場企業の財務諸表を用いて、テストデータにおける判別性能を評価した。

本研究は、先行研究がほとんどない中での嚆矢的な技術実験・評価であるため、3段階からなる実験デザインを構築した。実験 1 は、あるべき方法論の探索実験である。実験 2 は、実験 1 の結果をふまえた方法論の詳細比較実験である。実験 3 は、実験 1 と実験 2 の結果をふまえた、本研究が提案する方法論にかかる実験である。これら 3 段階の実験の結果、本研究の問題設定における財務諸表の会計数値を用いた機械学習において、以下の 5 つの主要な発見があった。

1. 生データのスケール処理は、判別性能の向上に大きな効果をもたらさないこと。

2. 同様の変量を束ねて新たな変量を生成する SVD のような次元削減手法の応用は、会計数値をそのまま特徴量として用いるよりも、判別性能の向上に大きな効果をもたらすこと。
3. ただし、次元削減手法の応用によって次元削減を行うことが必ずしも判別性能の向上に大きく貢献するわけではない。むしろ、本研究の実験結果では、目的変数との相関が高い変量を選ぶ特徴量選択の方が、判別性能の向上に大きな効果をもたらしたこと。
4. 全ての企業を同じモデルで学習するよりも、ポジティブ／ネガティブな融資状況のようなドメイン知識をふまえてデータを分割して学習する方が、判別性能の向上のために大きな効果をもたらすこと。
5. $t-1$ 年度と t 年度のデータから $t+1$ 年度中の長期借入金の増加を判別する問題に対し、どの年度にも一般的に有効な機械学習モデルを構築することは現時点で難しいこと。しかし、それぞれの年度に対して平均的に 80%以上の正解率をもたらすモデルは構築できそうであること。

ただし、本研究が用いたデータは、東証一部上場企業のものであることに注意が必要である。地域金融機関が本研究で提案した方法論を適用する際には、取引先の典型的な小規模性や、自己資本比率のマイナス方向への散らばり等を注意せねばならない。しかし、本研究が将来的に目指す地域金融機関における取引先の資金ニーズ予測に際して、財務諸表の一定の利活用可能性を示すことができたと思われる。

加えて、本研究が用いた機械学習のためのモデルは、最も単純な手法の1つであるロジスティック回帰モデルである。今後の本格的なラボ実験（TRL4）や地域金融機関が蓄積・保有する実データを用いた技術評価（TRL5）の際には、実用上の機械学習タスクを具体化・洗練化させるとともに、それに最もふさわしい高度な手法の利用も実験・検討すべきである。その際は、中小企業の倒産を予測する二値分類モデルにおいて、複数のモデル・サンプリング手法・特徴量選択手法を比較検討した研究（Papíková and Papík 2022）等を参考とし、機械学習の文脈から実験デザインを構築することが有効であろう（Kaur and Dharni 2022）。

最後に、地域機関が有する取引先の財務諸表データは、機密情報であることを忘れてはならない。また、地域金融機関の実情として、本研究で提示した方法論を適用するために必要な経営リソースが不足している可能性もある。その場合、我々研究機関を含む外部のデータ解析担当者に対して、実データを提供せずに、重要な情報内容を保持したデータ共有のスキームを併せて検討しなければならないかもしれない。この点について、実際の生データを変換した中間表現データを作成し、それを外部の解析機関へ共有してデータ解析を行い、解析結果やモデルを返すデータコラボレーション解析の方法論（Bogdanova et al. 2020; Takahashi

et al. 2021; Imakura et al. 2021) が提案されるようになった。そして、本研究の技術実験・評価の結果を見ると、そもそも生データではなく、一種の中間表現データである SVD を用いて新たに生成したデータを用いた方が、判別性能が高いという発見も非常に興味深い。本研究の技術実験・評価の方向性の先に TRL9 のサービス実装を目指すには、こうした新技術の応用も含め、経営リソースが限られた地域金融機関にとってリスクが小さく費用対効果が高いサービスモデルの検討も必要となる。

参考文献

- 金融庁. 2022. 「中小・地域金融機関向けの総合的な監督指針」.
- 市原直通・首藤昭信. 2017. 「FinTech×監査の現状：AI で見抜く不正会計」『企業会計』65 (2): 55-63.
- 首藤昭信. 2019. 「AI が会計学研究に与える影響」『会計』195 (2): 127-141.
- Ain, N., G. Vaia, W. H. DeLone and M. Waheed. 2019. Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. *Decision Support Systems* 125: 113113.
- Beaulieu, P. R. 2020. Contract-Based Cost Analytics. *Journal of Emerging Technologies in Accounting* 17 (1): 11-19.
- Bogdanova, A., A. Nakai, Y. Okada, A. Imakura and T. Sakurai. 2020. Federated Learning System without Model Sharing through Integration of Dimensional Reduced Data Representations. *Proceedings of FL-IJCAI 2020*: 1-7.
- Božič, K. and V. Dimovski. 2019. Business intelligence and analytics for value creation: The role of absorptive capacity. *International Journal of Information Management* 46: 93-103.
- Chen X., Y. H. T. Cho, Y. Dou and B. Lev. 2022. Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research* 60 (2): 467-515.
- Elliot, V. H., M. Paananen and M. Staron. 2020. Artificial Intelligence for Decision-Makers. *Journal of Emerging Technologies in Accounting*. 17 (1): 51–55.
- Guo K. H., X. Yu and C. Wilkin. 2022. A Picture is Worth a Thousand Journal Entries: Accounting Graph Topology for Auditing and Fraud Detection. *Journal of Information Systems* 36 (2): 53-81.
- Henry, E. 2006. Market Reaction to Verbal Components of Earnings Press Releases: Using a Predictive Algorithm. *Journal of Emerging Technologies in Accounting*. 3 (1): 1-19.
- Huang, A. H., A. Y. Zang and R. Zheng. 2014. Evidence on the Information Content of Text in Analyst Reports. *The Accounting Review*. 89 (6): 2151-2180.
- Imakura, A., H. Inaba, Y. Okada and T. Sakurai. 2021. Interpretable Collaborative Data Analysis on Distributed Data. *Expert Systems with Applications*. 177: 114891.
- Kaur J. and K. Dharni. 2022. Application and performance of data mining techniques in stock market: A review.

- Intelligent Systems in Accounting, Finance and Management* 29 (4): 219-241.
- Liang P. J., A. Wang, L. Akoglu and C. Faloutsos. 2021. Pattern Recognition and Anomaly Detection in Bookkeeping Data. *Asian Bureau of Finance and Economic Research Working Papers*.
- Lobo, J. M., A. Jiménez-Valverde and R. Real. 2007. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17(2) :145-151.
- Mandrekar, J. N. 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* 5 (9): 1315-1316.
- Marshall, T. E. and S. L. Lambert. 2018. Cloud-Based Intelligent Accounting Applications: Accounting Task Automation Using IBM Watson Cognitive Computing. *Journal of Emerging Technologies in Accounting*. 15 (1): 199–215.
- Müller R., M. Schreyer, T. Sattarov and D. Borth. 2022. RESHAPE: Explaining Accounting Anomalies in Financial Statement Audits by enhancing SHapley Additive exPlanations. *ICAIF '22: Proceedings of the Third ACM International Conference on AI in Finance*: 174-182.
- Papíková L. and M. Papík. 2022. Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises. *Intelligent Systems in Accounting, Finance and Management* 29 (4): 254-281.
- Rikhardsson, P. and O. Yigitbasioglu. 2018. Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems* 29: 37-58.
- Takahashi Y., H. Chang, A. Nakai, R. Kagawa, H. Ando, A. Imakura, Y. Okada, H. Tsurushima, K. Suzuki and T. Sakurai. 2021. Decentralized Learning with Virtual Patients for Medical Diagnosis of Diabetes. *SN Computer Science* 2: 239.
- Youden, W. J. 1950. Index for rating diagnostic tests. *Cancer* 3 (1): 32-35.

(付記) 本論文は、日本会計研究学会第 80 回全国大会における自由論題報告の内容を加筆・修正したものである。なお、本研究は、NEDO 委託事業「データコラボレーション解析による生産性向上を目指した次世代人工知能技術の研究開発」、および JSPS 科研費 (22H00895) による研究成果の一部である。

補足資料

参考 1 SVD 有り， 相関上位 100 特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.782	0.858	0.755	0.783
2006	0.766	0.845	0.753	0.784
2007	0.737	0.817	0.717	0.747
2008	0.756	0.828	0.727	0.756
2009	0.737	0.808	0.716	0.734
2010	0.760	0.837	0.740	0.783
2011	0.759	0.832	0.743	0.766
2012	0.741	0.821	0.718	0.749
2013	0.728	0.805	0.710	0.732
2014	0.734	0.808	0.733	0.742
2015	0.747	0.818	0.722	0.755
全期間	0.634	0.680	0.637	0.668

参考 2 SVD 有り， 相関上位 200 特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.836	0.905	0.774	0.787
2006	0.799	0.883	0.757	0.784
2007	0.779	0.863	0.723	0.743
2008	0.787	0.866	0.734	0.751
2009	0.771	0.852	0.719	0.723
2010	0.791	0.871	0.744	0.774
2011	0.791	0.867	0.742	0.762
2012	0.777	0.863	0.729	0.754
2013	0.767	0.846	0.714	0.732
2014	0.772	0.853	0.726	0.743
2015	0.782	0.857	0.728	0.746
全期間	0.642	0.693	0.636	0.671

参考3 SVD有り, 全特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.846	0.918	0.642	0.594
2006	0.810	0.896	0.633	0.605
2007	0.792	0.875	0.585	0.550
2008	0.793	0.877	0.621	0.572
2009	0.800	0.866	0.611	0.532
2010	0.803	0.885	0.605	0.580
2011	0.780	0.881	0.614	0.578
2012	0.794	0.879	0.614	0.589
2013	0.776	0.862	0.605	0.556
2014	0.783	0.865	0.620	0.573
2015	0.786	0.870	0.615	0.579
全期間	0.646	0.696	0.627	0.651

参考4 SVD無し, 相関上位100特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.686	0.755	0.662	0.648
2006	0.684	0.741	0.657	0.626
2007	0.641	0.700	0.640	0.624
2008	0.659	0.708	0.640	0.606
2009	0.656	0.708	0.631	0.580
2010	0.697	0.745	0.671	0.652
2011	0.673	0.739	0.648	0.639
2012	0.667	0.717	0.640	0.615
2013	0.666	0.720	0.642	0.618
2014	0.672	0.709	0.646	0.603
2015	0.687	0.749	0.674	0.662
全期間	0.615	0.653	0.622	0.641

参考5 SVD無し，相関上位200特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.719	0.797	0.650	0.632
2006	0.713	0.780	0.652	0.619
2007	0.655	0.733	0.629	0.602
2008	0.695	0.758	0.635	0.610
2009	0.688	0.759	0.613	0.573
2010	0.718	0.766	0.647	0.604
2011	0.703	0.782	0.643	0.633
2012	0.690	0.755	0.639	0.609
2013	0.693	0.753	0.635	0.607
2014	0.679	0.745	0.617	0.570
2015	0.726	0.791	0.653	0.647
全期間	0.630	0.670	0.626	0.648

参考6 SVD無し，全特徴量を選択

基準年度 t	Train		Test	
	ACC	AUC	ACC	AUC
2005	0.762	0.845	0.642	0.604
2006	0.759	0.832	0.645	0.603
2007	0.743	0.825	0.638	0.621
2008	0.745	0.821	0.639	0.607
2009	0.726	0.798	0.617	0.559
2010	0.751	0.817	0.641	0.590
2011	0.738	0.819	0.620	0.605
2012	0.739	0.818	0.626	0.606
2013	0.729	0.810	0.631	0.602
2014	0.717	0.795	0.617	0.570
2015	0.747	0.822	0.661	0.628
全期間	0.639	0.692	0.628	0.655

参考 7 売上増加企業の混同行列

基準年度 t	企業数	真陰性	偽陽性	偽陰性	真陽性
2005	784	327	103	84	270
2006	974	365	160	95	354
2007	891	291	104	111	385
2008	428	217	26	17	168
2009	254	158	7	4	85
2010	844	360	124	79	281
2011	802	336	109	78	279
2012	800	312	93	82	313
2013	1054	394	143	133	384
2014	955	402	133	123	297
2015	873	376	92	105	300
全期間	8659	3073	1559	1602	2425

参考 8 売上減少企業の混同行列

基準年度 t	企業数	真陰性	偽陽性	偽陰性	真陽性
2005	199	125	6	0	68
2006	177	94	7	2	74
2007	265	131	15	3	116
2008	735	291	102	55	287
2009	921	443	157	81	240
2010	333	172	21	1	139
2011	410	206	23	15	166
2012	424	184	26	21	193
2013	175	97	1	1	76
2014	303	161	9	6	127
2015	387	218	18	7	144
全期間	4329	1703	804	668	1154