



日本銀行ワーキングペーパーシリーズ

入出金情報を用いた信用リスク評価 —機械学習による実証分析—

三浦 翔*

井實 康幸**

竹川 正浩***

No.19-J-4
2019年6月

日本銀行
〒103-8660 日本郵便（株）日本橋郵便局私書箱30号

* 金融機構局 金融高度化センター（現みずほ第一フィナンシャルテクノロジー株式会社）

** 株式会社りそな銀行 リスク統括部

*** 一般社団法人CRD協会 企画部

日本銀行ワーキングペーパーシリーズは、日本銀行員および外部研究者の研究成果をとりまとめたもので、内外の研究機関、研究者等の有識者から幅広くコメントを頂戴することを意図しています。ただし、論文の中で示された内容や意見は、日本銀行の公式見解を示すものではありません。

なお、本ペーパーに対するご意見・ご質問は、金融機構局金融高度化センター（caft@boj.or.jp）までお寄せ下さい。

商用目的で転載・複製を行う場合は、予め日本銀行情報サービス局（post.prd8@boj.or.jp）までご相談下さい。転載・複製を行う場合は、出所を明記して下さい。

入出金情報を用いた信用リスク評価[※]

－ 機械学習による実証分析 －

三浦翔* 井實康幸** 竹川正浩***

【要 旨】

金融機関における信用リスク管理業務では、法人債務者のデフォルトに対する予兆管理が行われている。この点、債務者が大企業を中心とした上場企業であれば、企業の信用状態をタイムリーに反映しやすい株価などをデフォルトの予兆管理指標として活用できるが、中堅中小企業を中心とした非上場企業には、信用状態を即時に反映する指標が存在しない。また、信用リスク管理業務では、財務情報を用いた信用リスク評価が一般的に行われているが、財務計数には、信頼性及び即時性の面で一定の制約がある。

そこで、本稿では、中堅中小企業を中心とした非上場企業にも適用可能で、かつタイムリーなモニタリングを実現するためのデフォルト予測モデルを構築する。具体的には、金融機関の預金口座における入出金情報を用いて、機械学習モデルや統計モデルを用いた信用リスク評価モデルを構築し、モデル精度の検証を行い、そのうえで予兆管理実務への適用可能性について検討する。

モデル精度検証の結果、入出金情報を用いた場合において、機械学習モデルの精度は十分に実用可能な水準であることが確認された。また、機械学習モデル対比ではやや精度が落ちるものの、解釈性に優れたロジットモデルについても、実務で十分に活用可能な精度を有することが確認された。

キーワード：信用リスク、予兆管理、機械学習、入出金情報

※ 本稿の作成に際し、山下智志教授（統計数理研究所）及び著者らの所属組織スタッフから有益なコメントを頂いた。また、本稿の内容は日本銀行金融機構局金融高度化センター主催の「AIを活用した金融の高度化に関するワークショップ」の第3回「信用評価」（2019年2月18日開催）における著者（三浦）の講演内容を基に作成しており、同ワークショップの参加者からも有益なコメントを頂いた。この場を借りて感謝したい。なお、本稿の内容は著者ら個人に属するものであり、日本銀行及び著者らの所属組織の公式見解を示すものではない。

* 日本銀行 金融機構局 金融高度化センター（現みずほ第一フィナンシャルテクノロジー株式会社）

** 株式会社りそな銀行 リスク統括部

*** 一般社団法人 CRD 協会 企画部

1. はじめに

(入出金情報を用いた信用リスク評価モデルの構築)

金融機関の信用リスク管理業務における、法人債務者に対するモニタリング実務では、デフォルトの予兆^{※1}管理指標として、企業の信用状態をタイムリーに反映しやすい株価や Moody's KMV の EDF^{※2}などが活用されている。一方、中堅中小企業を中心とした非上場企業には、株価や EDF といった債務者の信用状態を即時に反映しやすい指標が存在しない。

また、多くの金融機関では、大企業を中心とした、相対的に大きな与信を供与している先に対しては、経営者面談等を含む、きめ細かいモニタリングを頻繁に実施することで、できるだけ早期にデフォルトの予兆を捕捉しようというインセンティブが働く。一方、中堅中小企業のように、比較的与信額の小さな企業に対し、担当者が同様のモニタリングを行うことは、費用対効果の面で見合わない。

さらに、信用リスク評価では、一般的に財務情報が用いられるが、財務情報には、会計知識が十分でない担当者（経営者）が作成したがゆえに誤った計数や、改竄された計数が含まれる可能性があり、データの信頼性が十分でない場合がある^{※3}。

加えて、財務情報は 1 年ごとなどの決算のタイミングでしか更新されないため、財務情報と足もとの経営実態との間にかい離が発生するおそれがあり、融資実務における信用リスク評価を行ううえでの制約がある。

こうした問題意識から、以下の要件を満たす予兆管理モデルの構築を検討する。

- ① 株価等のマーケット情報でなく、債務者が中堅中小企業を中心とした非上場企業の場合であっても入手可能な情報を用いたモデルであること
- ② 作成者の知見の有無や意図が影響しない情報を用いたモデルであること
- ③ 債務者の実態をタイムリーに反映する情報を用いたモデルであること

そこで、本稿にて扱う実証分析では、このような要件を満たすため、信用リスク評価モデルを構築するうえでのインプット情報として、預金口座における入

※1 本分析における「予兆」とは、「デフォルト発生の数ヶ月前の時点におけるデフォルト発生の可能性」と定義する。

※2 EDF (Expected Default Frequency) は、米国 KMV 社が開発した、構造型モデルをベースとした予想デフォルト確率モデル。株価とバランスシート等の情報を基に構成される。詳細は、Moody's Analytics 社ページ等参照 (<https://www.moodysanalytics.com/>)。

※3 この点は、例えば Hasumi and Hirata (2010) にて指摘されている。

出金情報を用いる。

(機械学習モデルと統計モデル)

内部格付制度などの信用リスク評価実務で利用されるデフォルト予測モデルとしては、ロジットモデルを中心とした統計モデルが広く採用されており、モデルの構築に際しては財務情報を用いることが一般的である。ロジットモデルを構築する際、財務指標から説明変数として採用する指標を選択するプロセスである指標選択アプローチ^{※4}では、AR 値（詳細後述）などの定量的な基準に基づいて指標選択が行われる。一方で、指標選択のプロセスでは、これまで蓄積・共有されてきた知見やノウハウを用いた指標選択（エキスパートジャッジ）が行われるケースも多い^{※5}。しかし、入出金情報を用いる場合、どのような情報（指標）が有効かという知見が蓄積・共有されていないため、エキスパートジャッジによる指標選択が困難である。

また、入出金情報には、取引金額や取引件数における計数の多寡だけでなく、融資に伴う入金取引や公共料金の支払による出金取引など、入金及び出金取引の種類・内容に関する様々な情報が含まれている（これらの入出金の種類・内容を表す情報は「摘要」という項目で整理されることが一般的である）。従って、これらの様々な情報の中から、より適切な情報を多く取り込んで信用力変化を予測するモデルが望ましい。

以上の理由から、今般のモデル構築では、多数の指標候補の中から、定量的な基準を用いて機械的に指標選択するモデルとして、機械学習モデルを採用する。

さらに本研究では、多くの指標を自動的に取り込める機械学習モデルだけでなく、これまで信用リスク評価モデルとして広く用いられてきたロジットモデルについても、入出金情報を用いて構築し、精度検証を行うこととした。この目的は、入出金情報を用いた信用リスク評価においては、モデルの非線形性の強い機械学習モデルのみが有効なのか、もしくは線形のロジットモデルでも精度が担保されるのかを確認するためである。

※4 指標数が多い場合、実務的には、定量的な単指標分析を行って有効性が認められた指標をいくつか採用候補として抽出し、(現実的に計算が可能な) 指標の組み合わせ数においてモデル精度を比較するなどして最終的なモデルを構築・決定するアプローチが採用されることが多い。

※5 財務情報については、融資審査実務における知見の蓄積を通じて、安全性（自己資本比率など）や支払能力（インタレストカバレッジレシオ）などが、一般的に有効な指標として知られている。

【表 1：一般的な信用リスク評価（デフォルト予測）モデル構築における比較】

	一般的な指標選択方法	採用の検討が可能な指標数	線形 or 非線形
機械学習モデル	定量的な基準を用いて機械的に選択	多	非線形のモデル
ロジットモデル	定量分析+エキスパートジャッジ ※本稿では定量基準（機械学習モデルの指標重要度を参照）のみで指標選択。詳細は3節(4)を参照。	少	線形のモデル

なお、本分析では、（指標選択プロセスにおける精度向上余地ではなく）モデルの精度比較を行うことを主目的としているため、ロジットモデル構築に際しては、機械学習モデルの構築を通じて得られた指標の重要度を参考として指標の絞り込みを行い、モデル構築を行った。

（先行研究の概要）

本稿と同様の入出金情報を基に信用リスク評価モデルを構築した分析として、本田ほか（2018）では、入出金情報の指標（特徴量）をクロスリカレンスプロットと呼ばれる手法を用いて加工し、画像情報に変換したうえで、CNN^{*6}などの機械学習モデルを構築している。一方、本稿は、実務への応用を念頭に、指標に定性的な意味合いを保持したうえで、指標重要度の観点から説明可能なモデル構築を企図している点や、より解釈性が高いロジットモデルを構築し、精度を検証している点で異なる。

また、Yao ほか（2017）は、入出金データに機械学習モデルを適用し、財務情報を用いた従来のロジットモデル対比でデフォルト予測精度が向上することを報告している。根本ほか（2019）では、財務及び入出金情報に主成分分析を施し、ロジットモデルによるモデル構築を行っている。当研究では、入出金情報のみでも相応のデフォルト予測水準を確保できることに加え、財務情報に入出金情報を加味することによって、デフォルト予測精度がさらに向上することが指摘されている。

このほか、財務以外の情報を用いてデフォルトを予測する分析として、山中・木下（2018）は、企業の受注情報を基にしたデフォルト予測モデルを構築している。この研究では、企業の受注情報を用いて利益及び資産価値を算出したうえで

^{*6} 畳込みニューラルネットワーク Convolutional Neural Network の略。ディープラーニングのモデルの一つ。一般に、画像認識に用いられることが多い。

構造型モデルをベースにしたデフォルト確率を推計する手法を提案し、実企業の受注情報を用いたデフォルト確率の推計を行っている。

(本稿の構成)

本稿の構成は、以下の通りである。まず、2節で使用するデータの概要を述べる。3節では、本稿で構築するモデルの概要とその検証方法をまとめる。4節では、2節で記述したデータと3節で示したモデルを使用した分析結果を示す。最後に、5節で分析結果をまとめ、分析を通じて得られた所見及び今後の検討課題について整理する。なお、補論1では、本稿で用いた機械学習のモデルについて、その概要をまとめる。補論2では、本文での分析を踏まえて追加的に検証した結果を示す。

2. 使用するデータ

(1) 入出金情報の概要と集約方法

本分析で使用した入出金情報は、①時間及び②摘要の2つの軸で集約を行ってモデル構築用の指標を作成した。以下では、この2点について集約を行う前後の入出金情報の性質等を記載し、具体的な指標作成方法を記述する。

今回の分析では、りそな銀行における2014年10月から2018年5月末までの債務者企業（匿名）ごとの預金口座の入出金情報（預金口座取引履歴データ^{※7}）を利用した。また、時間として「秒」単位、取引情報として「摘要コード」^{※8}と呼ばれる粒度の入出金明細（月次で約36百万件、csvファイルで約9GB）を、図1のイメージのように、「月次」単位で、「債務者」ごとに入出金項目の「摘要グループ」（詳細後述）別に集計した（月次で約3万件、csvファイルで約70MB）。

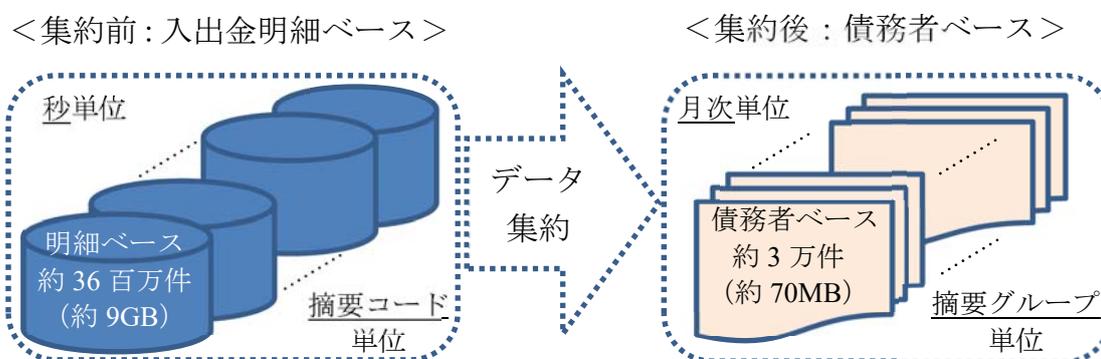
日次あるいは週次で集約することも可能であるが、この場合、直近に発生した特定の入出金取引に過度に反応するモデルとなるおそれがあるうえ、ノイズを含んだ（信用力の変化とは関わりのない）変動までも捕捉する可能性が高い。一方、月次単位での集約は、日次・週次といった短期間で集約した場合に生じるノイズが捨象されるため、関連部署への正確な情報提供という観点で実務的に運用しやすい。また、給与の振込みや公共料金の支払いなど、月次で発生する入出金取引、あるいは、月初・月末時点のみに発生する取引などが含まれるほか、より長い集約期間とすることで足元の変化を捉えにくくなる弊害を避けることが

^{※7} 本分析では、振込明細及び取立明細の粒度で情報を取得している。従って、例えば総合振込（総振）は振込先数の単位で取引件数を取得している。

^{※8} 入出金取引の取引内容を表すコード。例えば、振込や各種手数料の引落しなどの情報が分類されている。

できると考えられる。そこで、本稿では、月次単位での集約を行うこととした。

【図1：入出金情報の集約（概要）】



また、インプットとして用いる入出金情報の時系列が短すぎる場合、季節性を排除できず、適切な信用リスク評価が行えない可能性がある。このため、入出金情報を捕捉する期間としては、特定の季節性を控除するために1年間（12ヶ月間）以上は有すべき^{※9}であり、また、より長期間の情報を抽出すべく、今般の分析では24ヶ月間の入出金情報を基にデフォルト予測を行うこととした^{※10}。

（「摘要グループ」への集約）

今般の分析では、入金・出金の別（入払区分）と摘要内容を表す情報を用いて、リスク管理や融資審査などの実務経験に基づいた定性的判断を基に、入出金情報の摘要コードを、近しい情報を有すると考えられるグループに集約した（以下、「摘要グループ」）。また、分析には口座の預金残高（月中平残）も使用した。以上をまとめた結果は、表2の通りである。

※9 季節性を調整するモデル等を用いる方法も考えられるが、本分析では採用していない。

※10 補論2においては、使用する入出金情報の期間を短縮した場合の検証を行っている。

【表 2：入出金情報の摘要グループ構成】

No.	項目	摘要グループ名	概要（摘要コード等）
1	入金	売上	振込、でんさい、外為、など
2		投資	投資信託など
3		融資	融資
4		入金その他	カードなど
5	出金	原価	でんさい、送金、振込、など
6		変動費	カード、税金、など
7		固定費	光熱費、NHK 受信料、など
8		返済	貸出利息、返済、保証料、など
9		投資	投資信託など
10		不渡	不渡など
11	預金残高	預金残高	口座預金残高（月中平残）

(まとめ)

以上より、本稿で用いたデータ及びその集約方法をまとめると、表 3 の通り。

【表 3：使用するデータ概要】

項目		概要	備考
データ期間		2014/10～2018/5 (44 ヶ月分)	・当該期間の入出金情報及び対応する時点の内部格付（デフォルト情報含む）。
入出金 情報	集約頻度	月次	・集約前の情報は秒単位。
	集約単位	摘要グループ	・集約前の情報は摘要コード単位。 ・定性的判断を基に、近しい情報を有すると考えられる 10 グループ（及び預金残高）に集約。
	観測期間	24 ヶ月間	・季節性を排するため、直近 24 ヶ月分の入出金情報を使用。

(2) 指標作成

ここでは、モデル構築に用いる指標の作成方法として、本節(1)で説明した、集約後の入出金情報（月次・摘要グループ単位）を用いた指標作成について記述する。

まず、前述「摘要グループ」ごとに、金額及び件数（入出金件数）を月次集計して指標とする。本分析においては、「訂正^{※11}」オペレーションによる入出金情報も指標として含める。これらの指標の月ごとの指標数は、【金額・件数及び通

※11 訂正とは、一旦実行した入出金取引を事後的に訂正するオペレーション。手形引落しの訂正（依頼返却）などを含む。

常・訂正の全4項目】×【摘要グループ10項目】+【預金残高：1項目】の計41指標となる。また、本指標を直近24ヶ月分採用するため、指標数は、合計984となる（表4）。

【表4：各カテゴリの集約単位と指標数】

項目		期間	集約単位	摘要 グループ数	指標数	
					月次	全体
入出金 情報	(通 常)	24 ヶ月	取引の 金額及び件数	10	20	480 (=20×24)
	訂正				20	480 (=20×24)
預金残高			金額(月中平残)	1	1	24
合計				11	41	984 (=41×24)

このような指標に加えて、財務項目として売上高、総資産及び融資残高の3指標を用いる^{※12}。さらに、モデル精度の向上を企図し、表4の指標を売上高など3指標で除して基準化したもの（規模の情報を捨象することが目的）や、前月比や前年比、期間合計・期間標準偏差などの各加工を行って作成した指標バリエーションを追加し、最終的に合計5,989指標をモデル構築のための指標候補とした。

【表5：指標の加工方法の概要】

No.	指標数	概要	
①	984	入出金の月中合計【金額・件数】	以下指標の ベース
②	3	売上高、総資産、融資残高	
③	2,952	①/②で計算される指標（基準化）	
④	943	各指標の前月比	
⑤	492	各指標の前年比	
⑥	615	期間合計・標準偏差、期間合計の前年比など	
合計	5,989		

なお、表5の中で③～⑥の各指標数は、以下の通り計算される。

- ③ 売上高、総資産、融資残高で除して基準化（ $984 \times 3 = 2,952$ ）
- ④ 各指標の前月比（ $41 \times (24 - 1) = 943$ ）
- ⑤ 各指標の前年比（ $41 \times (24 - 12) = 492$ ）

※12 売上高、総資産は入出金情報の直近1ヶ月時点で得られる最新の財務情報より取得（更新頻度は概ね年次）。融資残高は月末時点の値を採用（更新頻度は月次）。

⑥ 各指標の期間合計・標準偏差、期間合計の前年比など ($41 \times 15^{*13} = 615$)

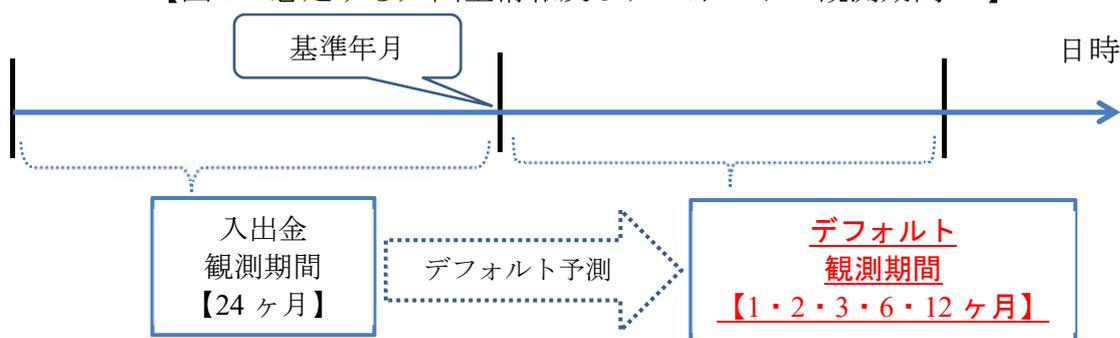
3. モデルの構築と検証

(1) 問題設定

基準時点より将来の特定期間内に発生するデフォルト^{*14}を予測するモデルを構築する。

モデル構築用のデータは、入出金データと内部格付の時系列データ（デフォルト情報含む）を紐付けして作成する。なお、信用リスク評価の分析では、デフォルト観測期間は一般的に1年間（12ヶ月間）とすることが多いが、本分析では、予兆管理への応用を企図し、より短い観測期間におけるデフォルト予測精度の検証のため、1・2・3・6・12ヶ月間の5パターンを設定する。

【図2：想定する入出金情報及びデフォルトの観測期間^{*15}】



(2) モデル構築のアプローチ

本稿の分析では、2節で記述したように、入出金情報に基づく約6,000指標を指標候補として用いる。本指標を用いたモデル構築に際し、以下の3点に留意して使用するモデルを検討した。

- ① すべての指標の組み合わせを考慮して最適な指標群を抽出すると膨大な組み合わせ数^{*16}となるため、一定の基準に基づいて機械的に指標選択が行わ

^{*13} 24・12・6・3ヶ月の期間における合計と標準偏差及び標準偏差を合計値で除して規格化した指標は、計 $4 \times 3 = 12$ 。直近12・6・3ヶ月の前年比(3)を加え、合計15となる。

^{*14} 本分析では、実務で広く採用されている基準として、内部格付における要管理先以下への格下げ事象をデフォルトと定義した。

^{*15} 実際には、2014年10月～2018年5月における入出金情報及び対応する内部格付のデータセットをMoving-Windowで対象期間を1ヶ月ずつずらしてモデル構築・検証用データを作成している。

^{*16} 例えば、6,000指標から20指標を選択する組み合わせは、 10^5 (1兆は 10^{12})

れることが望ましい。

- ② 入出金情報を用いた信用リスク評価において、どのような情報・指標が有効かという知見は一般に蓄積されていないため、エキスパートジャッジによる指標選択が困難であり、定量的に指標選択を行う必要がある。
- ③ 入出金情報には、取引金額や取引件数の多寡だけでなく、様々な種類の情報が含まれており、より適切な情報を多く取り込んで信用力変化を予測するモデルが望ましい。

こうした背景から、より多くの情報（指標）から定量的・機械的に適切な情報（指標）の抽出が可能である、機械学習モデルを検討対象とした。特に、本分析では、デフォルト・非デフォルトという 2 値の分類（判別）問題を取り扱うことや、機械学習モデルの中でも指標（説明変数）の重要度を評価可能であることなどを勘案し、決定木をベースとしたモデルを選択した。具体的には、「ランダムフォレスト」と「勾配ブースティングツリー」の 2 つの機械学習モデルの構築・検証を行った^{※17}。

実務でのモデルの利用に際しては、解釈性や説明性の高いデフォルト予測モデルが求められる。このため、現状の実務では、ロジットモデルが広く用いられている。本モデルは、信用スコア（デフォルト予測値／非デフォルト予測値の対数値）を被説明変数とし、選択した指標を説明変数とした線形回帰モデルであり^{※18}、選択した指標とデフォルト予測値との関係が明確である。リスク管理担当者に馴染みがあり、線形回帰モデルに基づくため解釈性や説明性が高いことから、実務的な観点で本モデルを採用することのメリットは大きい。但し、入出金情報を説明変数として用いる場合、①でも述べたように、指標選択については定量的・機械的な選択が必要になるため、本稿で取り扱う問題にロジットモデルを適用するには工夫が必要である^{※19}。

乗)の桁数で存在し、すべての組み合わせを探索することは現実的でない。指標の採用候補が多い場合、実務的には、AR 値などの単指標の精度等を基に、あらかじめ採用候補を絞ったうえで最適な組み合わせを探索するアプローチが採られることが多い。

※17 その他の機械学習モデルとして、サポートベクターマシン（SVM）やニューラルネットワーク（ディープラーニング含む）なども存在するが、信用リスク分野の特性上、結果の解釈を指標重要度を用いて行うことが可能な決定木をベースとするモデルを採用した。

※18 デフォルト予測値を p 、説明変数を X_1, \dots, X_k とすると、回帰係数を β_j ($j = 0, 1, \dots, k$)として以下の通り定式化される。

$$\log(p/(1-p)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

※19 ロジットモデルに対してもステップワイズ法（脚注 28 参照）などを用いて定量的・機

そこで、機械学習モデルにより指標を選択し、それをベースとするロジットモデルの説明変数として用いることで、精度が高く、かつ解釈性や説明力が高いモデルを構築できないか検証する。採用した 2 つの機械学習モデルは、指標重要度^{※20}の計算が可能であることから、これにより指標重要度の高い指標を選択し、それをロジットモデルの説明変数として用いることとする。すなわち、取り込める指標数が多く、非線形性の強い機械学習モデルに加え、同モデルで得られた指標重要度を踏まえ、従来のモデル構築アプローチで広く用いられているロジットモデルを構築する。

機械学習モデルとして本稿で採用するランダムフォレストと勾配ブースティングツリー、及び従来の統計モデルであるロジットモデルの概要は、表 6 の通りである。なお、機械学習モデル（ランダムフォレスト、勾配ブースティングツリー）の詳細については、補論 1 を参照。

【表 6：使用する 3 モデルの概要】

モデル		特徴等
機械学習モデル	ランダムフォレスト	決定木を構築する際、サンプル及び指標をランダムに抽出し、多数の決定木を <u>独立に構築</u> するモデル。 (機械学習の中では比較的) 過学習 ^{※21} しにくいモデル構造。
	勾配ブースティングツリー (XGBoost)	決定木を構築する際、それまでに作られたすべての決定木の予測結果を利用し、 <u>逐次的に決定木を構築</u> していくモデル。 過学習しやすい傾向があるため、クロスバリデーション等を用いてパラメータをチューニングする必要がある。
統計モデル	ロジットモデル	デフォルト確率を予測するモデルで、予め抽出した（一定数の）説明変数（指標）で回帰分析を行うモデル。現在、信用リスク評価モデルで最も広く採用されている。 線形回帰モデルのため（表 1、脚注 18 を参照）、一般に過学習しにくいモデル構造。

備考：ロジットモデルについて、本稿の分析では、指標のスコア化を行い、ステップワイズ法にて指標選択・モデル構築を実施している。詳細は本節(4)を参照。

械的に指標選択が可能であるが、約 6,000 指標を対象とした指標選択は計算時間の観点から現実的でない。

※20 各モデルにおける指標重要度の詳細は後述。なお、指標重要度としては、ロジットモデルのような回帰係数の有意性とは異なり、「各指標を用いて分類した場合、どの程度分類精度が向上するか」という観点から重要度の評価を行うことになる。

※21 「過学習」とは、モデル構築に用いたデータ（トレーニングデータ）へのフィットや的中率を高めるあまり、トレーニングデータ以外の未知データ（アウトオブサンプルデータ）に対するフィットや的中率（汎化性能）が低下する事象。非線形性の強いモデルで発生しやすい傾向がある。

(3) パラメータのチューニング

モデル構築用データとして、債務者ベースに集約した月次データ（図1右）から抽出した約7,000件（うちデフォルト約1,400件）の債務者データを用いる。これらのデータをトレーニングデータと検証データ（パラメータ選択用データ）^{※22}に2分割し、パラメータのチューニングを実施する。パラメータのチューニングは、AR値^{※23}が最大になるように調整する。

ランダムフォレストのパラメータとしては、決定木の数、指標の採用候補数（以下、「指標ユニバース数^{※24}」）のチューニングを行う。一方、勾配ブースティングツリーのパラメータとしては、決定木の深さ及び学習率のチューニングを行う^{※25}。

(4) ロジットモデル構築のアプローチ

ロジットモデルは、デフォルト予測モデルとして広く用いられている統計モデルの一つである。今般のモデル構築に際しては、指標値のスコア化^{※26}や、指標選択の際に機械学習モデルの構築結果（指標重要度）を参照することによって、できるだけ精度を保持しつつも高い解釈性や説明力を有するモデルの構築を試みた。本稿のロジットモデル構築における具体的なステップ（指標選択・パラメータ推計）は、以下の通りである。

【Step1】 ランダムフォレスト、勾配ブースティングツリーの各モデルの構築結果における指標重要度の上位（例えば100）指標をそれぞれ抽出し、

^{※22} 本分析では実施していないが、トレーニングデータ（約7,000件）に対してクロスバリデーションを用いてモデル構築（パラメータチューニング）を実施する方法もある。

^{※23} AR（Accuracy Ratio）値とは、指標値や信用スコアとデフォルトした先の序列性（デフォルト先が、その集団の中でより低信用と順位付けできているか）を測る統計量。-1～1の値をとり、1に近いほど、デフォルト先をより低位に評価できていると判断する。詳細は、山下・三浦（2011）を参照。

^{※24} 本分析では全体で5,989指標を採用候補としているが、ランダムフォレストの各決定木を生成する際には、5,989指標から（100指標、200指標など）特定数の指標をランダムに抽出して、その指標を基に決定木を生成する。本稿では、この特定数抽出された指標群を「指標ユニバース」と記述している。このようにランダムに指標を抽出することによってモデル精度を高めるための工夫がなされているが、詳細は補論1を参照。

^{※25} 決定木の数（反復回数）は、トレーニングデータを用いたクロスバリデーションによって決定した。

^{※26} 各指標をパーセント点で区切った際の、各区分におけるデフォルト率を参照しスコア化している。なお、ロジットモデルにおいてデフォルト率と指標の関係が適切に表現された場合、指標係数は負となるモデル構造を仮定している。

重複指標を削除。

【Step2】上記指標群を採用候補とし、トレーニングデータを用いて AIC^{※27}を評価関数とするステップワイズ^{※28}（指標削減）手法でモデルを構築^{※29}。

【Step3】Step2 で得られたモデルの指標係数が正の場合、多重共線性が発生しているとみなし^{※30}、本指標を採用候補から除外し、再度モデルを推計^{※31}。

【Step4】すべての指標係数が負となるまで、Step2・3を繰り返す。

なお、このステップは、2つの機械学習モデルの指標重要度でそれぞれ上位5～120（上記の例では100）となった指標を抽出したうえで実施する^{※32}。

（5）バックテスト

構築したモデルの検証として、バックテストを行う。具体的には、中堅中小企業の与信ポートフォリオに本モデルを運用した場合を想定し、月次の同与信ポートフォリオにおけるデフォルトの予測精度を確認する。このとき、バックテストにおけるモデル精度の検証はAR値を用いる。

一連のモデル構築とバックテストに使用したデータをまとめると図3のようになる。モデル構築用に使用したデータとバックテストに使用した入出金データの期間は、時点の重複がなく、バックテストは作成したモデルの汎化性能を評価するためのアウトオブサンプルでの検証となっている。

※27 Akaike Information Criterion（赤池情報量規準）の略。モデル選択（指標選択）時に用いられる基準の一つ。

※28 ステップワイズ法とは、統計モデルにおける説明変数の選択手法の一つ。変数を1つずつ追加・削除することによって評価関数（本稿ではAIC）がどのように変化するかを逐次計算し、最適な説明変数の組み合わせを探索する手法。変数を増やしていく変数増加法、減らしていく変数減少法などがある。

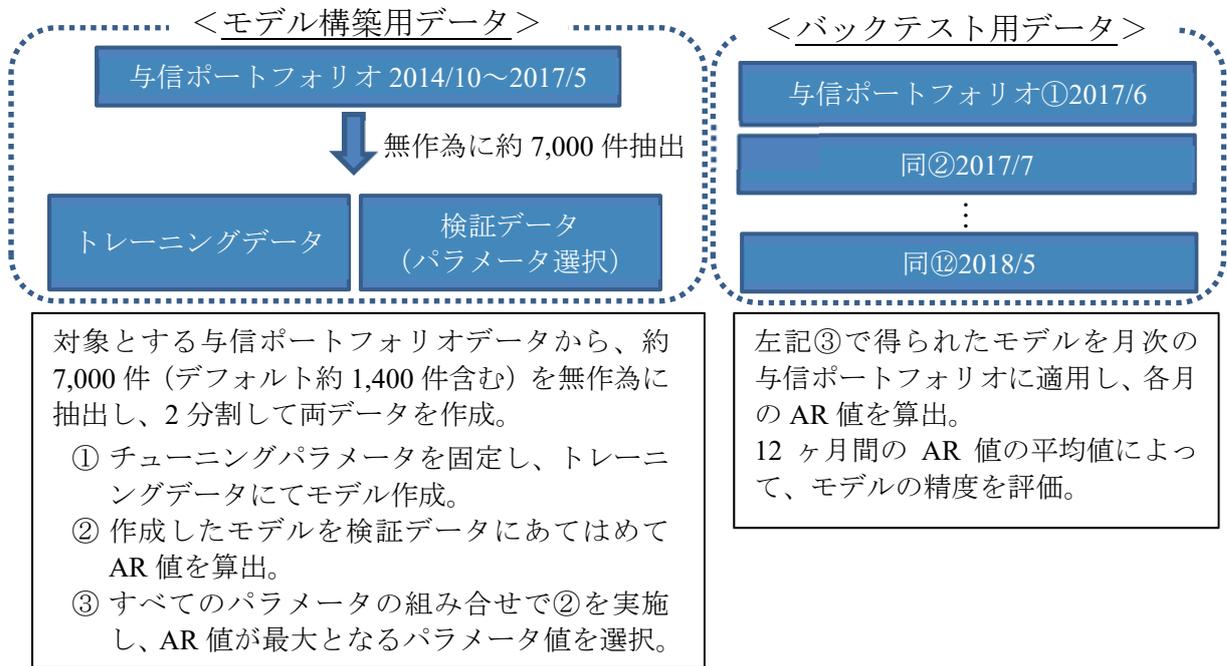
※29 別途、LASSO（Least Absolute Shrinkage and Selection Operator）による指標選択・モデル構築手法も試行した。その結果、正則化項の値の選択にも依存するが、本分析で採用しているステップワイズ・AIC手法とほぼ近いモデルが得られることを確認している。

※30 脚注26の通り、各指標値をスコア化する際、ロジットモデルにおいてデフォルト率と指標の関係が適切に表現された場合、指標係数は負となるモデル構造を仮定している。そのため、符号が正となる場合は、多重共線性の影響によるものと考えられる。

※31 最終的に得られたモデルにおいて、指標値の信頼度を検証するために p 値を確認している。本分析においては、一部の指標に5%で有意でない指標が含まれていたが、影響は軽微とみなし指標の除外は行っていない。

※32 5～120指標を5指標刻みで指標数を定めてモデルを構築する。構築したモデルを検証データに適用した際にAR値が最大となる指標数を特定している。

【図3：モデル構築用及びバックテスト用データの概要】



4. 分析結果

本節では、まず、モデル構築の前段として、単指標のデフォルト予測有効性について AR 値を用いて概観する。次に、ランダムフォレスト、勾配ブースティングツリー及びロジットモデルの各種モデルの構築結果^{※33}を示す。

また、各モデルについて、3 節(5)で示したモデルの検証を行ったうえで、さらに、債務者をメイン取引先・非メイン取引先で分けた場合のモデル精度の比較や財務スコアリング計数を併用した場合のモデル精度の影響度についての検証結果を示す。

(1) 単指標分析

モデル構築に際し、AR 値を用いた単指標分析を行った。対象データとして、2 節に記載したデータから約 7,000 件（デフォルト 1,400 件含む）を抽出したデータを対象とする。

まず、各カテゴリのデフォルト予測能力を確認するため、各カテゴリにおける

※33 尾木（2017）によれば、過去の分析では、企業規模によって信用リスク評価モデルの AR 値は水準が異なっており、一般的には、大企業では 0.8 程度、中小企業では 0.6~0.7 程度、小企業では 0.4 程度と報告されている。

AR 値の上位 20 指標の平均値を計算した結果が表 7 である。表 7 より、「預金残高」のデフォルト判別力が高く、続いて、「出金返済」、「入金売上」などの AR 値が高い傾向にあることがわかる。また、上記指標ではデフォルト観測期間が長いほど AR 値の水準は低下し、直感と整合的な結果となった。

【表 7：各指標カテゴリにおける上位 20 指標の平均 AR 値】

カテゴリ (摘要グループ)		デフォルト観測期間					
		1M	2M	3M	6M	12M	
1	入金	売上	0.2277	0.2160	0.2166	0.2119	0.1994
2		投資	0.1800	0.1794	0.1757	0.1822	0.1825
3		融資	0.1993	0.1998	0.1993	0.2029	0.1994
4		入金その他	0.1993	0.1993	0.1993	0.1994	0.1992
5	出金	原価	0.2059	0.2000	0.2007	0.2011	0.1989
6		変動費	0.2138	0.2050	0.2031	0.2001	0.1994
7		固定費	0.1334	0.1407	0.1364	0.1277	0.0962
8		返済	0.3050	0.2777	0.2747	0.2540	0.2377
9		投資	0.1990	0.1994	0.1992	0.1992	0.1992
10		不渡	0.1993	0.1993	0.1993	0.1994	0.1994
預金残高(月中平残)			0.4787	0.4644	0.4631	0.4561	0.4241

次に、全指標中、最も AR 値が高い指標（上位 5 指標）を示した結果が表 8 である。表 8 より、「預金残高（及び預金残高を加工した指標）^{※34}」がいずれのデフォルト観測期間においても上位であることがわかる。なお、表 8 中に記載の「預金残高(○)」の○数字は、表 5 の「No.」欄の指標加工バリエーションの番号に対応しており、同じ番号であっても異なる指標の場合もある（冗長のため、後続の表では記載を割愛）。

【表 8：全指標中の AR 値上位 5 指標】

順位	デフォルト観測期間									
	1M		2M		3M		6M		12M	
	指標名	AR値	指標名	AR値	指標名	AR値	指標名	AR値	指標名	AR値
1	預金残高(①)	0.5286	預金残高(①)	0.5187	預金残高(①)	0.5169	預金残高(①)	0.5022	預金残高(①)	0.4643
2	預金残高(①)	0.5145	預金残高(③)	0.5030	預金残高(⑥)	0.4997	預金残高(⑥)	0.4877	預金残高(⑥)	0.4572
3	預金残高(⑥)	0.5140	預金残高(⑥)	0.4977	預金残高(①)	0.4990	預金残高(①)	0.4857	預金残高(①)	0.4546
4	預金残高(③)	0.5134	預金残高(①)	0.4975	預金残高(③)	0.4924	預金残高(③)	0.4836	預金残高(①)	0.4517
5	預金残高(③)	0.5131	預金残高(③)	0.4931	預金残高(①)	0.4843	預金残高(①)	0.4788	預金残高(⑥)	0.4407

表 8 より、「預金残高（及び預金残高を加工した指標）」の AR 値が高い傾向を確認できたが、「預金残高」に次いで AR 値が高い指標を確認するため、「預金残高」を除いた上位 5 指標について検証した。その結果が表 9 である。上位指標に

※34 月次の預金残高（月中平残）だけでなく、預金残高（同）の前年比や期間平均などの指標の加工方法のバリエーションも含まれる。詳細は表 5 を参照。

は「出金返済」の指標が多く、表 7 と整合的な結果となった。これらの結果から、後述のランダムフォレストや勾配ブースティングツリーのモデル構築においては、「預金残高」と「出金返済」の指標を中心にモデルが構築されることが想定される。

なお、表 9 の指標順位から、デフォルト観測期間が長いほど、順位が低下する（上位に「預金残高」が集中している）傾向が確認された。このことから、より短い期間では「預金残高」以外にもデフォルト予測に有効な指標が存在する一方、長期になると「預金残高」の有効度が相対的に高いことを表している。

【表 9：預金残高を除いた場合の AR 値上位指標】

デフォルト観測期間									
1M		2M		3M		6M		12M	
順位	カテゴリ	順位	カテゴリ	順位	カテゴリ	順位	カテゴリ	順位	カテゴリ
30	出金返済	44	出金返済	64	出金返済	86	総資産	88	出金返済
40	出金返済	66	出金返済	77	出金返済	87	出金返済	89	出金返済
55	出金返済	80	出金返済	83	出金返済	94	入金融資	92	総資産
72	出金返済	82	出金返済	85	出金返済	96	出金返済	93	出金返済
76	出金返済	85	出金返済	86	出金返済	99	出金返済	96	出金返済

(2) モデル構築結果

3節で示した 3 モデル（ランダムフォレスト、勾配ブースティングツリー、ロジットモデル）を構築した結果を以下に示す。

なお、デフォルト観測期間は 5 パターン（1・2・3・6・12 ヶ月）でモデル構築しているが、パラメータのチューニング方法等はいずれも共通であるため、以下では、特段の記載がない場合は観測期間が 3 ヶ月の結果のみを示す。

イ. ランダムフォレスト

ランダムフォレストを用いて、チューニングパラメータ（決定木の数、指標ユニバース数）を変更しながらチューニングを行ったときの検証データにおける AR 値は、表 10 のようになった。表 10 の通り、決定木の数が 350、指標ユニバース数が 150 のときに AR 値が最大となった（0.7070）。なお、指標ユニバース数を大きくすると（表 10 の右側）、AR 値が低下する傾向にあり、やや過学習の傾向が見て取れる^{※35}。

※35 指標ユニバース数を大きくすると、より多くの指標を採用する結果としてトレーニングデータに過度にフィットし、検証データにおけるフィット（AR 値）が低下していると考えられる。

【表 10：検証データの AR 値（ランダムフォレスト）】

	50	100	150	200	250	300	350	400	450	500
50	0.6627	0.6796	0.6767	0.6812	0.6718	0.6857	0.6830	0.6786	0.6761	0.6752
100	0.6772	0.6953	0.6951	0.7013	0.6951	0.6907	0.6895	0.6892	0.6915	0.6914
150	0.6930	0.6983	0.6981	0.6970	0.6992	0.6970	0.6888	0.6940	0.6871	0.6844
200	0.6962	0.7055	0.6975	0.6976	0.6960	0.7029	0.6883	0.6935	0.6882	0.6936
250	0.6978	0.6985	0.7014	0.6967	0.7019	0.6952	0.6953	0.6971	0.6958	0.6935
300	0.7000	0.6988	0.6990	0.6977	0.6985	0.7031	0.6920	0.6973	0.6915	0.6880
350	0.7007	0.7014	0.7070	0.7041	0.6963	0.7009	0.6974	0.6971	0.6927	0.6982
400	0.7021	0.7040	0.7020	0.7025	0.7005	0.6985	0.6965	0.6957	0.6956	0.6906
450	0.7000	0.7059	0.7005	0.7034	0.7005	0.6996	0.6987	0.7001	0.6928	0.6931
500	0.6978	0.7061	0.7023	0.7011	0.7035	0.7006	0.6999	0.6971	0.6977	0.6959

備考：縦軸は木の数、横軸は指標ユニバース数。

AR 値が最大となったパラメータ（決定木の数：350、指標ユニバース数：150）における指標重要度^{※36}は、表 11 の通りとなった。上位 20 指標中、17 指標が「預金残高」で、その他は「出金返済」が 3 指標と、重要度の高い指標カテゴリが限定的であることがわかる。なお、表 11 には記載していないが、「預金残高」、「出金返済」以外の指標の最上位は「融資残高」で 85 位であった。すなわち、重要度の上位 84 指標は、すべて「預金残高」、「出金返済」の指標で占められており、主にこれら 2 カテゴリからモデルが構築されていることがわかる。これは、本節(1)の単指標の分析で想定された結果と整合的である。

【表 11：指標重要度（ランダムフォレスト）】

重要度順位	カテゴリ	Mean Decrease Gini(比率)	累積比率
1	預金残高	1.0%	1.0%
2	預金残高	0.9%	1.9%
3	預金残高	0.8%	2.7%
4	預金残高	0.7%	3.4%
5	預金残高	0.7%	4.1%
6	預金残高	0.7%	4.8%
7	預金残高	0.7%	5.5%
8	預金残高	0.6%	6.1%
9	預金残高	0.6%	6.7%
10	預金残高	0.6%	7.3%
11	預金残高	0.5%	7.8%
12	預金残高	0.5%	8.3%
13	出金返済	0.5%	8.8%
14	出金返済	0.5%	9.3%
15	預金残高	0.4%	9.7%
16	出金返済	0.4%	10.1%
17	預金残高	0.4%	10.6%
18	預金残高	0.4%	11.0%
19	預金残高	0.4%	11.4%
20	預金残高	0.4%	11.8%

ロ. 勾配ブースティングツリー（XGBoost）

勾配ブースティングツリーを用いて、チューニングパラメータを変更したときの検証データにおける AR 値は表 12 の通りであり、決定木の深さ 4、学習率

※36 ランダムフォレストにおける指標重要度は、OOB（Out-Of-Bag）誤り率を用いる計算手法と Gini 係数の減少度を用いる計算手法等があるが、ここでは後者を用いている。なお、表中「比率」は、Gini 係数の減少率平均（Mean Decrease Gini）の合計値が 100%となるように基準化した場合の各指標の占有率を示している。なお、OOB については脚注 67 参照。

0.01 のときに AR 値が最大^{※37} (0.7329) となっている。この結果は、ランダムフォレスト (最大 AR 値 : 0.7070) 対比で優位な値であった。なお、決定木の深さを深くすると (表 12 の下側)、検証データの AR 値が低下し、過学習の傾向が確認された^{※38}。

【表 12 : 検証データの AR 値 (勾配ブースティングツリー)】

	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1	0.6521	0.6641	0.6676	0.6755	0.6677	0.6685	0.6737	0.6668	0.6721	0.6671
2	0.7123	0.7165	0.7118	0.7106	0.7115	0.7087	0.7155	0.7140	0.7093	0.7100
3	0.7266	0.7261	0.7225	0.7246	0.7211	0.7254	0.7205	0.7249	0.7150	0.7227
4	0.7329	0.7263	0.7297	0.7183	0.7241	0.7228	0.7210	0.7231	0.7126	0.7194
5	0.7181	0.7186	0.7163	0.7177	0.7139	0.7107	0.7067	0.7003	0.6973	0.6977
6	0.6977	0.6999	0.6941	0.6941	0.7030	0.7060	0.6884	0.6999	0.7019	0.7012

備考 : 縦軸は木の深さ、横軸は学習率。

AR 値が最大となるパラメータ (決定木の深さ 4、学習率 0.01) における指標重要度^{※39}は表 13 の通りである。上位 20 指標中、最多は「出金返済」の 10 指標、次いで、「預金残高」が 8 指標という結果となった。その他、複数のカテゴリの指標が上位となり、「預金残高」と「出金返済」が上位 20 指標を占めていたランダムフォレストとは傾向が異なることが確認された^{※40}。ランダムフォレスト対比で AR 値の上昇が確認された理由として、こうした「預金残高」と「出金返済」以外のより多くの指標の情報も取り込むことによって、モデル精度の向上が果たされたと解釈される。

※37 AR 値が最大となるパラメータ設定 (決定木の深さ : 4、学習率 : 0.01) における決定木の数 (反復回数) は、クロスバリデーションにて 500 回としてモデル構築した。

※38 決定木を深くすると、より粒度細かくトレーニングデータに過度にフィットするようになり、検証データにおけるフィット (AR 値) が低下していると考えられる。

※39 XGBoost における指標重要度はいくつか計算方法が存在するが、ここでは「Gain」 (評価基準の改善度) を用いている。

※40 ランダムフォレストは単指標 AR 値の高い指標が上位であるのに対し、勾配ブースティングツリーでは、単指標の順位とは異なる重要度順位となる傾向があった。こうした傾向は、モデルのロジック (表 6 および補論 1 を参照) の差異に起因していると考えられる。

【表 13：指標重要度（勾配ブースティングツリー）】

重要度 順位	カテゴリ	Gain (比率)	累積比率	重要度 順位	カテゴリ	Gain (比率)	累積比率
1	預金残高	17.7%	17.7%	11	出金返済	1.5%	51.0%
2	出金返済	8.4%	26.0%	12	預金残高	1.3%	52.3%
3	預金残高	5.4%	31.4%	13	出金返済	1.0%	53.4%
4	出金返済	4.6%	36.0%	14	出金原価	0.9%	54.3%
5	預金残高	3.6%	39.6%	15	出金返済	0.9%	55.2%
6	出金返済	2.4%	42.0%	16	出金返済	0.9%	56.1%
7	預金残高	2.0%	44.0%	17	融資残高	0.9%	57.0%
8	預金残高	2.0%	46.0%	18	預金残高	0.8%	57.8%
9	出金返済	1.9%	47.9%	19	出金返済	0.7%	58.5%
10	出金返済	1.7%	49.6%	20	預金残高	0.7%	59.2%

ハ. ロジットモデル

3節(4)に記載した手法を用いてロジットモデルを構築した結果、24 指標^{※41}が採用された（表 14）。

【表 14：ロジットモデルの採用指標と係数推計値】

番号	カテゴリ	係数	番号	カテゴリ	係数
0	(切片項)	12.741	13	出金返済	-0.527
1	預金残高	-0.354	14	出金返済	-0.703
2	預金残高	-0.213	15	出金返済	-0.320
3	預金残高	-0.312	16	預金残高	-0.330
4	預金残高	-0.221	17	預金残高	-0.268
5	出金返済	-0.398	18	出金返済	-0.260
6	預金残高	-0.210	19	出金原価	-0.420
7	出金返済	-0.655	20	出金返済	-0.653
8	出金返済	-0.345	21	出金原価	-0.210
9	出金返済	-0.819	22	入金融資	-0.264
10	預金残高	-0.272	23	入金その他	-0.177
11	出金返済	-0.243	24	出金原価	-0.756
12	出金返済	-0.490			

このとき、検証データにおける AR 値は 0.7113 となり、ランダムフォレスト（最大 AR 値：0.7070）以上、勾配ブースティングツリー（最大 AR 値：0.7329）以下という結果となった。

以上、デフォルト観測機間が 3 ヶ月の結果を確認したが、すべてのデフォルト観測期間における AR 値は、表 15 の通りにまとめられる。

※41 複数指標が選択されたカテゴリは、出金返済（11）、預金残高（8）、出金原価（3）となった。なお、本結果はランダムフォレスト、勾配ブースティングツリーから指標重要度の上位 120 指標ずつを抽出し、係数符号を確認しながらステップワイズ法にて指標選択を行ったもの。

【表 15：検証データにおける各モデルの AR 値】

モデル	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダムフォレスト	0.7611	0.7246	0.7070	0.6830	0.6701
勾配ブースティング	0.7937	0.7441	0.7329	0.7104	0.7015
ロジットモデル	0.7828	0.7401	0.7113	0.6851	0.6820

表 15 より、いずれの期間においても、ロジットモデルはランダムフォレスト対比で優位な結果となり、勾配ブースティングツリー>ロジットモデル>ランダムフォレストの序列となった。ロジットモデルがランダムフォレスト対比で優位となった理由として、モデル構築における指標選択の工夫が効果を発揮していると考えられる。すなわち、ロジットモデル構築に際する採用指標候補は、勾配ブースティングツリーの指標重要度の結果を参照しており、組み合わせとして適切な指標群をベースに指標が選択されていることが理由と考えられる。

(3) モデルのバックテスト

構築したモデルの精度を確認するため、実データを用いたバックテストとして、中堅中小企業の与信ポートフォリオに適用した場合の AR 値を求め、直近 12 ヶ月間の平均値を用いてモデルを検証する。ここでは、すべてのデフォルト観測期間（1・2・3・6・12 ヶ月）についての結果を示す。

イ. 各モデルのバックテスト

表 16 は、バックテストの結果（直近 12 ヶ月間の平均 AR 値）である。いずれのモデルにおいても、デフォルト観測期間が長くなると AR 値が低下する傾向にあり、直感と整合的な結果となった。

機械学習のモデルでは、勾配ブースティングツリーがランダムフォレスト対比で僅かながら高い結果であるが、大きな差は生じていない。なお、機械学習対比、ロジットモデルは僅かながら低いが、水準自体は 0.65～0.75 程度と、一般的な財務スコアリングモデルと比較しても相応の水準を有すると考えられる。

また、デフォルト観測期間が長くなるほど、ロジットモデルと機械学習モデル対比の精度差分が大きくなる傾向が確認された。この結果は、デフォルト観測期間が長いほど、より将来のデフォルトを当てるために、様々な、かつ、多くの情報を使用する必要があることを示唆している^{※42}。

※42 ロジットモデルの採用指標数はいずれも 20 指標程度で、かつ、預金残高・出金返済が大宗を占めており、多くの指標の情報を取り込んでいる機械学習対比で、使用している

【表 16：バックテストにおける各モデルの AR 値】

モデル	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダムフォレスト	0.7749	0.7578	0.7482	0.7429	0.7166
勾配ブースティング	0.7876	0.7797	0.7728	0.7377	0.7244
ロジットモデル	0.7581	0.7608	0.7174	0.6853	0.6669

ロ. メイン・非メイン別に分けた場合の検証

メイン取引先^{※43}と非メイン取引先でモデルの予測精度に違いが生じているかを検証する。バックテストにおけるメイン取引先・非メイン取引先別の AR 値とその差分は表 17 の通りである。いずれのモデル・デフォルト観測期間においてもメイン取引先の AR 値の方が高い結果となっている。メイン取引先は、売上金などより多くの入出金情報が把握できる先であることから、直観と整合的な結果である。但し、非メイン取引先における AR 値も概ね 0.7 以上と相応の水準を確保しており、非メイン先（入出金情報では売上金などが十分に把握できない先）であってもモデルは有効であることが示唆される。

【表 17：メイン・非メイン別の各モデル AR 値】

ランダムフォレスト	デフォルト観測期間				
	1M	2M	3M	6M	12M
メイン	0.7835	0.7980	0.7848	0.7563	0.7542
非メイン	0.7622	0.7347	0.7259	0.7314	0.6866
差分	(+0.0213)	(+0.0633)	(+0.0590)	(+0.0249)	(+0.0676)

勾配ブースティング	1M	2M	3M	6M	12M
メイン	0.8150	0.8014	0.8067	0.7524	0.7468
非メイン	0.7760	0.7680	0.7490	0.7238	0.7012
差分	(+0.0390)	(+0.0334)	(+0.0577)	(+0.0287)	(+0.0456)

ハ. 財務スコアを併用した場合の検証^{※44}

法人債務者のデフォルト予測モデルとして、財務スコアリングモデルが広く

情報が限定されている。その結果、より多くの情報を用いて予測する必要があるデフォルト観測期間が長い問題設定で、モデル精度に大きな差が生じていると考えられる。また、モデルの表現力（非線形性など）でも機械学習モデル対比で劣後することが理由と考えられる。

※43 メイン取引先は、「（入出金情報から計算される）直近 1 年間の入金売上合計が、損益計算書上の売上高の 30%以上を占める先」と定義した。すなわち、決済取引のメイン取引先を特定しており、厳密には、融資取引のメイン取引先とは異なる可能性がある。

※44 各財務指標を多く用いて、入出金情報と共に指標候補として採用しモデルを構築する

使われている。本分析では、財務情報から計算されたスコア（以下、財務スコア）を追加指標候補とした場合の各モデルの精度向上を検証する。

なお、実際に、財務情報を用いて信用リスク評価を行うには、決算書の入手や格付付与作業などに相応の時間を要するため、本稿における財務スコアの更新は、決算時点から3ヶ月間のラグが発生するように設定している。それ以外の時点での財務スコアは、11ヶ月間、横置きとしている（同じ値の財務スコアが12ヶ月間保持される）。

財務スコアを併用したバックテストにおけるAR値は表18の通りとなり、3ヶ月のデフォルト観測期間では各モデルで+0.1程度AR値が上昇している^{※45}。AR値の水準については、勾配ブースティングツリーやロジットモデルで0.8台後半と、モデル精度としては極めて高い水準にあることがわかる。

【表18：財務スコアリング計数を併用した各モデルのAR値】

モデル	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダム フォレスト	0.8073 (+0.032)	0.8384 (+0.081)	0.8416 (+0.093)	0.8367 (+0.094)	0.8193 (+0.103)
勾配ブース ティング	0.8629 (+0.075)	0.8712 (+0.092)	0.8650 (+0.092)	0.8551 (+0.117)	0.8244 (+0.100)
ロジット モデル	0.8393 (+0.081)	0.8697 (+0.109)	0.8537 (+0.136)	0.8555 (+0.170)	0.8290 (+0.162)

備考：各セル下段（ ）内の数字は、財務スコアを用いない場合（表16）との差分。

機械学習モデルの中で優位な勾配ブースティングツリーを用いて、財務スコアを併用したモデルでの指標重要度を分析すると、表19の結果が得られた。

アプローチも考えられるが、ここでは財務情報を集約した財務スコア（1指標）のみを用いている。なお、財務スコアは、CRD協会の財務スコアリングモデルを利用した。
^{※45} ロジットモデルの方が上昇度が大きい理由として、以下の2点が考えられる。第1に、財務スコアを用いない場合の水準（表16）に差があり、ロジットモデルの方が上昇余地が大きいこと、第2に、ツリー系のモデルは指標をある閾値で二分割し、スコアリングの序列の情報を取り込みづらい（スコアを連続的な情報として捕捉しにくい）一方、ロジットモデルはスコア全体の序列の情報を反映できること、である。

【表 19：財務スコアを併用した場合の指標重要度（勾配ブースティングツリー）】

重要度 順位	カテゴリ	Gain (比率)	累積比率	重要度 順位	カテゴリ	Gain (比率)	累積比率
1	財務スコア	56.5%	56.5%	11	預金残高	0.7%	70.5%
2	預金残高	3.6%	60.2%	12	出金返済	0.7%	71.2%
3	出金返済	1.6%	61.8%	13	出金返済	0.5%	71.8%
4	出金返済	1.6%	63.3%	14	出金原価	0.5%	72.3%
5	出金返済	1.5%	64.8%	15	融資残高	0.5%	72.8%
6	出金原価	1.3%	66.1%	16	入金融資	0.5%	73.2%
7	預金残高	1.1%	67.2%	17	出金返済	0.4%	73.7%
8	出金返済	0.9%	68.2%	18	出金返済	0.4%	74.1%
9	預金残高	0.9%	69.0%	19	出金返済	0.4%	74.5%
10	出金返済	0.8%	69.8%	20	出金返済	0.3%	74.8%

財務スコアを併用しない表 13 の結果と比較すると、「預金残高」の指標重要度（順位）が低下する一方、「出金返済」の指標重要度は上位に残る傾向が確認された^{※46}。その理由としては、以下の2つの仮説が考えられる。

① 「預金残高」と財務スコアリングモデルとの関係

「預金残高」は、企業の「規模」の情報も有しており、現預金や企業規模を評価する財務モデルと相関が高い。従って、財務スコアを併用したことによって「預金残高」の指標の重要度が低下した。

② 「出金返済」と財務スコアリングモデルとの関係

入出金情報における「出金返済」は、貸出利息額だけでなく元本返済額も合算した値であり、金額ベースでみると、「出金返済」の情報の多くは、利息分よりも元本の比率が高いと考えられる。今般使用した財務モデルにおける支払能力を評価する指標は、インタレストカバレッジレシオ（インカバレシオ）のような収益と返済利息との関係性から支払能力を評価している。一方、入出金情報における「出金返済」は、返済金額の大小を直接評価したものであるため、財務モデルにおける支払能力を評価する指標との（情報としての）相関はそれほど高くはない。従って、「出金返済」の指標の重要度は低下しなかった。

5. おわりに

本稿の結びとして、分析結果のまとめ、分析を通じて得られた知見及び今後の課題を述べる。

※46 重要度上位 20 指標中、財務スコア追加前に含まれるのは、出金返済 10、預金残高 8 であったのに対し（表 13）、追加後は出金返済 11、預金残高 4 であった。

(分析結果のまとめ)

第 1 に、入出金情報を用いた信用リスク評価の分析結果として、機械学習モデルでは、いずれのデフォルト観測期間においても AR 値が 0.7 以上と、予兆管理モデルとして十分に活用可能な水準であることが確認された (表 16)。

バックテスト (表 16) の結果から、機械学習モデルとロジットモデルの AR 値の比較においては、ロジットモデルは機械学習モデルよりも僅かに低い結果であった。また、より多くの情報を使用してデフォルトを予測する必要があると考えられる観測期間が長いケースでは、機械学習モデルとロジットモデルの差分が大きくなる傾向が確認された。

但し、ロジットモデルは、モデル精度は機械学習対比で僅かに低いものの、AR 値の水準は (最もデフォルト期間が長く、精度が最も低い場合でも) 0.65 以上を有するなど、相応に高い水準であることも確認された。従って、モデルの解釈性の高さ、指標影響度のわかりやすさといった運用上の利点を勘案すれば、予兆管理モデルとしてロジットモデルを採用する、という判断も合理的と考えられる^{※47}。

第 2 に、メイン取引先・非メイン取引先別の精度比較では、メイン取引を行っている債務者は、非メイン債務者対比でデフォルト予測精度が高い結果が確認された (表 17)。メイン取引先では、より多くの入出金情報が捕捉可能となるため、本結果が得られたものと考えられる。

但し、非メイン取引先においても相応の予測精度が確保された。金融機関にとって、メイン行である場合に比べて与信額が小さいことが多い非メインの取引先に対し、営業担当者 (RM) が電話や訪問等を通じて行うモニタリングは、一般に費用対効果の点で見合わない。このため、本分析で検証した信用リスク評価モデルの算出結果をモニタリングすることで、担当先に電話・訪問等を行わずとも予兆管理が可能であるとする、非メイン先に対する予兆管理ツールとしても有用性が高いと考えられる。

第 3 に、入出金情報に加え財務スコアを併用した結果、AR 値が (入出金情報のみを使用した場合に比べて) 約 0.1 上昇し、0.8 台後半となり、極めて高い予測精度が実現された (表 18)。入出金情報のみでも予測精度として十分な水準 (AR 値 0.7 以上) であるが、財務情報を加えることにより、即時性と予測精度の双方を兼ね備えた予兆管理モデルの構築が可能であることがわかった。但し、

^{※47} 本分析におけるロジットモデル構築は、機械学習の結果 (指標重要度) をベースに定量的に指標候補を抽出した後に指標選択を行った。ロジットモデルの構築において、定性的な判断を介さず、定量的に入出金情報を選択するには、何らかの工夫が必要と考えられる。例えば、根本ほか (2019) では、主成分分析により指標の情報抽出を行っている。

財務情報は、(更新頻度が1年に1度などの)決算時点でのみ更新が行われる点に留意が必要である。

(本分析を通じて得られた追加的な知見)

また、本分析を通じて以下のような知見が得られた。

機械学習モデルは、パラメータのチューニングに試行錯誤が必要であり、また、一回の計算時間も相応に要するため、時間的コストを要した。特に、勾配ブースティングツリーは過学習しやすい傾向にあるため、チューニングを丁寧に行う必要があった。

一方、ロジットモデルの構築に際しては、今般の分析では機械学習モデルのアウトプット(指標重要度)をベースに定量的に指標選択を行い、指標選択にかかるコストを小さくした。しかし、一般には、定量的な指標精度や組み合わせの最適性だけでなく、エキスパートの知見や定性的な判断を要して指標選択を行ったり、また、より適切に情報を抽出するために指標加工に工数を要する場合もあるなど、様々なリソースが必要になる^{※48}。

このように、機械学習とロジットモデルのモデル構築においては、それぞれコストを要する観点・項目が異なるため、担当者・担当チームの各人の専門性やリソースの多寡などを踏まえつつ、採用するモデルを適宜検討していく必要があると考えられる。

(今後の検討課題)

今後の検討課題としては、以下の4点が挙げられる。

① 摘要グループの作成方法高度化

本稿の分析では、入出金情報から、より精緻に情報を抽出するために摘要グループを10個に分類しモデル構築を行った。このようなグルーピングは、入出金情報をいかに適切に抽出できるかという観点で重要である^{※49}。今般の分析における摘要グループへの集約は、リスク管理や融資審査などの実務経験に基づいた定性的な判断を基に集約したが、より定量的なアプローチを用いて高度化する余地があると考えられる。例えば、現在採用している摘要グループの中でデフォルト予測に有効なグループと有効でないグループの濃淡を確認し、有効なグ

^{※48} 影井ほか(2016)では、金融マーケティングのデータを用いて、ロジットモデルとディープラーニングを工数及び精度の面で比較を実施。工数全体の多くを占める、エキスパートによる指標作成の工数が大きく削減されつつ、精度はほぼ同程度であることを報告している。

^{※49} 補論2では、入金または出金の粒度に集約した場合のモデル精度への影響度を検証し、AR値が0.06~0.1程度低下することが確認されている。

ループの粒度をより細かくすることなどが考えられる^{※50}。また、クラスタリング^{※51}などの手法を用いて、入出金情報を定量的な手法にて集約するアプローチも考えられる。

② 景気後退局面におけるモデルの挙動確認

本稿の分析では 2014 年以降の入出金情報を分析対象としたが、この期間中、国内のデフォルト発生率は低水準にあったことから、景気循環によるモデルの挙動変化を検証できていない。特に、企業のデフォルトが多く発生したリーマンショック並みの景気後退局面におけるモデルの挙動変化や頑健性については検証が出来ておらず、当時の入出金情報及びデフォルトデータに基づいて確認するなどの検証の必要がある。

③ 業種・規模等によるセグメント別の分析

本稿のモデル構築では、メイン取引先・非メイン取引先別のモデル精度比較を行ったが、実務的にはメイン・非メイン以外のセグメントで分けた分析を行う必要がある^{※52}。例えば、本モデル構築では全業種共通のモデル構築を行ったが、業種間での入出金の傾向やデフォルト発生との関係性の差異を検証したうえで、業種を分けてモデル構築を行うことも検討した方がよい。その他、企業規模など、セグメントに分けた時のモデル精度や入出金の傾向等を考察し、モデルをセグメント別に構築するか検討が必要であろう。

④ ディープラーニングによるモデル構築

本稿の分析では、機械学習モデルとして、ランダムフォレスト及び勾配ブースティングツリーによるモデルを採用したが、ディープラーニングのモデル構築を試行しなかった。このため、表現力の高いディープラーニングをモデルとして採用することも検討の余地がある^{※53}。

※50 例えば、今般の分析結果では、摘要グループ「出金返済」のデフォルト予測が高いとの結果が得られている。「出金返済」には、融資返済だけでなく、貸出利息や保証料などの出金も含まれており、より細かい粒度の情報を用いてデフォルトとの相関等を検証しながら適切なグルーピングを探索することによって、より精緻な集約を行うことが可能と考えられる。

※51 教師なし学習の一手法。定量的に類似性の高い集団にグルーピングし、既定個数のグループに集約する手法。K-means 法などが広く使われている。詳細は、石井・上田（2014）などを参照。

※52 デフォルト予測モデルを構築する場合の多くは、デフォルト件数が限定的である。従って、セグメント分けした場合に一部セグメントでデフォルト件数が僅少となり、モデル構築が困難となる可能性がある。セグメント分けを検討する際は、デフォルト件数とセグメント分けの粒度や必要性との兼ね合いで検討がなされるべきであろう。

※53 例えば、本田ほか（2018）では、入出金データにクロスリカレンスプロットと呼ばれ

ディープラーニングは画像認識に用いられることがあるが、その場合、画像の特徴（量）を機械的に抽出し、特定のパターンの特徴に対する適合度によって画像認識を行う。但し、特徴量は解釈が困難で、意味を有しないことが多く、信用リスク管理において、このような画像認識的アプローチが適切か検討を行う必要がある。また、ディープラーニングは、機械学習以上にモデルが結果を導出した根拠を示すことが困難で、モデルがブラックボックス化しやすい^{※54}。解釈性、説明性が求められる信用リスク管理において実用化が可能か、また、そもそも問題設定として、デフォルト予測を行うために必要なモデルの複雑さ^{※55}はどの程度か、などの観点で検討を行う必要がある。

る指標加工を行い、入出金情報を画像認識として捉えたいうえで、CNNによるモデル構築を行っている。

※54 信用リスク評価の文脈では、大久保・尾藤（2018）にこの点が記載されている。

※55 一般に、ディープラーニングが有効な場合は、指標（ファクター）に対してデフォルト発生が複雑な非線形関数になっているケースである。このため、「預金残高」や「出金返済」など一部の（入出金）情報を捕捉すればデフォルトが概ね予想できるような状況では、深層化するメリットは大きくない可能性が高いと考えられる。

補論 1. 機械学習のモデル

ランダムフォレスト及び勾配ブースティングツリー (XGBoost^{※56}) は、いずれも決定木をベースにしたアンサンブル学習モデルである。以下では、最初に決定木、続いてアンサンブル学習の概要を説明した後、使用する 2 つのモデルの特徴等を整理する。

(1) 決定木^{※57}

特定の変数の値 (またはカテゴリ) を閾値 (分割点) に応じて 2 つの集団^{※58} に分類し、この 2 つの集団を、別の変数を用いてさらに 2 つの集団に分類することを繰り返し、最終的に 2 値 (例えばデフォルト・非デフォルト) に分類 (判別) する手法・モデルを「決定木」と呼ぶ。

各企業をデフォルト・非デフォルトに分類する場合の決定木の例 (イメージ) として、図 A-1 のような決定木が考えられる^{※59}。なお、終端のデフォルト/非デフォルトは (実際のデフォルト有無を問わず)、本集団に分割された場合に、当該企業をデフォルト/非デフォルトと分類 (判別) することを表している。

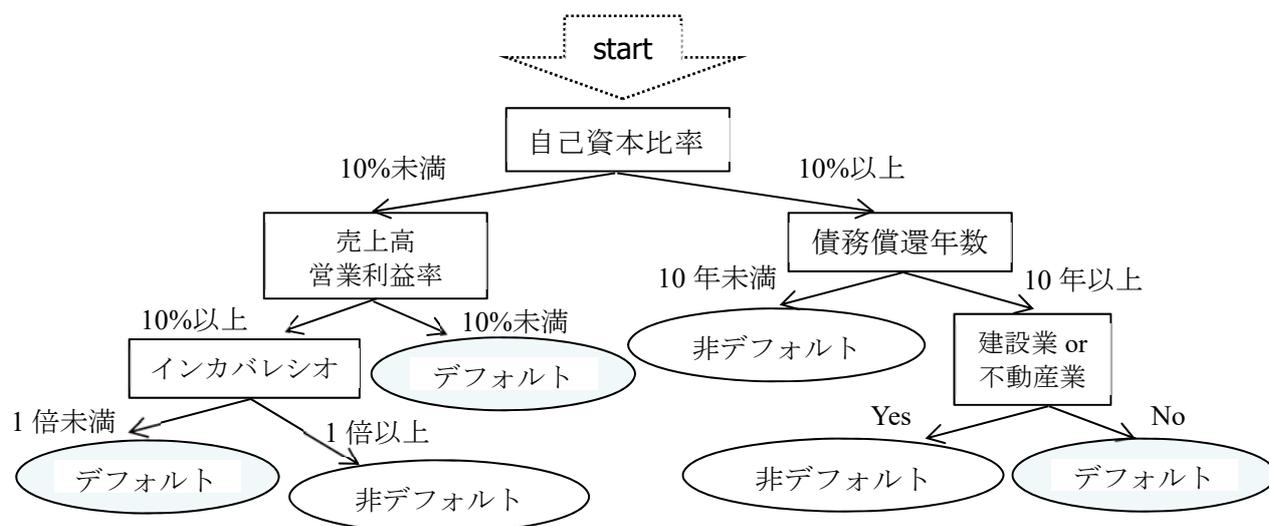
※56 XGBoost (eXtreme Gradient Boosting) は勾配ブースティングツリーのアルゴリズムが実装されているモデル・ライブラリ。通常の勾配ブースティング対比で高速処理が可能。詳細は Chen and Guestrin (2016) を参照。その他に勾配ブースティングツリーを高速かつ高精度に計算するアルゴリズム・モデルがいくつか提案、開発されている。例えば、米国マイクロソフト社が開発した LightGBM もその一つ。詳細は Ke ほか (2017) などを参照。

※57 分類 (判別) 問題に適用する決定木を分類木 (classification tree) と呼ぶ。また、回帰問題に適用する決定木は回帰木 (regression tree) と呼ばれる。

※58 ここでは、(デフォルト・非デフォルトの) 2 つの集団に分類することを想定している (2 分木と呼ばれる) が、3 クラス以上 (多クラス) への分類を行う決定木 (多分木と呼ばれる) も構築が可能。

※59 大久保・尾藤 (2018) を参考として作成。なお、記載の財務項目・属性及びその閾値 (分割点) 等はすべて架空のもの。

【図 A-1：デフォルト判定に用いられる決定木の例（イメージ）】



決定木を構築する際、指標【どの指標で区切るか】と閾値（分割点）【どの値で区切るか】を決める必要がある。実際には、採用候補のすべての指標について、取りうるすべての閾値で区切った場合^{※60}を想定し、判別精度がどの程度向上したかを「不純度（impurity）^{※61}」と呼ばれる評価関数を用いて評価し、分割すべき指標及び閾値を決定しながら決定木を成長させる^{※62}。

決定木を用いる利点として、①機械学習モデルの中では相対的に解釈がしやすい（可読性が高い）、②外れ値に対して頑健^{※63}（そのため、外れ値処理等を行う必要性が低い）、③数値変数とカテゴリ変数が混合したデータにも適用可能、

※60 ある変数が連続的な値をとる場合、データ数が N であれば、 $N - 1$ 個の離散的な分割候補点を取りうるが、隣り合った2点の間をその特徴軸の分割候補点とすることが多い。

※61 異なるグループ（デフォルトまたは非デフォルト）が多く混在する集団は不純度が高く、いずれか一方の集団で構成されるほど不純度が低い。一般に、決定木は分割後の集団の不純度が一番小さくなるような指標及び閾値を選んで分割していく。なお、不純度を表す主な関数として、①Gini係数、②交差エントロピー、③誤分類率、などがあり、①または②が使用されることが多い。詳細はHastieほか（2014）参照。

※62 決定木をどこまで成長させるか（決定木の深さをどこまで深くするか）という観点では、より決定木を深くした方が粒度細かく分類できる一方、過学習が発生しやすくなる傾向があるため、どこまで決定木を深くすべきか検討が必要である。この点に関する一般的な検討手法は平井（2012）を参照。

※63 例えば、図 A-1 における最初の分岐である自己資本比率において、10%以上であれば（大きな外れ値であっても）同じ評価を行うことが可能である。一方で、ロジットモデルなどで外れ値処理などをせずにモデル構築（最尤推定法などを用いたパラメータ推計）を行うと、外れ値が含まれている場合はパラメータが影響を受けやすい。なお、本分析におけるロジットモデル構築では、脚注 26 の通り、指標値をスコア化することによって外れ値の影響を軽微にするような加工を行っている。

といった点がある。

一方で、決定木の弱点としては、①判別性能の低さ、②過学習の問題がある。これらの課題を解決するためのアイデアとして、複数の決定木の出力した結果の合算（多数決）で最終的な出力を決定する「アンサンブル学習」と呼ばれる手法がある。

(2) アンサンブル学習^{※64}

一つの決定木だけでは精度が低く、過学習のおそれがあるため、多数の決定木を作成し、それらの多数決によって最終的な予測を行う手法がアンサンブル学習である。一般に、アンサンブル学習は、「弱い（判別精度の低い）学習器を組み合わせることによって強い（判別精度の高い）モデルを構築する」手法であり、学習器の構築方法の差異によって、表 A-1 のように、①バギング (Bagging^{※65})、及び②ブースティング (Boosting^{※66}) と呼ばれるモデル (アルゴリズム) がある。

【表 A-1：アンサンブル学習の代表例】

アンサンブル学習の種類	モデル構築手法の概要
バギング (Bagging)	トレーニングデータのブートストラップサンプル ^{※67} を用いて複数の決定木を（並列的に）学習する方法【例：ランダムフォレスト】。
ブースティング (Boosting)	複数の決定木を学習する際、前までの決定木の結果を参照しながら一つずつ（直列的に）決定木を学習する方法【例：勾配ブースティング】。

※64 アンサンブル学習の詳細は、平井（2012）、杉山（2013）などを参照。

※65 **Bootstrap AGG**regat**ING** の略。名称の通り、Bootstrap（ブートストラップ）法で得られた結果を合算（aggregate）する手法。なお、ブートストラップ法については、脚注 67 を参照。

※66 詳細は Hastie ほか（2014）参照。

※67 ブートストラップ法とは、 N 個のデータから重複を許して（「復元抽出」） $n(\leq N)$ 個データをサンプリングすることにより、元の集団と少し異なる集団を作成する手法。本手法を用いることにより、例えば、（集団の異なる）モデル構築用データを複数作成することができ、統計量の信頼区間の推計などに用いられる。

なお、復元抽出を行うので、 n 個のデータのなかには、複数回サンプリングされるデータと一度もサンプリングされないデータが生じ得る。一つのデータが n 回の復元抽出で少なくとも1度サンプリングされる確率は（ n を大きくした場合）約63.2%。逆に、 $1 - 63.2\% = 36.8\%$ の確率で一度もブートストラップサンプリングで抽出されないことになる。ランダムフォレストにおいて、これら全体の約1/3の抽出されないデータをアウトオブサンプルとみなし、決定木の汎化性能を疑似的に評価する Out-Of-Bag (OOB) 誤り率という考え方がある（詳細は、平井（2012）を参照）。

(3) 本稿で使用する機械学習モデル

本稿の分析で使用するアンサンブル学習のモデルである、ランダムフォレスト、勾配ブースティングツリーは以下のように整理できる。

イ. ランダムフォレスト

表 A-1 の通り、ランダムフォレストは決定木を用いたバギングの一つである。バギングは、決定木のようにトレーニングデータに対するバイアスが小さく、分散が大きな弱学習器に適した手法であり^{※68}、分散をより小さくするようなアルゴリズムを通じてモデルが構築される。

ランダムフォレストを構築する際、ブートストラップサンプリングによって対象サンプルを変えながら各決定木を生成する。但し、ブートストラップサンプリングによって対象サンプルを変えて各決定木を構築しても、生成される決定木間の相関は高いことが多い。そこで、ランダムフォレストにおけるモデル構築では、対象サンプルだけでなく指標採用候補もサンプリングすることによって、生成される決定木間の相関をより小さくするための工夫がなされている。

一般に、分散 σ^2 をもつ M 個の独立な確率変数 $X_i (i = 1, 2, \dots, M)$ の平均 $\bar{X} = (1/M) \sum_{i=1}^M X_i$ の分散は、 $\text{Var}(\bar{X}) = \sigma^2/M$ となるが、任意の2つの確率変数間に正の相関 ρ がある場合には、平均 \bar{X} の分散は、

$$\text{Var}(\bar{X}) = \frac{1-\rho}{M} \sigma^2 + \rho \sigma^2$$

となる。いま、 X_i を一つの決定木から算出される予測結果とすると、ブートストラップの回数 M を増やした場合、上式第1項は小さくなるが、第2項は不変である。そこで、ランダムフォレストは、 ρ をより小さくし、平均の分散を小さくするための仕組みが施されている。具体的には、対象サンプルだけでなく、対象指標のサンプリングも併せて行うことによって、 ρ をより小さくする、バギングを強化したモデルである。

ランダムフォレストは、以下①～③のプロセスを一定回数（以下では M 回）独立に繰り返し、（ M 個の）決定木を構築することによって作成される^{※69}（イメー

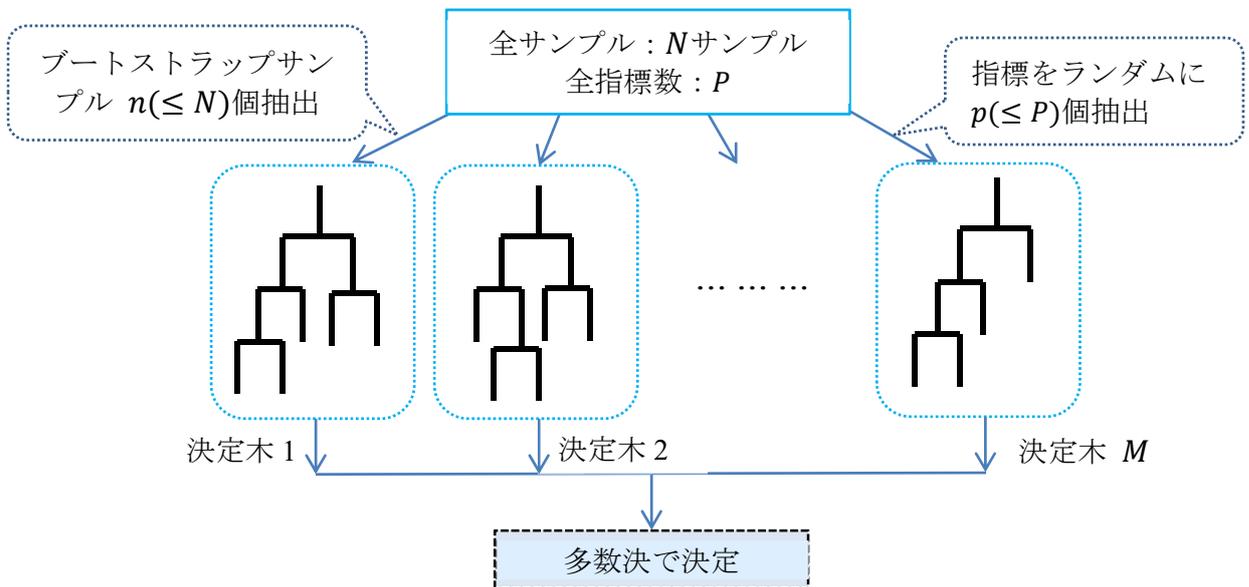
^{※68} モデルのバイアスとはトレーニングデータにおける予測値と実際の値との乖離具合、モデルの分散とはトレーニングデータが変化した場合に構築されるモデルの変化度を指す。一般に、複雑度が低い（例：線形な）モデルほどバイアス大・分散小、複雑度が高い（例：非線形な）モデルほどバイアス小・分散大となる（「バイアスと分散のトレードオフ」と呼ばれる）。

^{※69} ここではアルゴリズムのイメージを記載している。数学的な表現等の詳細は平井（2012）または Hastie ほか（2014）を参照。

ジは図 A-2)。

- ① 全 N サンプルから、ランダムサンプリングにより n 個のデータで構成されるサブサンプルを生成 (ブートストラップサンプル)。
- ② 全 P 指標から、一定の指標数 (p 個) の指標をランダムに抽出 (指標ユニバースの作成)。
- ③ 指定したノード数に達するまで、サブサンプルを最も良く分類する指標及び閾値を用いて、決定木を構築。

【図 A-2：ランダムフォレストのモデル構築イメージ】



ランダムフォレストにおけるチューニングパラメータは、決定木を由来とするもの^{※70}とランダムフォレスト特有のものがある。本稿では、後者のパラメータに着目し、決定木の数 (M) と指標ユニバース数^{※71} (p) をチューニングすることによって最適なモデルを構築することとした。

ロ. 勾配ブースティングツリー (XGBoost)

ブースティングの考え方は、大量の弱学習器 (判別力の低い学習器) の出力を組み合わせて強力な学習器を構築するというものである。具体的には、弱学習器で

※70 不純度の評価関数、決定木の深さ、などがある。

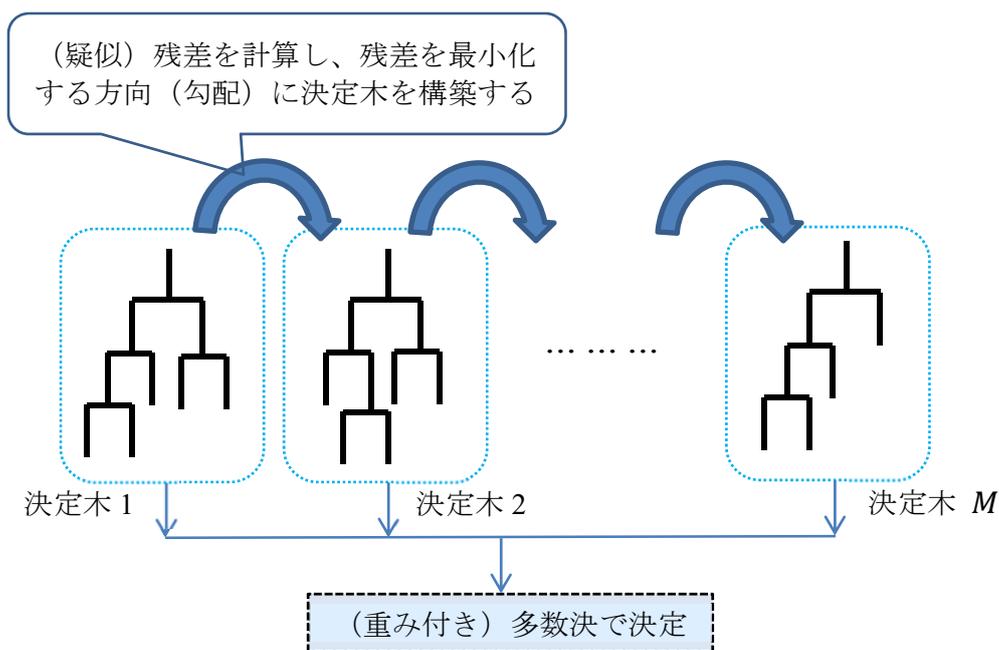
※71 一般的な経験則として、指標ユニバースは、分類では \sqrt{P} 、回帰では $P/3$ が妥当な水準とされている。

判別した結果から生じる残差^{※72}（予測値と実績値の差分）を損失関数とし、この損失関数（残差）を小さくするための弱学習器を逐次的に生成して、最終的に残差が極めて小さいモデルを構築する。なお、損失関数の「勾配」を用いて損失関数を減少させる決定木を追加し、モデルが更新され、加えて、弱学習器として決定木（ツリー）を用いているため、「勾配ブースティングツリー」と呼ばれる。

勾配ブースティングツリーは以下②、③のプロセスを一定回数（以下では M 回）逐次的に繰り返すことにより作成する^{※73}（イメージは図 A-3）。

- ①：初期値として、全サンプルに対して損失関数^{※74}を最小化する値を設定。
- ②：各時点でのモデルの予測値と実績値の残差を目的変数として、より残差を小さくする決定木を構築。
- ③：②で作成した決定木から推計される予測値を加算。

【図 A-3：勾配ブースティングツリーのモデル構築イメージ】



※72 一般に、残差とは回帰モデルにおける予測値と実績値の差分として定義されるが、用いる損失関数によっては、残差そのものでなく、概念的に残差に近い定義のものもあるため、その場合は疑似残差などと呼ばれる。損失関数として逸脱度を採用した場合は、実績値（デフォルトを1、非デフォルトを0とするフラグ）と予測値の差分であるが、ここでは（「疑似」を省略し）単に「残差」と表記している。

※73 脚注 69 と同様、ここではアルゴリズムのイメージを記載している。数学的な表現等の詳細は平井（2012）または Hastie ほか（2014）を参照。

※74 本分析では、「逸脱度」と呼ばれる（デフォルト・非デフォルトの実際の値（デフォルトを1、非デフォルトを0とするフラグ）とデフォルト確率の予測値との差分で定義される）関数を採用している。詳細は Hastie ほか（2014）を参照。

勾配ブースティングツリーにおけるチューニングパラメータには、決定木を由来とするものとブースティング特有のものがある。本分析においては、前者のパラメータとして決定木の深さ、後者のパラメータとして学習率 η ^{※75}をチューニングすることによって最適なモデルを構築することとした^{※76}。

※75 決定木を足し合わせるときに使う係数。学習率を大きくすると一つの決定木の予測値を強く反映するため、一般に収束するまでの決定木の数が少なくなり、学習にかかる時間は短くなる。一方で、学習率が大きいと残差に対して強く学習するモデルとなり、過学習が生じやすくなるため、チューニングが必要となる。

※76 各パラメータセットにおける決定木の数（反復回数）は、クロスバリデーションによって決定した。

補論 2. 追加検証結果

本文中の分析の追加検証として、デフォルト予測精度に与える影響が大きい情報を特定するために、現在使用しているデータ（情報）から、徐々に情報量を落とし、どのような影響が生じるかを検証した。追加検証項目は表 A-2 のようにまとめられる。

【表 A-2：追加検証項目】

No.	検証項目	備考
①	利用する入出金情報の期間短縮	・ 24ヶ月分用いている入出金情報データを、13・6・3ヶ月分のみ利用した場合のモデル精度の影響を検証
②	「預金残高」指標を除く	・ 指標重要度が最も高い「預金残高」（及び預金残高の指標バリエーション）を除いた場合のモデル精度の影響を検証
③	摘要グループを集約	・ 10個の摘要グループ（入金4・出金6） ^{※77} を採用しているが、「入金」及び「出金」に集約した場合のモデル精度の影響を検証（預金残高の指標は不変） ・ 指標バリエーションは、24・12・6・3ヶ月の各期間における平均・標準偏差を採用。また「出金÷入金」の指標バリエーションも採用

以下では、バックテストにおけるモデル精度の変化を記載することによって、各種変更の影響度を確認する。なお、比較を行うベースとなる結果は、本文の表 16 で与えられている。

① 利用する入出金情報の期間を短縮した場合

本文中の分析では 24ヶ月間の入出金情報を用いてモデル構築を行った。しかし、実務的なデータ利用を考えると、各金融機関が一時的に保存しているデータは 13ヶ月程度のことが多いと考えられるため、本期間の入出金情報を採用した場合のモデル精度を検証した。また、入出金情報をより短期間にした場合のモデル精度影響度を検証し、どの程度の期間の情報があればデフォルト予測精度が担保されるか確認した。

入出金情報を 13ヶ月、6ヶ月、3ヶ月と短縮した場合のモデルでの AR 値と減少幅は表 A-3 のようになり、AR 値の減少幅は最大（勾配ブースティングツリー：入出金 3M、デフォルト観測期間 3M の結果）でも 0.02 程度であった。すなわち、今般の分析では、入出金情報を短縮してもモデル精度の低下は限定的であ

※77 詳細は表 2 を参照。

ることが確認された^{※78}。

但し、この結果の解釈には留意すべき点がある。今般の結果は、一定期間の入出金取引実績がある先（ある程度入出金情報が安定的に推移していると想定される債務者）の入出金情報を断面で見た場合の結果である。すなわち、創業間もないスタートアップ企業のように、入出金の安定していない先においても、短期間の入出金情報で信用リスク評価が十分か否かについては、（入出金期間が3ヶ月など）より短いほど留意し、別途の検証が必要である。

【表 A-3：入出金情報期間を短縮した場合の影響度（上から、13M・6M・3M）】

入出金13M	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダムフォレスト	0.7644 (-0.01)	0.7595 (+0.002)	0.7548 (+0.007)	0.7318 (-0.011)	0.7172 (+0.001)
勾配ブースティング	0.7755 (-0.012)	0.7817 (+0.002)	0.7680 (-0.005)	0.7418 (+0.004)	0.7215 (-0.003)

入出金6M	1M	2M	3M	6M	12M
	ランダムフォレスト	0.7739 (-0.001)	0.7669 (+0.009)	0.7549 (+0.007)	0.7441 (+0.001)
勾配ブースティング	0.7710 (-0.017)	0.7732 (-0.007)	0.7617 (-0.011)	0.7279 (-0.01)	0.7085 (-0.016)

入出金3M	1M	2M	3M	6M	12M
	ランダムフォレスト	0.7658 (-0.009)	0.7640 (+0.006)	0.7578 (+0.01)	0.7379 (-0.005)
勾配ブースティング	0.7703 (-0.017)	0.7779 (-0.002)	0.7520 (-0.021)	0.7326 (-0.005)	0.7172 (-0.007)

② 「預金残高」指標を除いた場合

信用リスク管理において、債務者の預金残高はすでにモニタリングの対象とされており、預金残高に著変があった場合にアラートを出すような予兆管理システムが実装されている場合が考えられる。そのような場合、預金残高以外の情報を基にしたアラートがモニタリングツールとして機能すると、予兆管理上も有用であると考えられる。そこで、「預金残高」（及び預金残高を加工した）指標をすべて取り除いてモデル構築を行い、予測精度を検証した。

「預金残高」とその加工指標をすべて取り除いたモデルの AR 値と減少幅は

^{※78} 入出金情報の期間を短縮し、使用できる指標・情報を削減したにも関わらず、一部 AR 値がわずかに向上しているパターンが見られる。これは、バックテストに適用した際のわずかなモデル精度の向上や、乱数の影響などが考えられるが、いずれも上昇度は微小なものである。

表 A-4 のようになり、AR 値の減少幅は約 0.07～0.1 程度であった。相応のデータ件数を有する同一データを用いて AR 値が 0.1 の差分が生じることは誤差の範囲とは考え難く^{※79}、無視できないモデル精度の低下と考えられる。すなわち、「預金残高」を除いた場合、モデル精度は相応に低下することが確認された。

なお、「預金残高」を除いたことにより、指標重要度は「出金返済」や「入金売上」などが上位となった。この結果は、表 7 の結果と整合的であり、「預金残高」より下位の指標が繰り上がったことを意味する。このことから、「預金残高」を代替するような情報を有する指標は他に存在せず、預金残高を用いたアラートとは異なる情報（観点）に基づくアラートであることが示唆される。加えて、AR 値の水準は 0.6 台中盤を確保しており、デフォルト予測能力自体は一定程度確保されていることから、本アラートは実務的に有用と考えられる。

【表 A-4：預金残高の指標を除いた場合の影響度】

②預金残高除き	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダムフォレスト	0.6735 (-0.101)	0.6640 (-0.094)	0.6672 (-0.081)	0.6483 (-0.095)	0.6393 (-0.077)
勾配ブースティング	0.6916 (-0.096)	0.7101 (-0.07)	0.6893 (-0.084)	0.6671 (-0.071)	0.6566 (-0.068)

③ 摘要グループを集約した場合

本文中の分析では、入出金情報から、より精緻に情報を抽出することを企図して摘要グループを 10 個に分類しモデル構築を行った。摘要グループの分類に際しては、分類の仕方や区分数など、定量的な分析や、知見・ノウハウに基づいた判断を経て入出金情報を集約する必要がある、相応にコストを要する。

そこで、入出金情報の摘要グループへの集約がどの程度モデル精度に影響を与えるかを検証するための例として、「入金」及び「出金」の各摘要グループを集約し、「入金」または「出金」という 2 つの情報にすべて集約（表 2 の No.1～4 を入金、No.5～10 を出金と集約）した場合を想定しモデル構築を行った。なお、「預金残高」の取り扱い是不変とした。

結果は、表 A-5 の通り、AR 値の減少幅は約 0.06～0.1 程度であった。すなわち、本稿の分析では、摘要グループを集約した場合、モデル精度の低下は相応にあることが確認された。このことは、本稿で集約したような摘要グループに集約することはモデル精度向上に一定の寄与があり、「入金」または「出金」とい

^{※79} AR 値の信頼区間については、山下・三浦（2011）や Engelmann and Rauhmeier（2006）を参照。

た粒度の情報をを用いることではデフォルト予測に必要な入出金情報を十分に抽出できていないことを示唆している^{※80}。

但し、「入金」または「出金」の粒度でしか入出金情報を入手できない場合を想定すると、AR 値が 0.6 以上を有していることから、予兆管理において本モデルを用いたモニタリングも相応に有効であると考えられる。

【表 A-5：摘要グループを集約した場合の影響度】

③摘要グループ集約	デフォルト観測期間				
	1M	2M	3M	6M	12M
ランダムフォレスト	0.6903 (-0.085)	0.6843 (-0.074)	0.6807 (-0.067)	0.6687 (-0.074)	0.6289 (-0.088)
勾配ブースティング	0.7023 (-0.085)	0.7038 (-0.076)	0.7064 (-0.066)	0.6764 (-0.061)	0.6227 (-0.102)

^{※80} 例えば、出金返済（金額）が出金額全体に占める割合が大きい場合、出金額全体を指標として用いると、出金返済の情報を適切に抽出できない、など。

【参考文献】

- 石井健一郎・上田修功（2014）『続・わかりやすいパターン認識 教師なし学習入門』, オーム社.
- 大久保豊・尾藤剛（2019）『【究解】信用リスク管理』, 金融財政事情研究会.
- 尾木研三（2017）『スコアリングモデルの基礎知識 中小企業融資における見方・使い方』, 金融財政事情研究会.
- 影井智宏・友永康之・松下伴理（2016）「金融行動に対する人工知能の実証研究—金融マーケティングプロセスにおける人工知能の実務展開にむけて—」, 日本マーケティング学会 カンファレンス・プロシーディングス Vol.5, pp.197–208.
- 澤木太郎・田中拓哉・笠原亮介（2018）「機械学習による中小企業の信用スコアリングモデルの構築」, 金融情報学研究会第19回予稿集, pp.20–24.
- 杉山将（2013）『イラストで学ぶ機械学習』, 講談社.
- 日本銀行金融機構局金融高度化センター（2019）「AIを活用した金融の高度化に関するワークショップ」第3回「信用評価」（2019年2月18日開催）.
https://www.boj.or.jp/announcements/release_2019/rel190215d.htm/
- 根本直子・吉野直行・大久保豊・稲葉大明・柳澤健太郎（2019）「銀行口座動態情報を活用した中小企業の信用リスク分析」, 『現代ファイナンス』 No.40, pp.49–71.
- 平井有三（2012）『はじめてのパターン認識』, 森北出版.
- 本田大悟・大古田俊介・井實康幸（2018）「入出金データを用いた企業デフォルト予測—機械学習手法の有効性比較評価—」, JAFEE2017 冬季大会予稿集, pp.192-203.
- 山下智志・三浦翔（2011）『信用リスクモデルの予測精度—AR 値と評価指標—』, 朝倉書店.
- 山中卓・木下美咲（2018）「受注データに基づく構造型信用リスク評価モデル」, 日本銀行ワーキングペーパーシリーズ No.18-J-2.
https://www.boj.or.jp/research/wps_rev/wps_2018/wp18j02.htm/
- Chen T. and Guestrin C. (2016), “XGBoost: A scalable tree boosting system.”, *In Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and*

Data Mining, pp.785-794.

Engelmann B. and Rauhmeier R. (2006) *The Basel II Risk Parameters*, Springer, Heidelberg.

Hastie T., Tibshirani R. and Friedman J. (2014) 『統計的学習の基礎』, 共立出版.

Hasumi R. and Hirata H. (2010) “Small Business Credit Scoring: Evidence from Japan”, RIETI Discussion Paper Series 10-E-029.

Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q. and Liu, T.Y. (2017) “LightGBM: A highly efficient gradient boosting decision tree”. *In: Advances in Neural Information Processing Systems*, pp.3149–3157.

Yao J., Chapira M. L. and Margaryan M. (2017) “Checking account activity and credit default risk of enterprises: An application of statistical learning methods”, arXiv:1707.00757v1.