説明可能な AI（XAI）による機械学習モデルの特性分析

丹　波　靖　博*
原　口　健太郎**
新　原　俊　樹***

1．はじめに

近年，コンピュータの計算速度の向上，データ活用の重要性の認識，データ取得技術の進歩などにより，コンピュータで膨大な計算を行うことで柔軟かつ高精度の説明力を確保できる AI（人工知能）モデルの利用が広がっている。AI にはいくつかの計算手法が存在するが，機械学習モデルという呼称は，そのうちの複数のモデル群のことを指す。機械学習モデルには多くの種類が存在し，適用にあたっては目的や前提条件について検討を行い最適なモデルを選定する作業が重要となる。機械学習モデルの多くはモデル式を明示的に表すことができないものも多く，モデルの構造が非常に複雑である場合には，目的変数と特徴量の関係を理解することが困難なことが多い。分析を行う目的によってはモデルの説明可能性よりも予測精度を求める場合があるが，モデルの説明可能性がより重要となるケースも多く存在し，その際にはモデルの解釈を行いモデルの振る舞いに関する説明が要求される。しかし，モデルが複雑な機械学習モデルは，その仕組みがブラックボックス化されることも多い。このような背景から，予測精度と説明可能性を両立するため，機械学習モデルのデメリットである説明可能性を高める手法の研究が注目を浴びている。本稿では，得られ

*経済学部教授　**商学部准教授　***情報処理センター助教

た機械学習モデルの説明を行うための複数の「説明可能な AI（eXplainable AI : XAI）」の手法を紹介する。本稿では以下の内容で議論を進める。２章では，機械学習モデルを用いた分析の利用背景と本研究の目的について述べる。３章では，先行研究や主要な関連資料について紹介する。４章では，機械学習モデルの基本的な知識，いくつかのモデルの特徴，機械学習モデルの説明可能性について議論する。５章では，説明可能な AI の主要な手法の紹介と利用目的，および注意点などについて述べる。６章では，説明可能な AI の具体的な適用例と将来展望について考えをまとめる。７章では，最後のまとめを述べる。

２．本研究の目的

我々がデータを利用して分析を行う場合，単純な統計量などの記述を行うと同時に何らかの分析モデルを利用することが多い。導出されたモデルから目的変数とその他変数との関係性を理解したり，目的変数に何らかの影響を与えようとしたりする際には，それら変数間の関係性や因果関係を知ることが重要となる。説明可能性の必要の度合いは，分析の目的によっても大きく変動する。例えば，画像データによる男女の識別モデルでは，モデルの説明可能性は重要でなく，男女をうまく識別できたかという正答率などが重視される。文字の自動認識の問題も同様に正答率が重視される例である。他方，金融分野の分析においては，得られた結果の説明可能性が重要となることが多い。例として，銀行における住宅ローン審査の諾否判定において計量化モデルを用いる場合，個人の過去数年の収入，債務状況，属性情報などから返済能力が吟味され，将来にわたって支払い能力を維持できるかどうかが審査される。現在のところ，機械学習モデルの構築においてはモデル利用による判断の精緻性や業務の効率性に重きが置かれる傾向があるが，将来的にはモデルの精度や効率化だけでなく公平性，説明責任，透明性，信頼性などが求められる可能性が高いと考えられる。つまり，AI モデルの判断によって顧客に経済的な損失を被らせることを回避するとともに顧客サービス上の説明責任にも対応するため，なぜそのような結果になったのかというモデルの説明可能性を重要視しなくてはならない状

況が想定されるのである。また別の例として，金融機関の投資部門におけるポートフォリオの定期的リスクモニタリングのケースを考えると，保有対象資産にアラームが出た場合，能動的な投資戦略を立てるためにはアラーム発出の原因を知り，なぜそのような結果が導かれたのかというロジックをその後の対応や将来の行動に反映する必要がある。加えて，過去のリーマンショックなどの金融危機の経験から，このような具体的なリスクの把握は金融庁も重視している。モデルが導いた結果の理由を知ることは，その問題やデータに関する理解を深め，もしモデルが判断を誤った場合においても，その誤りの原因を探ることに役立つ。このように説明可能性が求められつつある社会的背景から，予測精度が劣っても説明可能性の高い回帰モデルが実務においてもいまだに利用されることが多いという現状がある。

しかし，近年機械学習モデルを用いた新たな試みは急速に広がっており，説明可能性のデメリットを克服し，より精緻な機械学習モデルを活用する重要性は増していくと考えられる。このような背景から，機械学習の課題である説明可能性の課題を解決するため，本研究では，機械学習モデルの特徴について概観し，説明可能性についての観点についてまとめる。また，近年急速に研究が進みつつある説明可能な AI の手法について，どのような教師あり機会学習モデルにも適用できるモデル非依存の主要な手法を紹介する。また，説明可能な AI 技術の具体的適用について展望する。

3．先行研究と関連資料

機械学習におけるモデルの解釈可能性ついての研究は近年急速に進められており，数多くの研究がなされているが，そのうちのいくつかの研究や関連資料について紹介する。Molnar（2020）は，機械学習において，モデルの解釈可能性に焦点を当て，決定木や線形回帰などの単純なモデル，および特徴量の重要度（feature importance）や ALE（accumulated local effects），LIME（Local interpretable model-agnostic explanations），SHAP（SHapley Additive exPlanations）などのモデルに非依存な XAI の各手法を紹介している。テーブルデータに焦点

を当て，それらの手法がどのように機能しているかや手法の長所，短所などについて解説を行っている。Biecek and Burzykowski（2022）では，分類や回帰の問題に適用するためのモデル非依存な手法が紹介されている。予測モデリングにおける本質的な課題は，データや計算能力，アルゴリズム，柔軟なモデルの欠如ではなく，モデルの探索（モデルが学習した関係の抽出），モデルの説明（モデルの決定に影響を与える主要な要因の理解），モデルの検査（モデルの弱点の識別とモデルの性能評価）のためのツールの不足であることを説明し，予測モデルを改善し変化する環境でモデルを監視するための手法について議論している。Mille（2017）では，研究者が良い説明と考えるものだけを使用する傾向にあることを問題視し，説明可能な AI の分野は哲学，認知心理学，社会心理学などの既存の研究に基づいて構築すべきであると主張している。これらの分野における関連論文をレビューし，重要な発見を抽出し説明可能な AI の研究に取り入れる方法について議論している。

大坪（2021）は，AI の説明責任についての重要性や課題の解説を行い，実務において誰にどのような説明が求められ，現在の XAI の技術を用いて何が可能かを議論している。LIME や SHAP などの XAI 技術を解説し，説明相手・性能要件・データ特性・AI アルゴリズムなどに応じた手法の選択について説明を行っている。森下（2021）は，高い予測精度を維持しつつ，機械学習モデルの解釈可能性を向上させるための様々な手法を取り上げている。機械学習の解釈可能性についての解説，線形回帰モデルの解釈可能性の理解について述べた後，複雑なブラックボックスモデルの挙動を理解するために，Permutation Feature Importance（以下 PFI と呼ぶ），Partial Dependence（以下 PD と呼ぶ），Individual Conditional Expectation（以下 ICE と呼ぶ），SHapley Additive exPlanations（以下 SHAP と呼ぶ）などの手法を紹介し，特徴の重要性，特徴と予測値の関係，インスタンス固有の挙動，予測値の理由について解説を行っている。吉田・田嶋・今井（2020）では，テーブルデータを用いて回帰問題として決定木ベースモデル（XGBOOST）で学習したモデルに対して SHAP 値を算出し，SHAP 値が概ね的確に特徴量のモデル学習への貢献を評価可能であることを示している。

４．機械学習モデルの特徴と説明可能性

4-1. 機械学習モデルの特徴

近年急速な発展を遂げている機械学習モデルは，コンピュータの計算機能を利用してデータから反復的に学習し，データに潜むパターンや規則性を見つけ出す手法全般のことを指す。そのような，モデルロジックの背景から，サンプル数が増加するにつれてデータに潜む規則性を見つけやすくなりデータに対して適応的に説明力を改善することが可能となるため，伝統的な統計分析手法に比べて，多くのデータを必要とすることが一般的である。

機械学習は，教師あり学習，教師なし学習，強化学習の３つの主要な種類に分類される。特徴を表すデータを特徴量，答えであるデータを目的変数またはラベル，ひとつひとつのデータのことをインスタンスと呼ぶ。教師あり学習は，目的変数（教師データ）と特徴を表すデータ（特徴量）をモデルに与え，目的変数の推定を行い予測値と教師データとの誤差を小さくすることで，機械学習モデルの推定精度を向上するために学習を行う。一方，教師なし学習は目的変数の正解を与えないモデルで，教師あり学習と同様，データの特徴を表す特徴量を使用するが，目的変数を推定するのではなく，データの中で特徴の近い部分集合を見つけることで入力データの構造を理解することなどを目的とする。強化学習は，設定環境下で行動するエージェントの報酬を定義し，報酬を最大化するようにエージェントが行動し学習していく手法である。本稿では，教師あり学習に焦点を当て議論を進めていく。

教師あり学習の機械学習には，回帰や分類タイプのモデルが存在する。広義の意味では回帰モデルも機械学習のモデルに含まれる。これまで回帰分析は伝統的な分析手法としてアカデミックのみならず実務においても広く一般的に利用されてきた。回帰モデルには線形回帰，ロジスティック回帰など様々なモデルが存在する。線形回帰は線形モデルによる回帰により連続値を予測する手法である。最小二乗法や最尤推定法によって係数と切片を決定し，それにより目的変数の値を予測する。ロジスティック回帰モデルは回帰モデルによりデータを分類する手法である。例えば線形回帰の出力をロジット関数に入力すること

で２値分類問題に対応することが可能である。回帰タイプと分類タイプの両方の問題に適用可能なモデルも存在する。サポートベクターマシンは線形回帰によってデータの分類を行うための機械学習アルゴリズムとして考案された手法である。その後，サポートベクターマシンは非線形の分類モデルへ拡張され，現在でもよく利用される手法である。分類を行うサポートベクターマシンは SVC（Support Vector Classification）と呼ばれている。また，サポートベクターマシンは回帰分析にも適用可能で，回帰を行うサポートベクターマシンは SVR（Support Vector Regression）と呼ばれている。サポートベクターマシンは計算過程でマージン最大化を取り入れることにより比較的少ないデータでも汎化性能が高い２値分類回帰モデルを構築することが可能である。分類モデルには，その他決定木，回帰ツリー，ランダムフォレスト，勾配ブースティングなど様々なタイプのモデルが存在する。ここではランダムフォレストモデルについて簡単な解説を行う。ランダムフォレストは決定木分析の手法の１つで，目的変数に影響する特徴量により樹木状のモデルを作成する分析方法で，予測，判別，分類などを目的として使われる機械学習の手法のひとつである。ある１つのインスタンスに対し，決定木における所属グループが決まるが，条件分岐によって分割されたグループの複数木における平均値を取ることで推測値を算出することもできる。ランダムフォレストは決定木モデルを複数利用することで，決定木単体よりも予測精度の向上を図る事ができるという利点を生かしたモデルである。その他にもニューラルネットワーク，ベイズ，時系列モデルなどのモデルが存在するが，ここでは各モデルの詳細についての説明は割愛する。詳しくは，Aurélien（2020），秋庭・杉山・寺田・加藤（2019），加藤（2018）などの資料を参照されたい。

4-2. モデルの説明可能性とは

社会的な機械学習モデルの活用の広がりにともない，モデル精度に加えて公平性，説明責任，透明性，信頼性などが社会的に求められるようになってきている。公平性とは，使用するユーザーの属性によらず公平な結果を提供できるよう不公平を生じさせるバイアスを排除することをいう。説明責任とは，モデ

ルの結果の誤りの原因や責任の所在を明確にすることである。つまり，そのモデルが入力データから，いかにして結果を導いたのかを示す根拠を提示することや，どこにその要因があったのかを明確にするための仕組みのことをいう。透明性とは，モデルの利用者が理解できるようシステムの情報を提示できることをいう。これには元となるデータ，検証内容，モデルの処理プロセスにおける基準や根拠などが含まれる。透明性を高めるためにはこのような構成要素についての情報を公に公開する必要がある。信頼性とは，データの小さな変化が目的変数の予測に大きな変化をもたらさないことをいう。これらの要件を満たすためにはモデルが「どのような処理を行いどのような根拠に基づいて結果を導いたのかを説明する」必要がある。本稿では，「データのどの部分を重視し，どのような要素判断基準により，どのような根拠に基づいて結果を導いたのかを説明できること」を説明可能性†と呼ぶことにするが，以下では前述の公平性，説明責任，透明性，信頼性のうち，主に説明責任を念頭に置いて議論を進める。

4-3. 主要モデルの説明可能性

モデルの説明可能性を分解すると，①目的変数との関係性における特徴量の重要度の定量化，②特徴量と予測値の平均的な関係性の把握，③平均的な関係だけでは捉えきれないインスタンスごとの特徴量と予測値との関係性の把握，④モデルが目的変数に対して行った予測に対する理由の解明，などの観点が挙げられる。

---

† 「説明可能性」と近い概念で「解釈可能性」という用語が区別して使われることもある。説明可能性とは，予測の判断理由を人間が理解できるように説明する技術のことを言う。つまり，得られた結果の理由を説明できることを包括しているため，必ずしもモデルの内部構造を精緻に解析できる必要はなく，原因を問う疑問への回答として説明できるという観点に焦点をおいている。一方，解釈可能性とは，モデルの内部構造を解析することで予測に至る計算過程を理解できることをいう。例えば高い解釈可能性のモデルとはパラメータの変更や入力データの変化により予測値がどのような影響を受けるかを予見できるモデルのことをいう。例えば，多くの解釈可能なモデルでは，ひとつのインスタンスに対する予測とデータの平均に対する予測の違いを説明することができる。本稿では，厳密な違いには踏み込まず両者の観点から議論を行う。また，特定の目的変数の予測や要因による説明をする必要がある場合は，因果関係の有無を検証する必要がある点に注意する必要がある。

ここではまず，シンプルで一般的によく利用されている線形回帰モデルを例にとって議論する。広い意味で機械学習モデルの一つである線形回帰モデルは予測精度が相対的に低い一方，構造がシンプルなためモデルの説明可能性が高い特徴がある。一般的には，予測精度と説明可能性はトレードオフの関係が存在する。線形回帰モデルには，長年の研究成果や多くの適用ケースから，モデルの振る舞いについて多くの知見が蓄積されているという利点や大規模なデータセットに対しても計算負荷が非常に小さいという利点もある。線形回帰モデルは機械学習モデルと比較すると予測精度が相対的に低いという特徴があるが，予測精度を高めようとすると交互作用や非線形な関係を組み込むなど複雑なモデルとなるため，説明可能性を犠牲にしなくてはいけないという問題が生じる。

線形回帰モデルの高い説明可能性には，上記の４つの観点が満たされているという特徴がある。①～④の番号は上述のそれぞれの観点に対応している。①一つ目の観点として，線形回帰モデルでは回帰係数による特徴量の相対的な重要度を測ることができることが挙げられる。線形回帰モデルにおける回帰係数は各特徴量に対する感応度を表すため，係数が大きいほど予測値に与える影響が大きい。ただし，予測値に対する影響の大きさは特徴量の絶対値の大きさにも依存するため，より精緻な影響度を測定するためには特徴量の標準化が必要となることに注意する必要がある。特徴量の重要度が得られれば，各特徴量と予測値への影響をコストパフォーマンスや実現可能性の観点から比較し，重要な特徴量から優先して操作し予測値へ影響を与えることも可能となる。②二つ目は特徴量と予測値の平均的な関係が解釈できることが挙げられる。重回帰モデルにおける係数はその特徴量と予測値の感応度を表しており，その特徴量が１単位増加した場合の予測値の平均的な変化を表している。また回帰係数の符号はその特徴量と予測値に対する正負の関係を表している。③三つ目の観点として特徴量と予測値の各インスタンスにおける関係が解釈できることが挙げられる。例えば線形回帰モデルにおいて特徴量のべき乗項や交互作用項がある場合は，各インスタンスにおいて特徴量が予測値に与える影響が非線形となるモデルが構築される場合がある。その場合には，インスタンスごとに特徴量の増加が予測値に与える影響が異なることとなるが，線形回帰モデルにイン

スタンスのデータを与えることで特徴量と予測値の各インスタンスの関係を解釈できる。④四つ目の観点として，インスタンスごとの予測値に対する理由が解釈できることが挙げられる。一つのインスタンスを見た場合，得られた線形モデルの回帰係数とそのインスタンスの特徴量の値から，各特徴量の予測値に対する貢献度を正確に計算することができる。そのため線形回帰モデルでは各インスタンスに対して，その予測値になっている理由の解釈が可能となる。

次に，線形回帰モデル以外の機械学習モデルである決定木を例に取ると，決定木は分岐条件ルールを積み上げツリーを作ることで入力データを分類していくモデルであるため，最終的な予測結果が得られるプロセスにおいて各インスタンスがどの条件をたどってその予測結果に至ったのかを確認することができる。このため決定木モデルは機械学習モデルの中でも比較的説明可能性が高いモデルであるといえる。ただし条件ルールが幾層にも重なった複雑な決定木を構築した場合は条件が複雑になるため，説明可能性が低くなってしまうという特徴がある。またランダムフォレストモデルのように複数の決定を利用し予測精度を向上させるモデルは単体木よりも複雑な構造となるため，説明可能性は単体木よりも低くなる。

ブースティングモデルは多くのケースにおいて比較的高い性能を示すことが知られているが，学習しなくてはならないパラメータを多数持ち，モデル構造もより複雑なためモデルの根拠を人間が理解するのは困難なことが多く，ランダムフォレストモデルよりも説明可能性がさらに低いモデルとなる。また，ニューラルネットワークの一種であるディープラーニングは複数の層構造の機械学習アルゴリズムであり，内部パラメータの多さからモデルの説明可能性が難しい技術であることが知られている。このように，一般的に機械学習モデルは高い予測精度を達成できるというメリットがある一方，その構造が複雑になればなるほどモデルの構造が複雑となるため，モデルの説明可能性が低いというデメリットがある。

５．XAI の各手法の概要

これまで述べてきたように，機械学習モデルは線形回帰などの手法に比べ非線形な関係性も表現できるため，説明力の高いモデル構築が可能である場合が多いが，モデルの解釈が困難である点が分析の目的によっては大きな課題となることもある。特に上記で述べたように金融の分野では，モデルの説明責任が問われるケースも多く，機械学習の分析においてモデルの解釈を行うことは分析の内容をブラックボックス化させないという観点から重要な場合が多い。前述したモデルの説明可能性の４つの観点において，機械学習モデルの説明を行うためのそれぞれに対応する説明可能な手法に，①Permutation Feature Importance（PFI）②Partial Dependence（PD）③Individual Conditional Expectation（ICE）④SHapley Additive exPlanations（SHAP）が挙げられる。これらのモデルは，様々な機会学習モデルにも適用できるモデル非依存の手法であることも大きなメリットである。

5-1. 特徴量の重要度

上記４つの説明可能性の観点のうち，Permutation Feature Importance（PFI）の手法を用いて，特徴量の重要度の定量化を行うことができる。PFI は，各特徴量の目的変数に対する重要度を予測誤差の増加分で定量的に判断するものである。具体的には，ある特定の特徴量の値をシャッフルし，その特徴量が使えない状態にして，目的変数に対する説明力の変化を計量化する手法である。各特徴量に対して，この作業を繰り返し，全ての特徴量の情報が使える場合と各特徴量の情報を使わない場合との予測誤差を計算し比較を行う。ある特徴量をシャッフルして使わない場合に予測誤差が大きく増加する場合，モデルがその特徴量を重視していると考えることができる。一方，もし予測誤差が変化しないなら，その特徴量は重要でないと判断できる。PFI の利点に，データをシャッフルして適用すれば予測誤差は計算可能なため適用するモデルに対する制限がないこと，アプローチが直感的で理解しやすいことなどが挙げられる。一方，PFI の注意点として，特徴量の相関が高い場合，特徴量同士で重要度を

相互に打ち消しあってしまうことがあるため適切に予測誤差を計測することができない場合があること，因果関係としての解釈を行うことには危険を伴うことなどが挙げられる。

5-2. 特徴量と予測値の平均的な関係

前述の PFI では，特徴量の重要度を計量するが，特徴量の値と目的変数の正負の方向性の関係を調べる手法に Partial Dependence（PD）がある。線形モデルのように係数の符号で正負の関係を容易に知ることのできる手法と違い，ランダムフォレストのようなツリーモデルでは，特徴量と目的変数の関係が非線形で複雑なブラックボックスとなりがちで，特徴量と予測値の関係を解釈することは容易ではない。PD は他の特徴量を固定してある特徴量の水準だけを変化させ，各インスタンスの予測値を平均して可視化する方法である。ある特徴量の値と目的変数の値に正の関係があることがわかれば，その特徴量を増加させることにより目的変数に正の影響を与えることができる。PD の利点には，機械学習モデルの種類によらず適用できること，各特徴量がモデルの予測値にどのような影響を与えているかを確認できること，他の特徴量の影響も考慮に入れて特徴量と目的変数の関係を捉えることができることなどが挙げられる。一方，PD の注意点として，因果関係として解釈することはできないこと，特徴量とモデルの予測値がインスタンスごとに異なっている場合でもその影響を無視してしまうことなどが挙げられる。その他の注意点として特徴量に相関がある場合には理論的な関係を復元できない場合があることが挙げられるが，その場合には，条件付き期待値をとる Marginal Plot や特徴量の値の範囲を分割されたブロックにし，各ブロックに含まれるインスタンスのみを対象にして，その区間の両端での予測値の差分により特徴量と目的変数の関係を分析する Accumulated Local Effects（ALE）などがある。ALE については，Apley and Zhu（2020）などを参照されたい。

5-3. インスタンスごとの特徴量と予測値との関係

Individual Conditional Expectation（ICE）は，PD のように特徴量と目的変数の平均的な関係を見るのでなく，インスタンスごとに他の特徴量を固定してある特徴量の水準を変化させ，特徴量と予測値の関係を確認する方法である。PD は特徴量と目的変数の平均的な関係に注目しているため，インスタンスごとに特徴量とモデルの予測値の関係が異なる場合にはその影響を考慮できないという課題がある。一方，ICE ではインスタンスごとに特徴量と目的変数との関係性を可視化できるので，インスタンスごとの異質性を把握することでモデルの振る舞いをより深く解釈することが可能となる。ICE の利点として，各インスタンスに対して各特徴量のモデル予測値への影響を確認できること，機械学習モデルの種類によらず適用できること，特徴量の交互作用をとらえることができる場合があることなどが挙げられる。ICE の注意点として，因果関係として解釈することはできないこと，PD のように平均を取っていないので値が安定しない傾向があること，実際の特徴量の値から大きく離れた範囲では推定が大きくはずれる可能性があるため実際の値の近傍で解釈する必要があることなどが挙げられる。

5-4. 目的変数に対して行った予測の理由

ICE では各インスタンスに対して各特徴量のモデル予測値への影響を確認できる一方，モデルが予測値を出した理由を知ることができない。各インスタンスの予測値に対して解釈を与えることのできる手法に SHapley Additive exPlanations（SHAP）がある。SHAP では，ある特徴量の値の増減が目的変数に与える影響や特徴量の貢献度の分布を色分けによる出力図で可視化することができる。また，SHAP の絶対値の平均をとることで特徴量の平均的な重要度を数値で評価することも可能である。SHAP は協力ゲーム理論における複数プレイヤーの協力によって得られた利得を各プレイヤーに公正に分配するための評価手法であるシャープレイ値（Shapley Value）を機械学習に応用したものである。Lundberg et al.（2020）が Nature Machine Intelligence に掲載されて以降，SHAP は自然科学のみならず社会科学の分野においても説明可能な AI における有効

なツールの一つとして利用され始めている。SHAP の利点は，協力ゲーム理論の Shapley 値の考え方を応用しており，貢献度の分解における望ましい性質を持っていること，SHAP は各インスタンスに対して使えるミクロ的な解釈とマクロ的な解釈のどちらにも用いることができることなどが挙げられる。SHAP の注意点として，特定インスタンスに対してモデルの予測値算出理由が分かるものの，ICE のように特徴量が変化した際の予測値の変化については分からないこと，計算負荷が高いこと，理論面の難易度が他の手法に比べると高いことなどが挙げられる。

６．説明可能な AI の具体的活用例

説明可能な AI 技術の具体的活用例として，機械学習モデルの内容についての詳細を知ることにより，イベントの前後における目的変数と特徴量との関係性の変化についての考察を行うことが可能となることが挙げられる。イベント発生前後のそれぞれの期間において，機械学習により機械学習モデルを構築し，説明力の高いモデルの特定や，選択された特徴量の変化，特徴量と目的変数との関係性の変化を理解することにより，イベント発生前後でどのような変化がもたらされたかなどの分析が可能となると考える。

例えば，新型コロナウイルス発生前後や東日本大震災発生前後のそれぞれの期間において，丹波・原口（2022a，2022b）で実践した手法により機械学習モデルにより地方債のスプレッド推定モデルを構築し説明力の高いモデルと選択された特徴量を確認する。得られた機械学習モデルについて説明可能な AI の手法を活用し，地方公共団体間のスプレッド差の要因について新型コロナウイルス発生前後の変化について考察を行うことが可能となる。それにより投資家がイベント発生に伴い投資行動をどのような理由からどのように変化させたかなどの投資行動の変化を読み取ることができると考える。このような大きなイベント発生時においては，投資家はリスク回避的な行動を取ると考えられるが，リスク回避的な行動とは信用リスクの回避なのか，それとも中長期から短期への時間的リスクの回避なのかなどのより詳細な検証が行える可能性がある。信

用リスクの回避行動を行ったとしても，投資家がリスクと考えるリスク許容度の程度を知ることが可能なのか，つまり投資家はどの特徴量を参照しどの投資対象であればよしとするのかなど，具体的な対象商品までの深掘りも可能となる可能性がある。さらに，新型コロナウイルスと東日本大震災では，投資行動の変化の大きな違いはあるのか，あるとすればその理由は何かなどの考察の手がかりになると考える。このような大きなイベントに対して，売り買いの多かった具体的な銘柄の特定まで行うことができれば，今後の投資戦略の立案にも応用することができ，実務上も有意義な情報となり得ると考える。

7．ま と め

本稿では，各種機械学習モデルの特徴について概観し，機械学習モデルの説明可能性の必要性やそれに関するいくつかの観点についてのまとめを行った。それら観点について，線形モデルと明示的なモデル式のない機械学習モデルにおける説明可能性について議論を行った。また，近年急速に研究が進みつつある「説明可能な AI（XAI）」の手法について，どの機会学習モデルにも適用できるモデル非依存の主要な手法を紹介した。さらに，経済的イベント発生前後における投資行動の変化について，説明可能な AI を活用するアイデアについての言及を行った。

参考文献

秋庭伸也，杉山阿聖，寺田学，加藤公一（2019）『見て試してわかる機械学習アルゴリズムの仕組み　機械学習図鑑』翔泳社。

大坪直樹他（2021）『XAI（説明可能な AI）― そのとき人工知能はどう考えたのか？』リックテレコム。

加藤公一（2018）『機械学習のエッセンス ― 実装しながら学ぶ Python，数学，アルゴリズム』SB クリエイティブ。

丹波靖博，原口健太郎（2022a）「わが国における地方債スプレッド推定モデル構築に対する機械学習の適用可能性」『経済学論集（西南学院大学）』第56巻第1・2合併号，pp75−91。

丹波靖博，原口健太郎（2022b）「機械学習を用いたわが国における地方債のスプレッド推定モデルの構築」『JAFEE ジャーナル』近刊。（査読採択済み）。

森下光之介（2021）『機械学習を解釈する技術～予測力と説明力を両立する実践テクニック』技術評論社。
吉田秀穂，田嶋優樹，今井優作（2020）「決定木モデルの解釈における SHAP 値の有用性の検証」『人工知能学会第34回全国大会論文集』1～3頁。
Apley, D. W., and J. Zhu. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society*: Series B (Statistical Methodology) 82.4, pp.1059−1086.
Aurélien, G.（2020）『scikit-learn，Keras，TensorFlow による実践機械学習』下田倫大監修，長尾高弘訳，オライリージャパン。
Biecek, P. and T. Burzykowski (2022) Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models, *Chapman and Hall/CRC*.
Lundberg, S. M. G. Erion, H. Chen, A. DeGrave, J. M Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. Lee (2020) Explainable ai for trees: From local explanations to global understanding, *Nature Machine Learning*, 2, pp.56−67.
Miller, T. (2017) Explanation in artificial intelligence: Insights from the social sciences. *arXiv Preprint* arXiv:1706.07269.
Molnar, C. (2020) Interpretable Machine Learning, *Lulu.com*.

謝　辞

本研究は科学研究費補助金（JSPS KAKENHI Grant Number JP19K23214，JP20K02058，JP21K13412），日本経済研究センター研究奨励金および一般財団法人ゆうちょ財団の研究助成の交付を受けて行ったものである。また，ロンドン証券取引所・FTSE グループ The Yield Book Inc. には，債券分析ソフトイールドブックを通じて効率的な分析ツールと貴重なデータを提供いただいた。この場を借りて御礼申し上げたい。