法務省保護局 御中

保護観察におけるアセスメントへのAI導入に向けた基盤整備業務（最終結果報告書）

令和6年2月29日

コンサルティング事業本部 ココロミルラボ

三菱UFJリサーチ&コンサルティング

世界が進むチカラになる。

MUFG

目次

I.	本調査研究の概要	2
II.	再犯リスク予測の初期モデルの構築・実装	10
1.	アセスメント項目の設計・運用の検討	14
2.	再犯リスク予測モデルの要件整理	26
3.	初期モデルの構築	37
III.	試験用画面サンプルの作成とユーザーからの意見聴取	64
1.	発見～定義フェーズ	67
2.	定義～開発フェーズ	87
3.	提供フェーズ	114
4.	まとめ	122
IV.	来年度以降に向けた活動の整理	126

I.　本調査研究の概要

背景・目的

「保護観察におけるアセスメントへのAI導入に向けた基盤整備業務の調達に関する仕様書（以下「仕様書」という）」第3の1　記載内容より抜粋

■ 背景

- 保護観察においては、犯罪をした人や非行のある少年の再犯・再非行を防ぎ、その改善更生を図るため、指導監督及び補導援護を実施
- 近年のAI技術の進展を踏まえ、法務省保護局（以下「貴省」という。）では、より的確なアセスメントに基づく保護観察処遇を行うことを目的として、令和3年1月から本格実施を開始した「CFP」等のアセスメントに今後AI （人工知能）を導入・実装することを検討

■ アセスメントへのAI導入の目的

1. 現状CFPにより判定している長期再犯リスクについて、保護観察官ごとの判断のばらつきを減少させ、情報の見落とし等を防ぎ、より適切な分析を支援するためのAI技術開発と実装・運用を行い、保護観察官によるアセスメントの質の向上と書類精査等に係る負担軽減を図る。
2. 急性再犯リスクについて、短期的な再犯リスクを判断するモデルを確立し、急性再犯リスクの高まりの検知・発報が期待できるAI技術開発と実装・運用を行い、保護観察官による適時適切な介入の判断を支援する

■ 調査研究事業

令和4年度
「保護観察におけるアセスメントへのAI導入に関する調査研究」
（以下「令和4年度調査研究」という）の実施ポイント

1. 長期の再犯リスクに影響する要因についてデータ解析を行い、再犯リスク予測モデリングに必要な項目や特徴量について基礎的な知見を洗い出すこと
2. 長期再犯リスク及び急性再犯リスクの予測等に必要な開発・実装を行い、令和9年度頃の運用を目指すこととした場合の具体的なロードマップについての検討

令和5年度
本調達に係る調査研究　（以下「本調査研究」という）
における実施ポイント

(A) 既存のデータから適切なアセスメント項目の設計

(B) 再犯リスク予測の初期モデルを構築

(C) 画面上で動作するサンプルを試験的に作成

(D) 運用上の課題を洗い出し、システム実装に向けた検討

本報告書の構成

本報告書では、本調査研究の実施ポイント(A)～(D)はそれぞれ下記のように各章内に記載。

Ⅰ章：本調査研究の概要
- 背景と目的
- 本報告書の構成
- サマリー
- 活動計画

Ⅱ章：再犯リスク予測の初期モデルの構築・実装 ⇐ (A) (B)
- アセスメント項目の設計・運用の検討
- 再犯リスク予測モデルの要件整理
- 再犯リスク予測の初期モデル構築

Ⅲ章：試験用画面サンプルの作成とユーザーからの意見聴取 ⇐ (C)
- 発見～定義フェーズ
- 定義～開発フェーズ
- 提供フェーズ
- まとめ

Ⅳ章：来年度以降に向けた活動の整理 ⇐ (D)
- 業務に実装・運用するための課題の整理
- AI-CFPの運用に固有の課題の整理
- AI-CFPの実現に向けたロードマップ案の策定

(A) 既存のデータから適切なアセスメント項目の設計
(B) 再犯リスク予測の初期モデルを構築

(C) 画面上で動作するサンプルを試験的に作成

(D) 運用上の課題を洗い出し、システム実装に向けた検討

サマリー

(A) 既存のデータから適切なアセスメント項目の設計 （Ⅱ章）

- 要因分析の要因記載欄（いわゆるCFPテキストデータ）の分析のみから、サブカテゴリ等のアセスメント項目の設計は難しかったが、法務省の知見からサブカテゴリの分類ラベルを作成して、CFPテキストデータの大部分を説明可能なアセスメント項目を作成した
- これらアセスメント項目を、リスク比などを活用して検証したところ、有用である可能性を示すことができた

(B) 再犯リスク予測の初期モデルを構築 （Ⅱ章）

- 本調査研究における要件を整理し、予測性能を重視し、CFPの有無など複数パターンでリスク予測モデルを構築した
- 予測モデルの性能については、パターン1（全データ使用）の号種1では、モデル評価指標（ROC）0.8、適合率0.9、再現率0.7など総合的に、令和4年度調査研究よりも高かった。CFPデータを予測性能に取り込むパターン2では、短期間のデータにも関わらず上記と同等の性能を示しており、CFPテキストデータも定量的には予測性能の向上に有効であること示唆された

(C) 画面上で動作するサンプルを試験的に作成 （Ⅲ章）

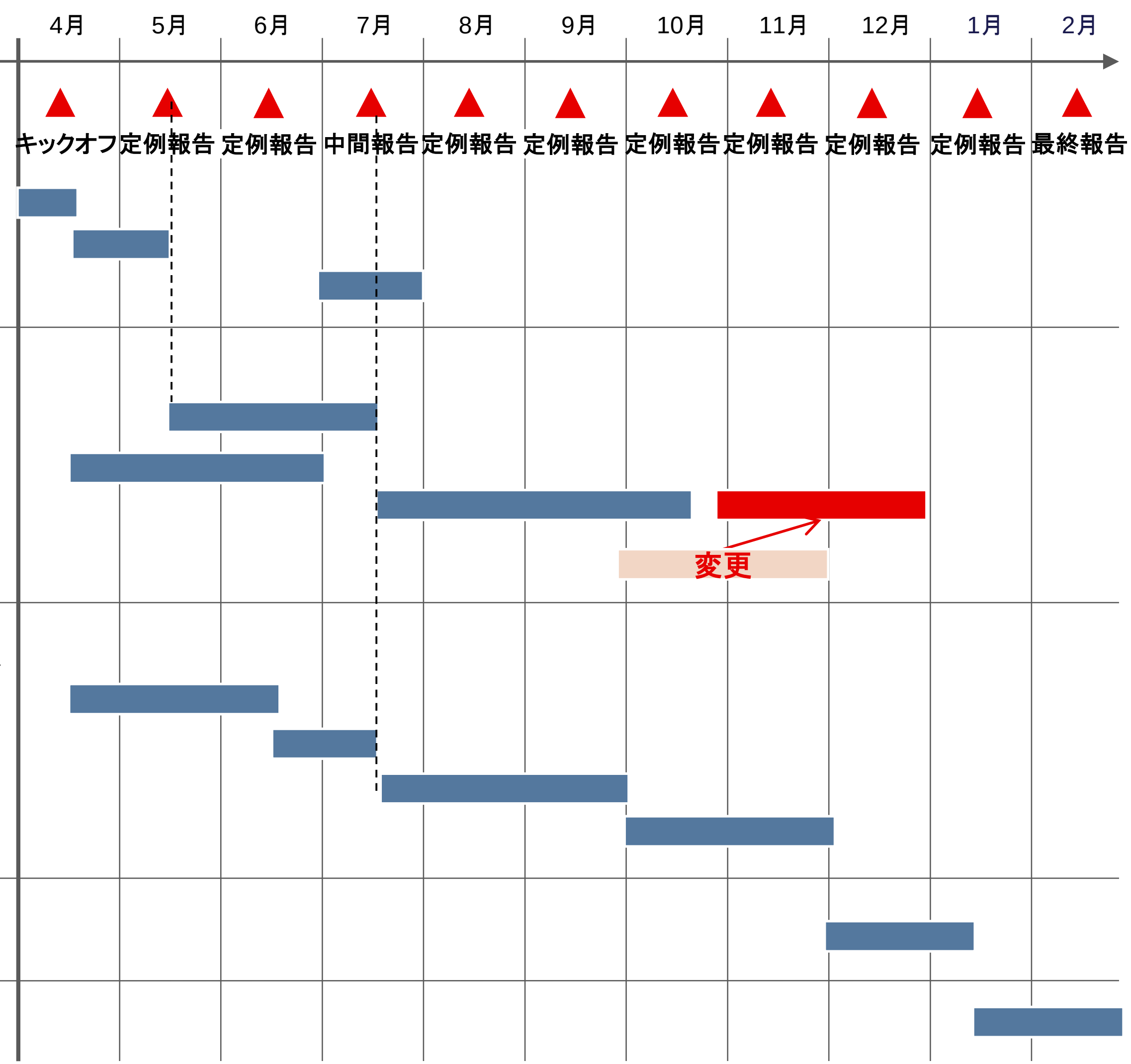
- 再犯リスク予測モデルをアセスメントへの活用について保護観察官へのリサーチを行った結果、保護観察官のアセスメントの質の向上には、パス図の作成の支援、個々の対象者の犯罪に至るパターン抽出に寄与することで、長期再犯リスク及び急性再犯リスクのどちらにも有効であることが示唆された
- リサーチ結果に基づいて、画面を試作し、予測結果の提示方法について、議論及びUIの評価を行った。①業務システムの改善による書類精査等に係る負担を軽減、②アセスメントに予測モデルの結果を活用することで、データ登録→分析→活用の循環を作ることにより、長期再犯リスクのサイクルをベースにした、急性再犯リスクの検知が可能になることが示唆された

(D) 運用上の課題を洗い出し、システム実装に向けた検討 （Ⅳ章）

- ITシステムを業務に実装・運用するための課題、AI-CFPの運用に固有の課題をそれぞれ整理した
- 令和4年度調査研究において検討されたAI-CFPシステムの構築だけでなく、AI-CFPに関連する既存のシステムやデータ管理など様々な課題が明らかになったため、それらを段階的に解決する方針としてのロードマップ案を作成した

活動計画：スケジュール概要

■ 基本的な作業項目（1～5）と活動計画について下記に示す。 作業項目2の赤字部分は当初計画より合意の上変更（本章で後述）



【作業項目】

1. 活動計画の詳細化・具体化

1. CFPモデルや業務フロー等の詳細理解
2. 実行可能な活動の抽出と計画修正
3. 中間結果報告と修正計画の策定

2.再犯リスク予測の初期モデルの構築・実装

1. アセスメント項目の設計・運用の検討
2. 再犯リスク予測モデルの要件整理
3. 初期モデルの構築
4. ~~初期モデルの活用方法の検討~~

3.試験用画面サンプルの作成とユーザーからの意見聴取

1. 発見
2. 定義
3. 開発
4. 提供

4.来年度以降に向けた活動の整理

5.報告書の作成

活動計画：作業項目一覧(1/2)

作業項目		概要	想定アウトプット	必要なインプット	依頼事項（例）
1. 活動計画の詳細化・具体化	**1. CFPモデルや業務フロー等の詳細理解**	■ 仕様書　第3の2(1)カ、同(3)アに記載される情報、作業項目2に必要なデータなど、契約後に受領する。	—	■ CFPモデルや業務フロー、期初統計情報・CFPデータなど	■ CFPモデルや業務フロー、データなどの提供
	2.実行可能な詳細計画の策定	■ 上記を受けて、作業項目2～3の実行性等を検討、詳細計画を策定する。 ■ 他に定例打合せ等の実施方針を確定する。	■ 詳細計画	—	■ 詳細活動計画のレビュー・承認
	3.中間結果報告、詳細計画の修正	■ 活動進捗及び結果見通し（必要なら修正計画）を中間報告としてA4用紙1枚にまとめる。 ■ 来期以降に必要となる事務等の概要についてまとめ、提案する。	■ 中間報告書 ■ 修正詳細計画	■ 作業項目2～3に関する7月までの実行状況	■ 中間報告等のレビュー・承認
2.再犯リスク予測の初期モデルの構築・実装	**1. アセスメント項目の設計・運用の検討**	■ 仕様書 第3の1「既存のデータから適切なアセスメント項目の設計を進める」を既存のCFPデータから支援する方法（統計的検定手法や機械学習手法）の適用を試行する。	■ 疑似的なアセスメント項目・評価結果	■ CFPデータ（要因テキスト）、期初統計情報（再犯有無）	■ CFP要因テキストへのカテゴリ該当ラベルの付与
	2. 再犯リスク予測モデルの要件整理	■ 拡張性や発展性として求められる業務要件を抽出・整理、その優先順位等も確認する。 ■ それらに合致したアルゴリズム（機械学習や確率モデリング）を選定する。	■ 初期モデルのアルゴリズム候補リスト	■ ヒアリングによる業務要件等	■ 貴省ご担当者とのディスカッション
	3. 初期モデルの構築	■ 作業項目2.2で選定されたアルゴリズムで、初期モデル構築を試行する。 ■ データや整理した要件によって実施内容の詳細は変更の可能性がある。	■ 初期モデル ■ 評価結果（予測精度など）	■ 期初統計情報・CFPデータなど ■ （要件によって現場知見ヒアリングなど）	■ （要件によって貴省ご担当者とのディスカッションなど）
	~~4. 初期モデルの活用方法の検討~~	■ ~~業務における有用性の評価、使い方の検討など、活用に向けた課題を検討・抽出する。~~	■ ~~今後に向けた課題など~~	■ ~~初期モデル~~ ■ ~~評価結果~~	■ ~~貴省ご担当者とのディスカッション~~

活動計画：作業項目一覧(2/2)

作業項目		概要	想定アウトプット	必要なインプット	依頼事項（例）
3. 試験用画面サンプルの作成とユーザーからの意見聴取	1. 発見	■ 第3の2(3)ア記載の業務フローヒアリング後、アセスメント～保護観察官による分析、書類作成の業務内容の詳細を把握し、適時適切な介入のためのシステムのあるべき姿を検討する。	■ ステークホルダーマップ ■ 価値マップ ■ カスタマージャーニーマップ（AsIs） ■ ペルソナ	■ 貴省内における業務フロー及び、当該業務に関するユーザーの行動や状況等	■ インタビュー等における貴省内の調整 ■ ワークショップ形式での開催の場合、日程や場所・参加者の調整
	2. 定義	■ ユーザの価値の視点から、優先して解決すべき課題を選定する。 ■ 戦略を策定し、シナリオを作成する。	■ カスタマージャーニーマップ(ToBe) ■ ストーリーボード	■ 作業項目3.1のアウトプット	■ ワークショップ形式での開催の場合、日程や場所・参加者の調整
	3. 開発	■ プロトタイプを開発する。	■ ワイヤーフレーム ■ プロトタイプ	■ 作業項目3.2のアウトプット	—
	4. 提供	■ ユーザー要求が解決できているか評価を実施する。（ユーザビリティ調査、エキスパートレビューなど）	■ 評価結果	■ 作業項目3.1-3.3のアウトプット	■ 評価場所の確保 ■ 評価者への依頼
4 来年度以降に向けた活動の整理		■ 本番開発や利用に向けた課題を抽出する。 ■ 必要に応じてロードマップや中間報告内容（事務等の概要など）を改版しつつ、次年度以降の活動計画を策定する。	■ 課題等一覧 ■ 改版ロードマップ等	■ 本調査研究における全てのアウトプット	■ 貴省ご担当者とのディスカッション
5. 報告書の作成		■ 本調査研究の結果を報告書の形でまとめる。また、その内容を要約した概要紙をA4用紙1枚にまとめる。 ■ これまでの打合せ資料・議事録などもとりまとめ、納品物とする。	■ 納品物	■ 本調査研究における全てのアウトプット	■ 最終報告・納品物のレビュー・承認

活動計画： コミュニケーション方針・結果と作業項目の変更について

定例打合せなどのコミュニケーション方針・結果

- 定例打合せ
  - 月1回、Microsoft Teamsによるオンライン会議（60～90分）にて実施
  - その時点における進捗状況の共有、各作業項目について必要なディスカッションなどを議論
    - 予定通り実施。各回の詳細は別添「会議資料・議事録」を参照
  - その他必要に応じて、定例以外で打合せを実施。（実施打合せは以下の通り）
    - 5/12 データ確認に関する打合せ
    - 5/19 予測モデルの要件整理に関する打合せ
- 観察官などへのヒアリング実施 （実施完了）

作業項目2の変更について

- 法務省内でのデータベース上のデータに関する確認や差替え、それらに伴う当社でのデータ再加工の作業に想定以上の時間・工数を要したため、双方合意の上、本調査研究で提案していた作業計画を下記のように修正
  - 作業項目2の4で計画した「リスク予測モデルの検証」は作業項目4における総合的な評価に留め、当初の作業で予定した期間は「データ再加工～リスク予測モデル構築」の作業に充てることに変更
    （本章・スケジュール概要の頁の赤字で修正部分）

II.　再犯リスク予測の初期モデルの構築・実装

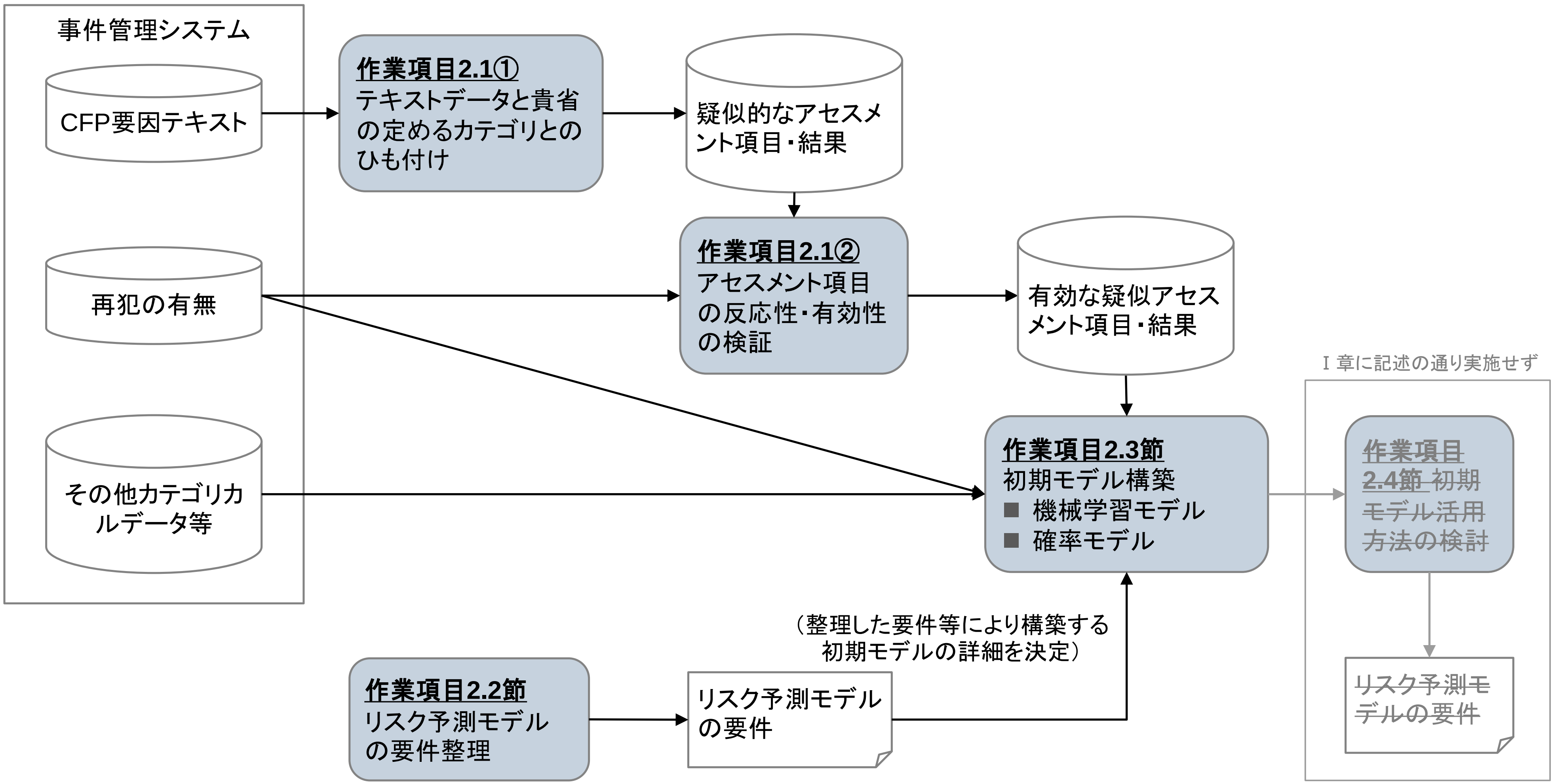
本調査研究における検証方針

- 事件管理システムに入力されているデータから再犯の有無に関する保護観察対象者の属性や特徴を分析し、再犯リスク予測の初期モデルを構築するプロセス（作業項目2）について、仕様書 第3の2(1)や令和4年度調査研究を踏まえた検証のポイントと対応方針（下図）
- 目的変数・説明変数を下記のように定義
  - 目的変数： 一定期間内での再犯の有無（2値）
  - 説明変数： 保護観察開始時に事件管理システムに入力する各種カテゴリカルデータ（罪名、刑期、生計状況、精神状況など。CFPに係る入力データを含む）、その他同システムに入力されたデータ

検証のポイント（仕様書 第3の2(1)から抜粋）	本提案における対応方針
可能な範囲で、既存のテキストデータについて当局の定めるカテゴリとひも付けるなどの整理を行い、モデル構築に盛り込むこと	**2.1 アセスメント項目の設計・運用の検討** 令和4年度調査研究よりCFPアセスメント情報の項目化と理解。統計的仮説検定を活用した検討により、仕様書 第3の1「既存のデータから適切なアセスメント項目の設計を進める」にも貢献
可能な範囲で、複数時点でのCFPデータを用いた時系列のモデルを検討	**2.2 再犯リスク予測モデルの要件整理** 拡張性や発展性として求められる業務要件を抽出・整理、それらに合致したアルゴリズムを選定する。なお幾つかの要件と予測性能のトレードオフが想定されるため、その優先順位等も確認する
今後のデータの拡張性や発展性（急性リスク予測やある介入を行った場合と行わなかった場合での再犯リスクの差違のシミュレーション等）を考慮、解析アルゴリズムについても当局と協議	
統計的仮説検定や、機械学習、確率モデリングなどを用い、学習データ自体の基礎的な評価や複数モデルでの性能の比較検討等を行って、予測性能のより高いモデルを採用すること	**2.3 初期モデル構築、活用方法の検討** 機械学習や確率モデリングを用いたモデル構築を試行する。両者はモデリングの手順が異なるため別タスクとしているが、データや整理した要件によって実施内容の詳細は変更可能性あり
実装環境について複数の選択肢を前提として、汎用的なアルゴリズムを用いるとともに、特定事業者の製品に過度に依拠せずAIエンジンの変更や拡張が容易となるような仕様とすること	本調査研究におけるモデル構築では、OSSのRまたはPythonを活用する

初期モデル構築プロセスの全体像

- 本提案において対応するタスクと、それらの順番・依存関係の整理（下図）
- 作業項目2.1と2.2は並行して検討し、その結果を基に再犯リスク予測の初期モデルの構築を試行



分析対象のデータ

■ 今般の業務の分析対象として法務省から提供のあったデータ

- 平成２０年６月１日～令和５年２月末日の期間に保護観察を開始した事件（１～４号別。１号短期・交短除く）のうち下記の情報

  ①保護観察開始時に事件管理システムに入力されるデータ一式
  （CFP開始以降は開始時統計的分析及び要因分析のデータ含む）
  ②保護観察開始後５年以内の保護観察又は生活環境調整の再係属の有無と再係属日
  （令和５年３月末日までに再係属したものに限る）

- 提供されたデータ（全27テーブル、838項目）の詳細は、別添のExcelファイル「chapter2-2.xlsx」の「全データ項目リスト」シートに記載

1. アセスメント項目の設計･運用の検討

CFPテキストデータにおける出現単語の可視化（手法について）

前頁の分析結果では、簡易的な分析としてCFPのテキストデータをそのまま使用しており、助詞など些末な違いで同じ意図のものも、別の内容として扱っていたが、各カテゴリ（要因区分）における主要なトピック（話題）を理解し、8カテゴリの下に「中カテゴリ」等を作成するヒントを得ることを目的に、以下のような分析を行った。

■ 実施内容

- CFPテキストデータに「形態素解析」を実施し、各カテゴリで出現する単語やその頻度について分析
- 要因区分の8カテゴリごとに、ポジネガ等を考えず話題を検討するため、1文字以上を含む約22.2万レコードから、「名詞」のみ件数上位100件を単語を抽出

■ 実施環境

項目	概要
形態素解析	MeCab (ver 0.996)
表現辞書	mecab-ipadic-neolog (ver 102)
使用言語	R (version 4.2.2)
パッケージ	RMeCab (ver1.10)
Platform	x86_64-pc-linux-gnu (64-bit)

- ただし、本分析では辞書はデフォルトのまま使用 （個別単語登録等のカスタマイズは行わない）
- 環境構築における参考情報： **2019年末版 形態素解析器の比較 - Qiita**

アセスメント項目の反応性・有効性の検証

検証の目的

- **アセスメント項目のスクリーニングについて**
  - 前頁まで検討したテキストデータ項目以外にも、「アセスメント項目」の対象となり得る項目は多いと考えられるが、これら項目の中には、再犯リスクと相関が無い、影響を与えないものもある
  - リスク予測モデル構築では、このような項目を除外する(スクリーニングする)ことで、予測性能を向上させたり、処理時間を短縮させることができる

- **アセスメント項目の運用フロー設計について**
  - 追加当初は有効と思われるアセスメント項目も、長期間の運用を通じて有効性が低下するものもある
  - また、アセスメント項目を追加し続ければ現場の負荷が増す可能性もある
  - こうした事態を避けるため、アセスメント項目の反応性・有効性について評価・検証する方法を検討し、アセスメント項目の修正・削除などを考慮した運用フローを設計することが重要と考える

そこで本調査研究では、

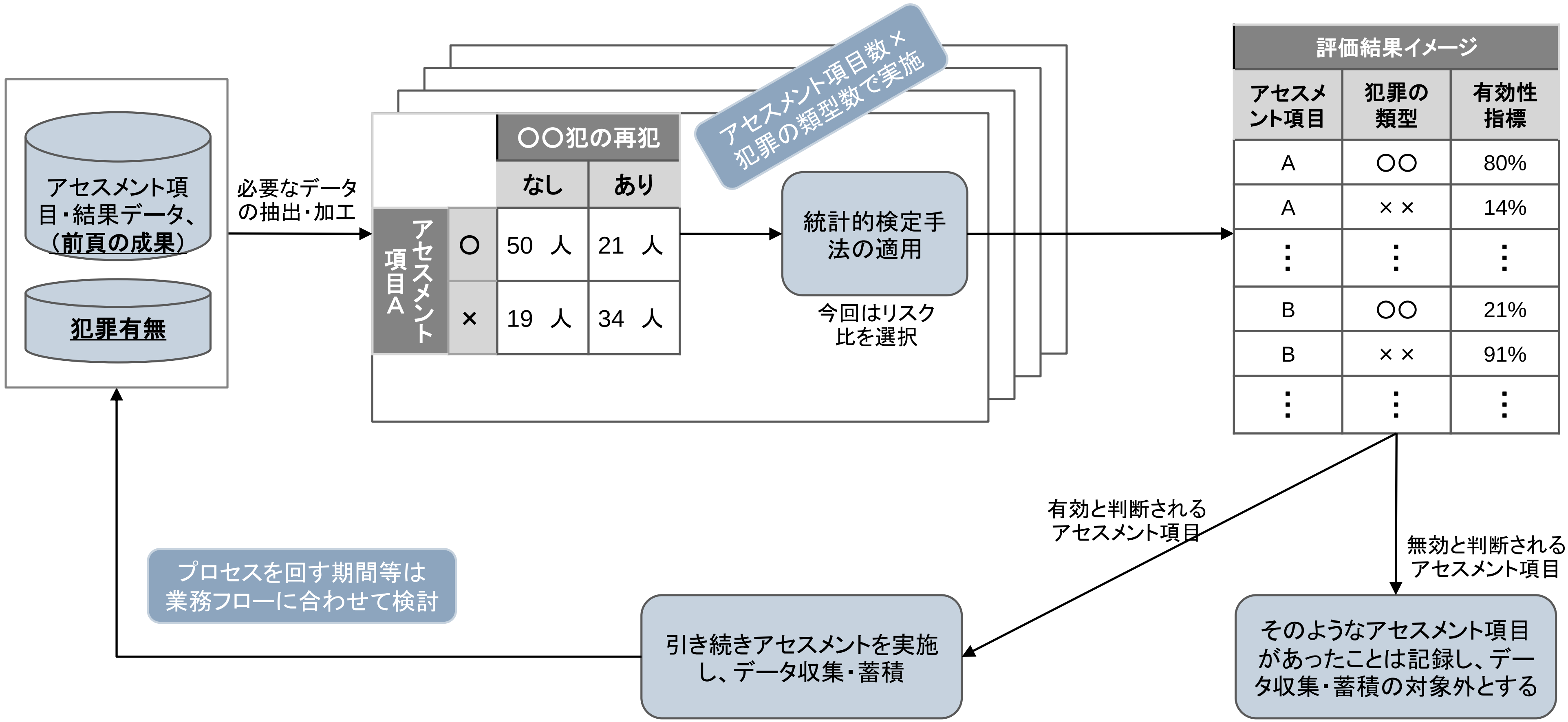
- アセスメント項目と犯罪種別の関係性を統計的に検定する手法を用いた評価プロセスを設計
- アセスメントについては仮想的に実行可能性を検証

これらにより反応性・有効性が認められるアセスメント項目を今回の「リスク予測モデルの構築」において活用することに。

アセスメント項目の反応性・有効性の検証

検証の手順

- 運用イメージを含めた検証手順（下図）
- 統計的検定手法としては、今回はリスク比を採用（次頁）



アセスメント項目の反応性・有効性の検証

リスク比とは

- 「リスク比」は疫学における指標の1つであり、1を超えていればリスクがあると言える（下図）。
  - 喫煙における発がんリスクなどの評価で使われる指標
  - 他の指標には、オッズ比、リスク差、連関係数、点相関係数Φ、カイ二乗検定のp値などがあるが、これらと比べて概念的にわかりやすいのが特徴の1つ
  - 本件は業務に近い方々による活用が想定されるため、わかりやすさを重視してリスク比を採用した

		○○犯の再犯	
		なし	あり
アセスメント項目A	○	A	B
	×	C	D

<u>リスク比とは、</u>アセスメントに該当しなかった群の再犯リスクに対する、該当した群の再犯リスクの比率のこと

$$\text{リスク比} = \frac{\frac{B}{A+B}}{\frac{D}{C+D}}$$

← アセスメントに該当したときの再犯リスク

← アセスメントに該当しなかった時の再犯リスク

- リスク比は得られたデータによる推定のため、例えば真の値があったとして、それと厳密には一致しない可能性がある。このため点ではなく区間で推定する統計的な手法を用いる。95%信頼区間とは、今回と同様にデータを収集して計算すれば100回に95回はこの区間に値が含まれる、と推定される区間を指す。
- リスク比の95%信頼区間は $e^{ln(RR) - 1.96\sqrt{1/a + 1/c - 1/(a+b) - 1/(c+d)}} \sim e^{ln(RR)+1.96\sqrt{1/a + 1/c - 1/(a+b) - 1/(c+d)}}$ で計算できることが知られている。

（参考） リスク比の信頼区間を求める #データ分析 - Qiita

アセスメント項目の反応性・有効性の検証

検証結果

■ 検証結果

- リスク比やその有意性について、当初想定との大きな乖離はなく、スクリーニングとして有効であることを確認
- 詳細は別添のExcelファイル「chapter2-1.xlsx」の「リスク比ほか」シートを参照
  - B～D列はデータ全体でリスク比と95%信頼区間を算出
  - その右の列から順次1～4号観察それぞれに絞ったデータでリスク比と95%信頼区間を算出
  - 最右列に各データ項目におけるNAの比率や1の比率を算出。詳細はⅡ章の第3節を参照

1		リスク比-	下限95%C	上限95%C	リスク比-	下限95%C	上限95%C	リスク比-	下限95%C	上限95%C	リスク比-	下限95%C	上限95%C	リスク比-	下限95%C	上限95%C	NA比率	1の比率
2	要因101	12.2592	10.1731	14.7731	0.8230	0.4457	1.5197	0.9720	0.7912	1.1942	1.0640	0.7832	1.4456	1.3661	0.4914	3.7977		
3	要因102	16.1364	14.5597	17.8838	2.7900	1.4264	5.4574	1.2406	0.8740	1.7610	1.4363	1.2988	1.5883	0.6978	0.3598	1.3535		
4	要因103	15.9915	15.3737	16.6341	1.2434	1.0753	1.4379	1.1015	1.0368	1.1704	1.2048	1.1501	1.2622	1.3260	1.1012	1.5967		
5	要因105	9.4569	7.7530	11.5353	0.7404	0.4509	1.2156	0.9961	0.7606	1.3046	1.2192	0.9491	1.5662	0.7854	0.3105	1.9871		
6	要因106	10.8544	9.9676	11.8201	1.3863	1.1316	1.6984	1.1985	1.0850	1.3237	0.8545	0.7552	0.9668	0.8485	0.5831	1.2347		
7	要因107	9.8427	8.3304	11.6296	1.2545	0.6950	2.2643	1.1097	0.8457	1.4562	1.1833	0.9900	1.4143	0.3680	0.1782	0.7597		
8	要因108	18.7764	18.1906	19.3810	1.3249	1.2024	1.4599	1.1761	1.1111	1.2449	1.1979	1.1553	1.2421	1.0551	0.9122	1.2204		
9	要因201	18.6735	17.9751	19.3989	1.6299	1.4493	1.8329	1.3158	1.2488	1.3864	1.3217	1.2528	1.3944	2.1488	1.7851	2.5866		
10	要因202	9.1868	4.8548	17.3843				0.9563	0.4672	1.9571	1.8159	1.0310	3.1984					
11	要因203	8.4827	7.4851	9.6131	0.7280	0.4350	1.2183	1.1411	0.9538	1.3651	0.9761	0.8488	1.1225	0.3717	0.2260	0.6113		
12	要因204	12.8260	12.0023	13.7063	1.3579	1.1417	1.6149	1.1023	1.0074	1.2061	1.1601	1.0600	1.2696	1.2955	1.0058	1.6688		
13	要因205	10.7057	8.6431	13.2604	1.9991	1.4524	2.7516	1.0465	0.8132	1.3468	0.4534	0.2408	0.8537	0.8727	0.3012	2.5287		
14	要因206	9.8451	5.8123	16.6760	1.4857	0.4373	5.0478	1.2757	0.8223	1.9790	0.2723	0.0424	1.7479	1.8230	0.6825	4.8695		
15	要因207	17.2429	16.6742	17.8309	1.3731	1.2394	1.5213	1.1133	1.0542	1.1758	1.2485	1.2004	1.2985	1.3663	1.1821	1.5791		
49	要因707	18.2682	13.1146	25.4469				0.7967	0.2989	2.1236	1.6351	1.1867	2.2529					
50	要因708	23.8377	21.8551	26.0001	5.5004	4.6296	6.5349	1.6149	1.5700	1.6612	1.4780	1.2717	1.7177	2.3447	1.3184	4.1702		
51	要因709	18.9561	18.3491	19.5832	1.8322	1.6630	2.0187	1.3027	1.2353	1.3738	1.3154	1.2664	1.3662	1.2013	1.0370	1.3917		
52	要因710	23.2454	19.7616	27.3433							1.8995	1.6236	2.2224	2.4321	0.9629	6.1435		
53	要因711	14.8546	14.2276	15.5093	1.3606	1.1708	1.5810	1.1929	1.1226	1.2675	1.1798	1.1205	1.2422	0.8428	0.6809	1.0431		
54	要因801	12.7948	11.7806	13.8962	1.7472	1.2591	2.4245	1.0337	0.9176	1.1646	1.0455	0.9426	1.1595	0.8292	0.5774	1.1910		
55	要因802	7.7021	3.5462	16.7282	1.1139	0.1860	6.6701	1.1957	0.6786	2.1069				1.2132	0.2024	7.2705		
56	要因803	13.9654	9.1436	21.3300	0.5567	0.0852	3.6373	0.9561	0.6319	1.4466	2.0429	1.1599	3.5981					
57	要因804	14.4671	13.4852	15.5205	0.9842	0.6482	1.4943	1.0336	0.9114	1.1722	1.2121	1.1188	1.3132	1.4866	1.1269	1.9612		
58	要因805	14.7733	10.0336	21.7520	1.1797	0.4221	3.2973	1.4514	1.2016	1.7531	1.0894	0.3723	3.1875					
59	要因806	13.9217	12.0573	16.0744	1.8411	1.2640	2.6817	1.3335	1.1493	1.5473	1.0818	0.8789	1.3317	1.6110	0.9514	2.7280		
60	要因807	13.3291	12.2933	14.4522	1.2875	0.9742	1.7016	1.2527	1.1363	1.3811	1.0820	0.9769	1.1985	1.1123	0.7584	1.6314		
61	要因808	13.4858	11.3122	16.0771	0.7155	0.2454	2.0864	1.3149	1.0534	1.6414	1.1827	0.9754	1.4341	0.5809	0.1535	2.1991		
62	要因809	14.1490	12.2166	16.3871	1.7091	1.0668	2.7382	1.1580	0.9549	1.4043	1.0194	0.8392	1.2383	1.6824	0.6220	4.5503		
63	要因810	14.4979	12.7389	16.4997	0.6071	0.1618	2.2775	1.1698	0.8611	1.5893	1.3301	1.1629	1.5213	1.4086	0.8889	2.2321		
64	要因811	10.6365	6.9626	16.2490				1.5948	1.5516	1.6393	0.8809	0.5417	1.4327	1.4561	0.2519	8.4168		
65	刑基本情報_ICHIBU_YUYO_FLG	1.8207	1.6007	2.0709							2.4483	2.3289	2.5738	1.0158	0.4729	2.1821	91.39%	1.15%
66	刑基本情報_1_刑区分ｺｰﾄﾞ.093._01	3.9254	3.7420	4.1178				0.8209	0.7526	0.8953							1.12%	3.87%
67	刑基本情報_1_刑区分ｺｰﾄﾞ.093._02																1.12%	4.75%
68	刑基本情報_1_刑区分ｺｰﾄﾞ.093._03	4.0604	3.3221	4.9628				1.6026	1.5451	1.6623							1.12%	0.16%
69	刑基本情報_1_刑区分ｺｰﾄﾞ.093._04	1.4236	1.0286	1.9701				0.7934	0.6246	1.0079							1.12%	0.19%
70	刑基本情報_1_刑区分ｺｰﾄﾞ.093._05	0.1903	0.1796	0.2016	0.7354	0.6648	0.8136										1.12%	36.24%
71	刑基本情報_1_刑区分ｺｰﾄﾞ.093._06	0.2325	0.2076	0.2604										0.1806	0.1580	0.2064	1.12%	9.40%
72	刑基本情報_1_刑区分ｺｰﾄﾞ.093._07	2.7320	2.6342	2.8334							0.6452	0.5919	0.7032	5.3922	4.7194	6.1609	1.12%	43.73%
73	刑基本情報_1_刑区分ｺｰﾄﾞ.093._08	1.3620	0.7629	2.4316							1.4987	1.0079	2.2286				1.12%	0.06%
74	刑基本情報_1_刑区分ｺｰﾄﾞ.093._09																1.12%	0.00%
75	刑基本情報_1_刑区分ｺｰﾄﾞ.093._10																1.12%	0.00%
76	刑基本情報_1_刑区分ｺｰﾄﾞ.093._13																1.12%	0.00%
77	刑基本情報_1_刑区分ｺｰﾄﾞ.093._15																1.12%	0.00%
78	刑基本情報_1_刑区分ｺｰﾄﾞ.093._17	1.8323	0.9331	3.5979							0.9078	0.5154	1.5989				1.12%	0.03%
79	刑基本情報_1_刑区分ｺｰﾄﾞ.093._18																1.12%	0.00%
80	刑基本情報_1_刑区分ｺｰﾄﾞ.093._27																1.12%	0.00%
81	刑基本情報_1_刑区分ｺｰﾄﾞ.093._28																1.12%	0.01%
82	刑基本情報_1_刑区分ｺｰﾄﾞ.093._40	6.1759	5.6537	6.7463	1.3637	1.2327	1.5085										1.12%	0.55%
83	刑基本情報_1_刑区分ｺｰﾄﾞ.093._41																1.12%	0.19%
84	刑基本情報_1_刑区分ｺｰﾄﾞ.093._42																1.12%	0.29%
85	刑基本情報_1_刑区分ｺｰﾄﾞ.093._50	23.3927	18.3673	29.7931				1.0635	0.8342	1.3559							1.12%	0.01%
86	刑基本情報_1_言渡.決定裁判所大分類コード_090																1.12%	0.00%
87	刑基本情報_1_言渡.決定裁判所大分類コ	0.6533	0.4746	0.8994							0.4149	0.3015	0.5709	1.3520	0.6109	2.9920	1.12%	0.44%
88	刑基本情報_1_言渡.決定裁判所大分類コ	1.0694	1.0122	1.1298	0.8888	0.7525	1.0498	0.8809	0.7909	0.9811	0.8784	0.8325	0.9268	1.4182	1.1455	1.7559	1.12%	10.04%
89	刑基本情報_1_言渡.決定裁判所大分類コ	0.8825	0.8196	0.9503	1.4147	1.2137	1.6490	1.0060	0.9002	1.1241	0.9428	0.8676	1.0246	0.8868	0.6527	1.2048	1.12%	6.29%

chapter2-1.xlsxの「リスク比ほか」シート
より一部抜粋

アセスメント項目の反応性・有効性の検証

活用方針

- 前頁の結果を受けて、リスク予測モデル構築へのアセスメント項目の活用の方針を以下のように策定
  - 95%信頼区間の下限が（すなわちリスク比も上限も含めた3つ組が全て）1より大きければ、
    そのアセスメント項目が再犯リスクを高めることに有意に影響があると考え、リスク予測モデル構築に活用する
  - 95%信頼区間の上限が（すなわちリスク比も下限も含めた3つ組が全て）1より小さければ、
    そのアセスメント項目が再犯リスクを低くすることに有意に影響があると考え、リスク予測モデル構築に活用する
  - 95%信頼区間の上限と下限の間に1を含む際は、
    そのアセスメント項目が再犯リスクに対して有意に影響を与えないと考え、リスク予測モデル構築の際には除外する
  - 値が算出されていない箇所は、算出に必要なデータが無い（NAが多い、1が少ないなど）ケース。
    リスク予測モデル構築の際には除外する。（後述する欠損値に関する前処理のロジックにおいても除外）

- リスク比等の算出結果を上記方針に基づいて、アセスメント項目の反応性・有効性を検証する手順に活用する
- 反応性・有効性があるアセスメント項目を、以降のリスク予測モデルの構築に活用する。詳細はⅡ章の第3節を参照

2. 再犯リスク予測モデルの要件整理

本節の実施目的

再犯リスク予測モデルの要件整理（作業項目2.2）

- 仕様書上に挙げられるキーワードの中には、リスク予測モデル等において実現すべきものか、実現するとしたらその方式などについて、曖昧なものがある。
- これらについて対応方針を明確にすることを目的とし、以下の検討を実施した。

1. 業務における利用シーンと要件の整理

- リスク予測モデルの利用シーンを確認
- 仕様書上に挙げられる**キーワードが、どのような利用シーンのニーズに該当するか**の観点から、要件として整理・具体化

2. 法規制・倫理等に関する要件の整理

- 仕様書上で提示された「法規制・倫理等」の観点についての考慮
- 対応すべき具体的な法律・条文は示されていないため、**観点としてAIプロダクト品質保証ガイドライン(QA4AI)**を活用

3. 要件に基づく対応方針等の具体化

- 左記の**2つを統合**した上で、リスク予測モデル、またはその周辺でどのように実現するかについての**対応方針等を検討・具体化**
- リスク予測モデル構築（作業項目2.3）の実施手順を定義

1. 業務における利用シーンと要件の整理

再犯リスク予測モデルの要件整理（作業項目2.2）

- 法務省内におけるコミュニケーションのシーンやタイミングを基に得られた要件を整理（下図）
- ヒアリング詳細は別添のExcelファイル「chapter2-2.xlsx」の「利用シーンから抽出される要件の検討」シートを参照

保護観察対象者 ⇔ 現場担当者（保護観察官など） ⇔ 法務省保護局担当者

<table>
<tr><th>タイミング＼箇所</th><th>対象者⇔現場担当者</th><th>現場担当者⇔保護局担当者</th></tr>
<tr><td>保護観察開始時</td><td>■ 開始時点での再犯リスクの把握</td><td rowspan="3">■ 政策等による再犯可能性の抑止効果を数値で予測・検証
　■ 政策との相関有無にかかわらず、CFPアセスメント項目は説明変数とする。
　■ 施策に関する変数（施策の対象であるかどうかや施策の結果等）を説明変数とし、施策の再犯抑止効果を予測したい。

■ アセスメント項目の設計への活用
　■ 令和5年度からCFPテキストデータがチェック項目化されている。これらは、理論上または現場感覚として有意と考えられる。この有意性を検証・改善していきたい。</td></tr>
<tr><td>保護観察中</td><td>■ 観察中に新たな情報を把握した時点での再犯リスクの把握
　■ ある時点において対象者に介入する必要があるかを判断したい（その時点に至るプロセスは必ずしも考慮しなくて良い）。</td></tr>
<tr><td>仮釈放時</td><td>■ 仮釈放時点の情報による再犯リスクの把握
　■ 担当官が「今、仮釈放して良いか」を考える意図</td></tr>
</table>

2. 法規制・倫理等に関する要件の整理

再犯リスク予測モデルの要件整理（作業項目2.2）

- AIプロダクト品質保証ガイドライン(QA4AI)を観点としてヒアリング、利用シーンから得られなかった要件を抜粋（下図）
- ヒアリングの詳細は別添のExcelファイル「chapter2-2.xlsx」の「法規制・倫理等に関する要件の検討」シートを参照

- 再犯リスク予測モデルはアセスメントの補助的な位置づけ（計2か所）
- 予測性能の充実（計2か所）
  - 予測性能に重きを置いている。
  - 仕様書から想起された「人の仮説を追加できるモデルの構造である」といった、目的に応じて個別に改編が求められるような再犯リスク予測モデルの利用シーンは想定されない。また仕様書にある確率モデル（ベイジアンネットワーク等）の採用についても同様である。

- 説明性、解釈可能性
  - 解釈可能性（予測ロジックが数式等で可視化できる、white-boxであること）は不要
  - 説明可能性（予測に影響する要因の可視化）は必要
  - 大事なのは再犯リスク予測モデル全体についてだけではなく、各対象者の何について介入したらリスクが下がるかについての知見。対象者ごとの処遇方針の検討に役立てたい。

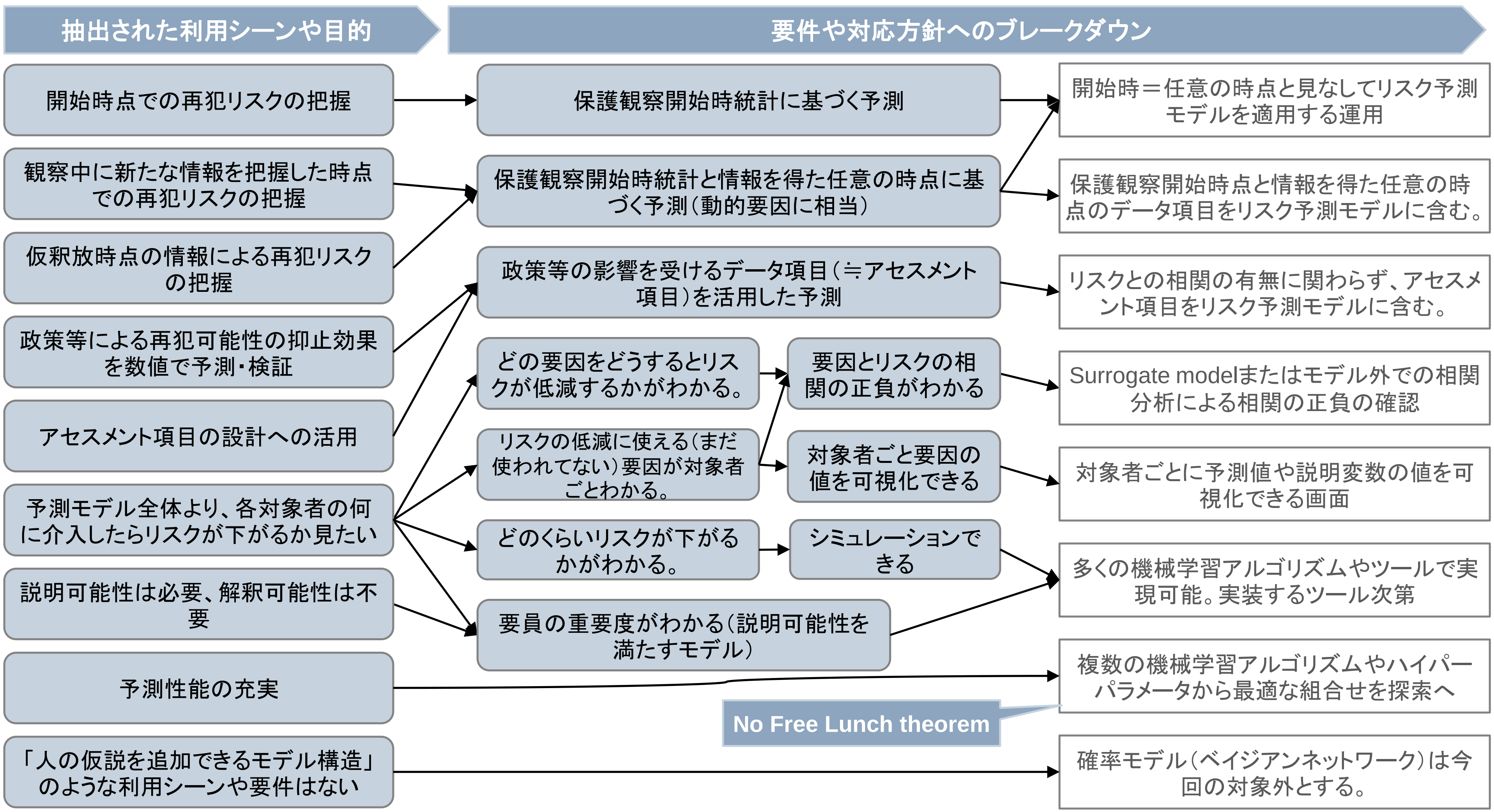
- 動的要因（1か所）
  - 前頁「観察中に新たな情報を把握した時点での再犯リスクの把握」等に対応

- 要配慮個人情報への配慮（計2か所）
  ⇒ 再犯リスク予測モデルに限らず、省内でのデータの扱いとして共通する要件のため割愛
- 試験用画面サンプルとして記載された、再犯リスク予測モデルの要件に関連しない議論も本節では割愛

3. 要件に基づく対応方針等の具体化(1/2)

再犯リスク予測モデルの要件整理（作業項目2.2）

■ 1～2から抽出された要件を整理、対応方針へのブレークダウンを実施



3. 要件に基づく対応方針等の具体化(2/2)

再犯リスク予測モデルの要件整理（作業項目2.2）

■ 仕様書で出現するキーワード（確率モデル等）も、利用シーンや目的に紐づかないため、要件として不要と考える

対応方針へのブレークダウン

- 開始時＝任意の時点と見なしてリスク予測モデルを適用する運用
- 保護観察開始時点と情報を得た任意の時点のデータ項目をリスク予測モデルに含む。
- リスクとの相関の有無に関わらず、アセスメント項目をリスク予測モデルに含む。
- Surrogate modelまたはモデル外での相関分析による相関の正負の確認
- 対象者ごとに予測値や説明変数の値を可視化できる画面
- 多くの機械学習アルゴリズムやツールで実現可能。実装するツール次第
- 複数の機械学習アルゴリズムやハイパーパラメータから最適な組合せを探索へ
- 確率モデル（ベイジアンネットワーク等）は今回の対象外とする。

実施する作業の詳細

■ データは再犯有無が判定できる時点のもののみであり、**保護観察開始時と最終時点のデータによるリスク予測モデルの構築**
■ 実務運用は後者に開始時や任意の時点を投入する「使い方」

■ **現時点のアセスメント項目8カテゴリ**を活用
■ アセスメント項目（有効性評価も）の深堀は翌年度以降

■ 上記を実施できるかデータ加工・特徴量作成が最優先課題
■ そのために**再犯有無との相関を見た特徴量の絞り込み等**

■ **リスク予測モデルを構築してから詳細を検討**
■ 大半の説明変数がカテゴリ値（yes-noなど）であるため、ロジスティック回帰をSurrogate modelとすることも有力な選択肢
■ shap値でも影響の正負を確認できる。

■ **サンプル画面の要件「対象者ごとに見る」**

■ 原理的には可能だが、様々な試行錯誤を一気通貫で効率的かつ柔軟に実現できるツール調査も実施

■ 仕様書から想定された解釈可能性やデータの無い仮説の追加などの要件は不要。加えてデータ項目が大量であり、データ起点の構造推定は計算量的に非現実的、仮に実装できても人の目で確認しきれない等の課題があるため、不採用とした。

運用

説明変数

Uncle Bernie's Rule / one in ten rule

画面

アルゴリズム

リスク予測モデル構築の手順

■ 一般的な機械学習アルゴリズムを適用する予測モデル構築は、以下の4ステップ

Step1. 基礎分析と詳細設計

- 分析対象となるデータの分布や基本統計量、欠損値の発生状況、相関関係などの確認する。
- 上記の確認を踏まえて、特徴量の設計や利用するツール、実装方法、評価指標など、分析手順の詳細を設計する。

Step2. データ加工・前処理

- データのクリーニング、名寄せ、欠損値への対応など、分析できる状態にデータを加工する。
- 必要であれば特徴量を設計する。また事前に目的変数との相関分析などによって利用する特徴量・説明変数を選別する。

Step3. リスク予測モデル構築

- 後述で選定するアルゴリズムによってリスク予測モデルを構築
- 必要に応じてハイパーパラメータのグリッドサーチ、設計・実行

Step4.リスク予測モデル評価

- 評価指標（F-値、precision、recall、accuracyなど）の算出
- ホールドアウト／交差検証などによる汎化性能の評価
- 必要に応じてPR 曲線やROC曲線、最適カットオフ値の算出など

評価結果を受けて、データ加工などの見直しも

■ リスク予測モデル構築においては、判別問題に適用可能な複数のアルゴリズム（次頁）を活用する

- 複数並行してリスク予測モデルを構築し、比較検討を通じて予測性能のより高いモデルを採用することとした

リスク予測モデル構築：二値分類の代表的なアルゴリズム

本調査研究で試行するもの

アルゴリズム		概要
ロジスティック回帰分析		■ 医学・疫学からマーケティングまで、多くの分野における二値分類問題において最もよく使われる多変量解析・統計の手法の1つ。機械学習の枠組みにおいても基本的な手法として用いられる ■ 学習データに過度に適合するすることで本来の傾向から外れ、新たなデータに適合しなくなる「過学習」という問題がある。これに対する解決策の1つとして、AIC（赤池情報量基準）に基づいた変数選択法（ステップワイズ法）を適用する
elastic net (リッジ回帰、ラッソ回帰)		■ 「過学習」を解決する技術的な手法「正則化」を実装した回帰分析の手法に、リッジ回帰（大きな特徴量を抑制に主眼）、ラッソ回帰（不要な特徴量の削減に主眼）がある。それぞれの手法を個々に適用することも有用 ■ この両者のバランスを取って組み合わせた手法がelastic net。両者のバランスはハイパーパラメータにより設定。 ■ 本手法以降は、複数あるハイパーパラメータについて、グリッドサーチよって最適な組合せを探索し、予測性能の向上を図る
決定木分析		■ ツリー構造を作成しながら、目的変数に影響を及ぼしている説明変数を見つけ出し、データを分類していく手法 ■ 多変量解析・統計の分野ではなく、情報科学の分野で開発され、多くの分野における二値分類問題において最もよく使われる機械学習の基本的な手法の1つ
ランダムフォレスト		■ 分析対象のデータからランダムに復元抽出したデータを複数作成し、そのデータごとに複数の決定木分析を実施、それらの推定結果をアンサンブル（多数決や平均値による結果の集約）することで分類等を実現する手法 ■ 単一の決定木分析よりも高い予測精度を持ち、過学習を防ぐことができる。機械学習の分野で広く用いられており、特に分類問題において高い性能を発揮することで知られている
勾配ブースティング	**xgbTree**	■ ランダムフォレストが全ての決定木を並列に処理するのに対して、作成した決定木の誤りを次の決定木が修正するように学習を進める。直列的な処理となるため、ランダムフォレストより時間はかかるが、性能は高くなる傾向が知られている。
	xgbDART	■ Dropout（正則化手法の1つ）を使用することで、xgbTreeの過学習のリスクを軽減するアルゴリズム。 ■ xgbTreeよりも予測精度は高いが、学習に時間がかかる。
	xgbLinear	■ xgbTreeの決定木の代わりに線形モデルを使用するアルゴリズム。特徴量の線形結合で予測できるものには効果的で学習速度が速い、メモリ使用量が少ないという利点がある。しかし非線形の構造の学習はxgbTreeの方が優れている。
LightGBM		■ xgbTreeを軽量化（計算速度向上、メモリ消費少）した、Microsoft開発のアルゴリズム。前年度調査研究でも活用されている。 ■ 大量データの扱いに優れる一方、過学習を起こしやすい。データによってはxgbTreeよりも予測性能が劣ることがある。
ニューラルネットワーク		■ 人間の脳神経系のニューロンを数理モデル化したもの、ディープラーニングの一種 ■ 問題に応じて様々な手法があるが、本調査研究では最も単純な手法(nnetパッケージnnet関数)を利用する。

（参考）決定木系のアルゴリズムを中心に選定した理由

リスク予測モデル構築：二値分類の代表的なアルゴリズム

- 常に最善の予測性能を示すアルゴリズムが無いことを「ノーフリーランチ定理」と呼ぶ。
- しかし、最近の論文では、表形式データ（本調査研究のようにExcelで扱うタイプのデータ）については
  もう少し強い傾向が示されている。

  [2207.08815] Why do tree-based models still outperform deep learning on tabular data? (arxiv.org)
  （訳：表形式データにおいて、ツリーベースのモデルが依然としてディープラーニングを凌駕するのは何故か）

  - ディープラーニング（深層学習）は、画像、テキスト、音声データなどの処理を大幅に進化させた。最近のChatGPTなどの大規模言語モデル（LLM）も、このディープラーニングをベースとしている。しかし、表形式データに対しては、ディープラーニングはあまり有効ではなく、勾配ブースティングなど決定木系（ツリーベース）のモデルの方が予測性能が高い、という調査結果が上記の論文で示されている。

- 上記が機械学習アルゴリズムの候補としてディープラーニング系からはnnet一つしか選定しなかった理由でもある。

- 技術的な詳細については以下の解説記事を参照
  - Why do tree-based models still outperform deep learning on typical tabular data ? なぜ、テーブルデータでツリーベースのモデルがディープラーニング（深層学習）モデルを凌駕するのか？ – セールスアナリティクス (salesanalytics.co.jp)
  - 【論文メモ】Why do tree-based models still outperform deep learning on tabular data? – 行李の底に収めたり[YuWd] (yuiga.dev)

リスク予測モデル構築における利用ツールや実行プログラム

- リスク予測モデル構築では、OSSのRとPythonをツールとして活用
  - 基礎的なデータ加工（後述）については、システム開発などの汎用性を考慮して、Pythonを活用した
  - 複数手法の試行錯誤やハイパーパラメータのサーチなど、リスク予測モデル構築自体を効率的に実施するツールとして、Rのcaretライブラリを採用した
    - 後述するデータの加工や前処理を予測モデル構築の段階でも細やかに実施する必要があるため、AutoML（機械学習の自動化）を志向する他のツールより適すること、また当社が従前より整備・保持していたプログラムを転用できるためデータ授受～加工における時間ロスを取り戻せることの2点が採用の理由

- 別添の「本調達で作成したプログラム一式」フォルダには、最終的な分析結果の作成に使用したプログラムのみを整備し、格納している
  - データ加工からリスク予測モデル構築の一連の流れに対応するプログラムを対象とした
  - データの確認や差替え、再加工や再分析を繰り返す中で生じた、最終的な分析結果にはつながらないプログラムは対象とはしていない

- 本稿でも主要な分析プロセスおよび最終的な分析結果を記載
  - 差替え前のデータによる分析や考察の詳細は、別添の「会議資料・議事録」フォルダの定例打合せごとの該当する成果物を参照

基礎的なデータ加工の実施内容

3. 要件に基づく対応方針等の具体化

- 一般に、ITシステムに格納されているデータを、そのままの形では分析には使えない
- 今回はヒアリング等を通じてテーブル間の複雑な構造などを解き明かしつつ、分析可能となるようにデータ加工・特徴量作成を実施した

- 最終的なデータ加工の内容については別添のExcelファイル「chapter2-2.xlsx」の「データ処理状況」シートを参照
  - 裁判身柄IDを基準に各テーブルをマージした
  - カテゴリカルデータのダミー変数化といった機械的な処理のほか、CFPに関するデータは法務省へのヒアリング結果を受けて実務での扱いに対応した処理を実施し、特徴量は約4,300項目。
  - この特徴量の絞り込みは、前述のリスク比や後述の欠損率などによって実施

- データの確認や差替え、再加工や再分析を繰り返す中で生じた、最終的な分析結果にはつながらない分析結果などの詳細は、別添の「会議資料・議事録」フォルダを参照のこと

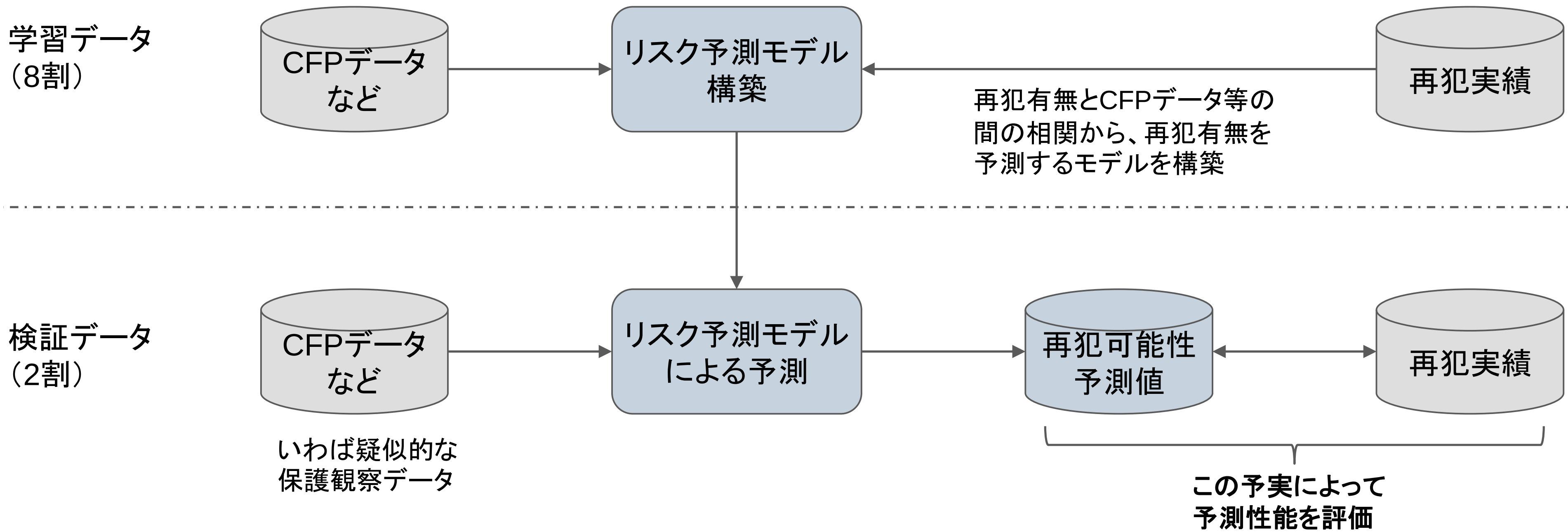
(*) ダミー変数化とは：

カテゴリカルデータ（質的データ）を、機械学習にかけられるように数値データ化する処理のこと。
例えばA、B、O、ABという4つの値をとる「血液型」というカテゴリカルデータがあったとする。
このとき「血液型A」「血液型B」「血液型O」「血液型AB」の4つの項目を設け、その血液型に該当すれば1、該当しなければ0というデータに変換することで、数値データとして扱えるようにする処理を指す。

3. 初期モデルの構築

予測モデルの構築（学習）と評価（検証）

- 前述「リスク予測モデル構築の手順」step3〜4の実施イメージ（下図）
  - データを8:2に分け、前者の8割で再犯有無とCFPデータ等の間の相関から再犯有無を予測する機械学習モデルを構築する。このリスク予測モデルを後者の2割のデータに適用し、その予実差から妥当性を評価する

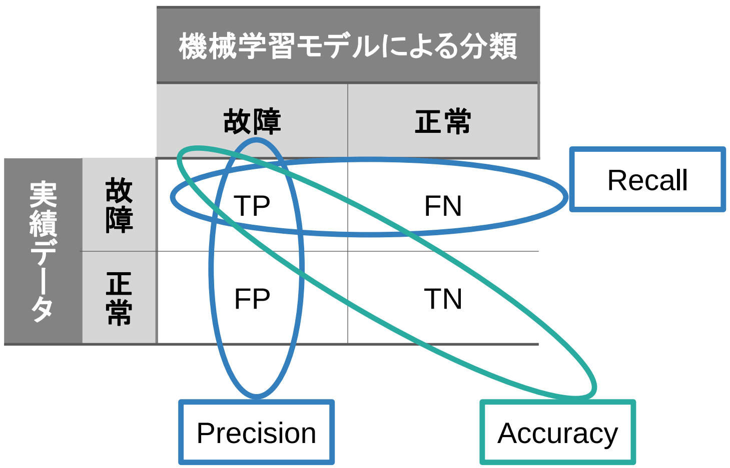


- 再犯率は全体で約26.5%程度であり、再犯有のデータと無しのデータの間で大きな偏りがある。この偏りを放置すると次頁のような適切でない予測モデルが構築されるなどの事象が生じる可能性がある

  ⇒今回はデータ量が十分にあるため、**再犯無しのデータを、再犯有のデータと同量にまで間引く（ダウンサンプリング）**ことで、この事象を回避しつつ予測性能を高める工夫をしている。

予測性能の評価における論点

- 予測性能を評価する観点や指標は多種多様であり、時に複合的に判断する必要があり。また状況によって使えない指標もあり。例えば、故障有無に偏りがある（不均衡データ：左下図）場合には、AUCのような評価指標は安定的に使えないことが知られている。



<u>偏りが大きいデータで解釈を誤る例</u>
頑丈な機械を常に「正常」と予測すれば正解率は高い。
だが業務的には故障の予兆を取りこぼしなく欲しいはず。

		機械学習モデルによる分類	
		故障	正常
実績データ	故障	0	1
	正常	0	99

Recall = 0%

Accuracy = 99%

指標	定義・意味
Accuracy（正解率）	■ (TP+TN) / (TP+TN+FP+FN) ■ 分類結果がどれだけ正しいか直感的に理解・解釈しやすい。 ■ **<u>偏りが大きいデータでは解釈を誤る可能性大</u>**
Precision（適合率）	■ TP / (TP+FP) ■ 故障と予測したもののうち、故障であった割合 ■ 分類したものの正確さ
Recall（再現率）	■ TP / (TP+FN) ■ 実際に故障であったもののうち、故障と予測できた割合 ■ 分類すべき対象の取りこぼしの無さ
F1 score	■ PrecisionとRecallの調和平均 ■ PrecisionとRecallはトレードオフ関係（片方の値が高くなると、もう片方の値が低くなりやすい）、両者をバランスよく見る指標

（出典） [評価指標]AUC（Area Under the ROC Curve：ROC曲線の下の面積）とは？：AI・機械学習の用語辞典 - @IT (itmedia.co.jp)

令和4年度調査研究報告書における予測性能の評価指標と評価結果

予測性能の評価における論点

- 令和4年度調査研究の報告書（令和４年８月）より抜粋
  - [法務省：保護観察におけるアセスメントへのAI導入に関する調査研究結果について (moj.go.jp)]

表 3.10 号種 4 における異なる項目条件での予測性能指標一覧

AUC-ROC には 95%信頼区間、AUC-PR には基準値(検証データの短期再係属該当率)を付記。閾値は感度と特異度の和が最大となるものを記載しており、Accuracy から Negative Predictive Value までの指標は、当該閾値を用いて算出している

性能指標	基礎データのみ	CFP データのみ	基礎データ+CFP
AUC-ROC	0.652[0.494, 0.810]	0.687[0.551, 0.823]	0.692[0.542, 0.841]
AUC-PR	0.092(0.05)	0.137(0.05)	0.095(0.05)
Threshold(閾値)	0.428	0.451	0.492
Accuracy	0.762	0.512	0.858
Recall(感度)	0.533	0.800	0.533
Specificity(特異度)	0.773	0.498	0.874
Precision	0.103	0.072	0.170
Negative Pred. value	0.972	0.981	0.975

- 元の再犯率（0.2～0.3）を下回る評価指標があることは、予測性能という観点では改善の余地があることを示す
  - 例えば実務上、precisionは元の再犯率を下回ることを理解した上で、recallを最優先に調整する、といったことはある
  - しかし、前述のヒアリングでも、令和4年度調査研究の報告書においても、そのような要件はないため、改善のポイントと考える

予測性能を評価する指標・枠組みに関する論点

予測性能の評価における論点

- 前頁の内容に加え、報告書中に特段の記載は無いため、あくまで推測として、令和4年度調査研究では不均衡データ（今回で言えば再犯有無に偏りがある）のまま予測モデル構築を実施した可能性がある
  - 不均衡データの場合、AUC-ROCのような評価指標は安定的に使えないことが知られている。この場合、最終的な評価指標としてAUC-ROCは適切でないことに加えて、令和4年度調査研究の報告書p18では「AUC-ROCが最大になるように調整」とあるため、モデル学習の段階でも不具合が起きていた可能性がある
  - 前述のダウンサンプリングなどの措置をしないまま機械学習アルゴリズムを適用し、十分な予測性能を得られない機械学習アルゴリズムの適用事例は他所でも見受けられるため、この可能性を思慮

- 上記を踏まえ、以下をリスク予測モデル構築の方針とした
  - 学習データは、ダウンサンプリングによって不均衡を是正した後に機械学習アルゴリズムを適用した
  - 検証データは、実態をそのまま反映させるべく、特に措置をせず、そのまま活用した
  - AUC-ROCの活用を限定的にした
    - 学習時は、F-値が最大になるように調整
    - 検証時も、AUC-ROCに依存せず、F-値やPrecision、Recallなどの指標で複合的に評価することとした

リスク予測モデル構築手順の詳細設計

- 前述「リスク予測モデル構築の手順」step3～4について、前頁の検討に従って、以下の詳細手順を設計し、リスク予測モデルを構築・評価した
- これに加えて、号種別や欠損値の扱いのパターン別に、モデル構築～評価も実施した（後述）

Step3. リスク予測モデル構築

- 裁判身柄IDごとに再犯有無を予測するリスク予測モデルの構築
- 整理した要件に沿ったアルゴリズムとそのハイパーパラメータの選定
- 汎化性能を高めるための交差検証 (10-fold cross validation)
- F-値による評価（カットオフ値=0.5）で最適なモデルを選定
- リスク予測モデルに学習データを適用し、最適なカットオフ値を算出

Step4.リスク予測モデル評価

- 構築したリスク予測モデルに検証用データを適用
- 前ステップで算出した最適カットオフ値を適用して再犯有無を判定
- リスク予測モデル自体の評価は、各種評価指標（F-値、precision、recall、accuracy、ROC-AUCなど）を算出し、複合的に評価
- 予測モデルの評価
  - 予測モデル全体で、どの特徴量が影響を与えているかImportance（変数重要度）によって可視化
  - 裁判身柄IDごとには、特徴量の再犯リスクに対する貢献度を「shap値」で評価

リスク予測モデル構築手順の詳細設計（説明補足：1/2）

前頁における専門用語の補足説明

■ <u>**交差検証（cross-validation)**</u>

- 機械学習のモデルの汎化性能を評価する手法の一つ
- データを複数の分割パターンで学習データと検証データに分けてモデルの性能を評価する。その際、データを分割する方法によって、k分割交差検証や層化k分割交差検証などがある
- 本調査研究ではグリッドサーチ（後述）と併用する。具体的には、グリッドサーチで総当たりで構築するモデルの1つ1つで交差検証を実施

■ <u>**10-fold cross validation**</u>

- 上記の交差検証の中で最も一般的な手法の一つ
- データを10個のグループに分割し、そのうち1つをテストデータ、残りを学習データとした交差検証を10回実施する この試行を通じて、10個のテストデータに対する予測精度を平均化することで、モデルの汎化性能を評価する

■ <u>**ハイパーパラメータ（Hyperparameter）**</u>

- 機械学習アルゴリズムの挙動を設定するパラメータのこと
- モデルの性能に大きな影響を与えるため、適切な値を選択することが重要。一般に、ユーザが設定することが多く、機械学習アルゴリズム自体が自動的に最適な値を見つけることはしない

リスク予測モデル構築手順の詳細設計（説明補足：2/2）

■ **グリッドサーチ（grid search）**

- 機械学習のハイパーパラメーターをチューニングする手法の一つ。他のチューニング手法にベイズ最適化など
- ハイパーパラメータを人が恣意的に決めるのではなく、機械的に探索・設定するための方法論
- パラメータの値（の組み合わせ）をグリッド（格子）状に設定し、それらの組み合わせ全てに対して総当たりでモデルを構築・評価することで、モデルの性能を最大化する最適なパラメータの組み合わせを見つけ出す
- パラメータの探索範囲を広く細やかにカバーするほど、予測性能を向上できる可能性が高くなるが、計算時間とのトレードオフがある

■ 今回対象とする機械学習アルゴリズムのハイパーパラメータ数と、グリッド数（下表）

- 例えばelastic netであれば、11^2=121のグリッドをサーチする、つまり121件のリスク予測モデルを構築して（今回はF値の観点で）ベストな予測モデルを選択

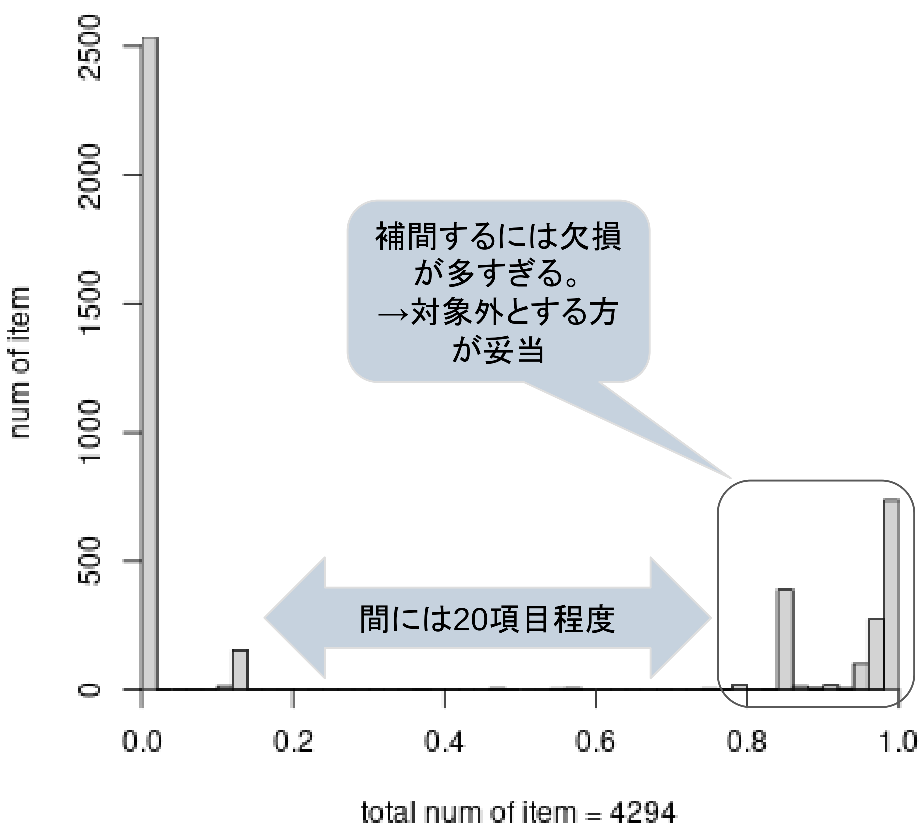
機械学習アルゴリズム	elastic net（リッジ回帰）	決定木分析	ランダムフォレスト	勾配ブースティング			LightGBM	ニューラルネットワーク
				xgbLinear	xgbTree	xgbDART		
caretパッケージのmethod名	glmnet	rpart	rf	xgbLinear	xgbTree	xgbDART	（個別対応）	nnet
チューニング対象のハイパーパラメータ数	2	1	1	4	7	9	3	2
各パラメータあたりの値の数	11	10	10	5	3	3	3	5

令和4年度調査研究に準じて設定

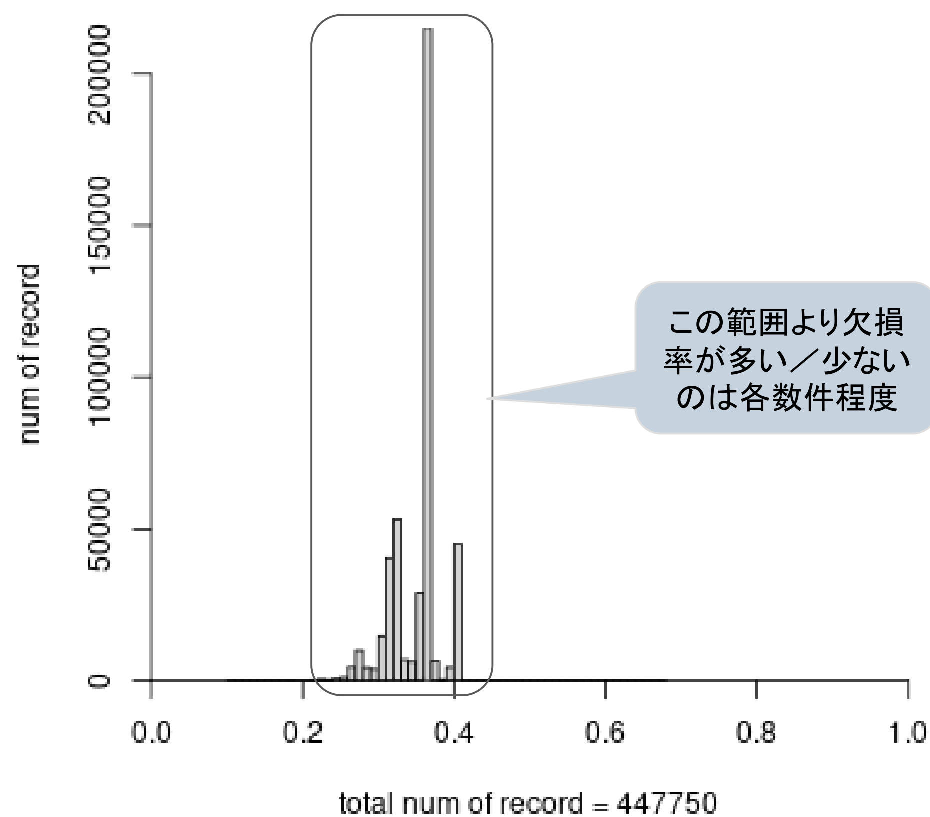
データ欠損率

- リスク比の算出の際にデータ項目（列）単位で算出していた欠損率に加え、行方向でも算出して可視化（下図）
- 「データを増やしさえすればよい」と考えがちだが、入力や管理の状態が悪いテーブルばかり追加しても有用ではない。保護観察IDと紐づかない項目や値が多ければ、データの前処理で作業量が増えるのみ

項目（列）単位の欠損率

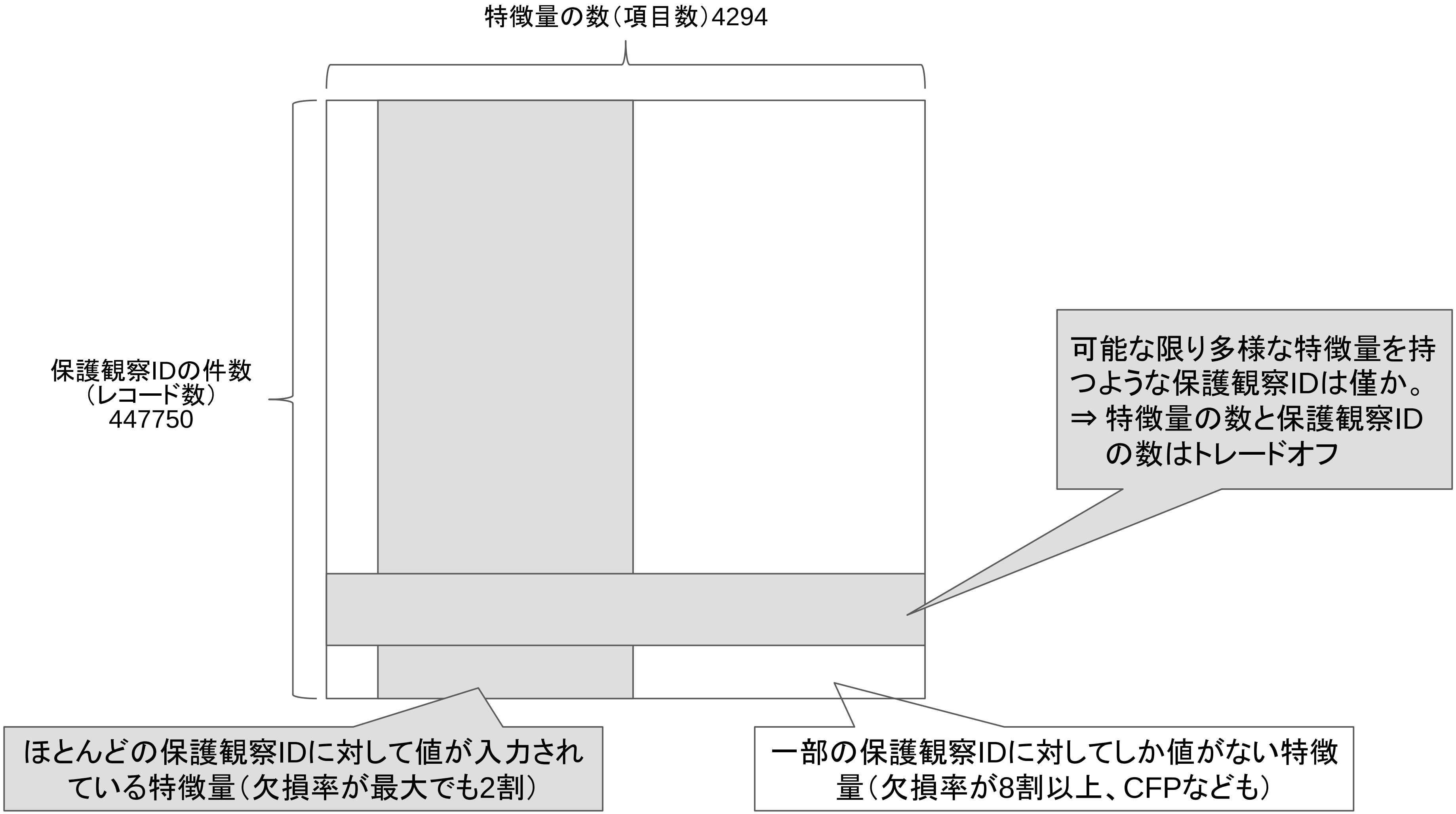


レコード（行）単位の欠損率



データ欠損率（前頁の図解）

データ全体としては、レコード数を確保しようとすると使える（欠損率の低い）特徴量が減る、特徴量を確保しようとすれば
レコード数を確保できない、というトレードオフ状態であった。



データ欠損率の考察

- 欠損率が高い原因は以下の2つが考えられる
  - 個々のテーブルで元々欠損が多かった
  - あるIDが片方のテーブルにあり、もう一方に無いと、マージ後のテーブルにおける後者に由来する項目では、そのIDの分が欠損値となる。今回は30以上のテーブル間で共通の裁判身柄IDが多くなかった

- 極度に欠損値の割合が高い項目は分析対象外とし、データ量を確保することで予測性能を追求することが一般的この方針でのリスク予測モデルの構築も試行した（後述パターン1）。
- 一方、本件で重要視されるCFP情報（含むテキストデータ）や時間的要因を考慮した項目の大半が、欠損値の割合95%を超えるデータ項目であった。これらを全て除外して予測性能を高めても、説明性に欠けるリスク予測モデルとなるため、CFP情報や時間的要因を考慮した項目が欠損でないレコードに絞り込み、リスク予測モデル構築を試行することを検討（後述パターン2）。

<u>リスク予測モデル構築・運用に向けた課題</u>

- 上述の原因の2つ目（テーブル間のIDの対応問題）に該当するものが多い場合、データ加工などの処理に時間やマシンリソースを要する割に予測モデル構築に使えないデータが多くなり、実運用ではボトルネックとなる。実際、使えないテーブルもあった　⇒　省内での適切なデータベース管理が求められる

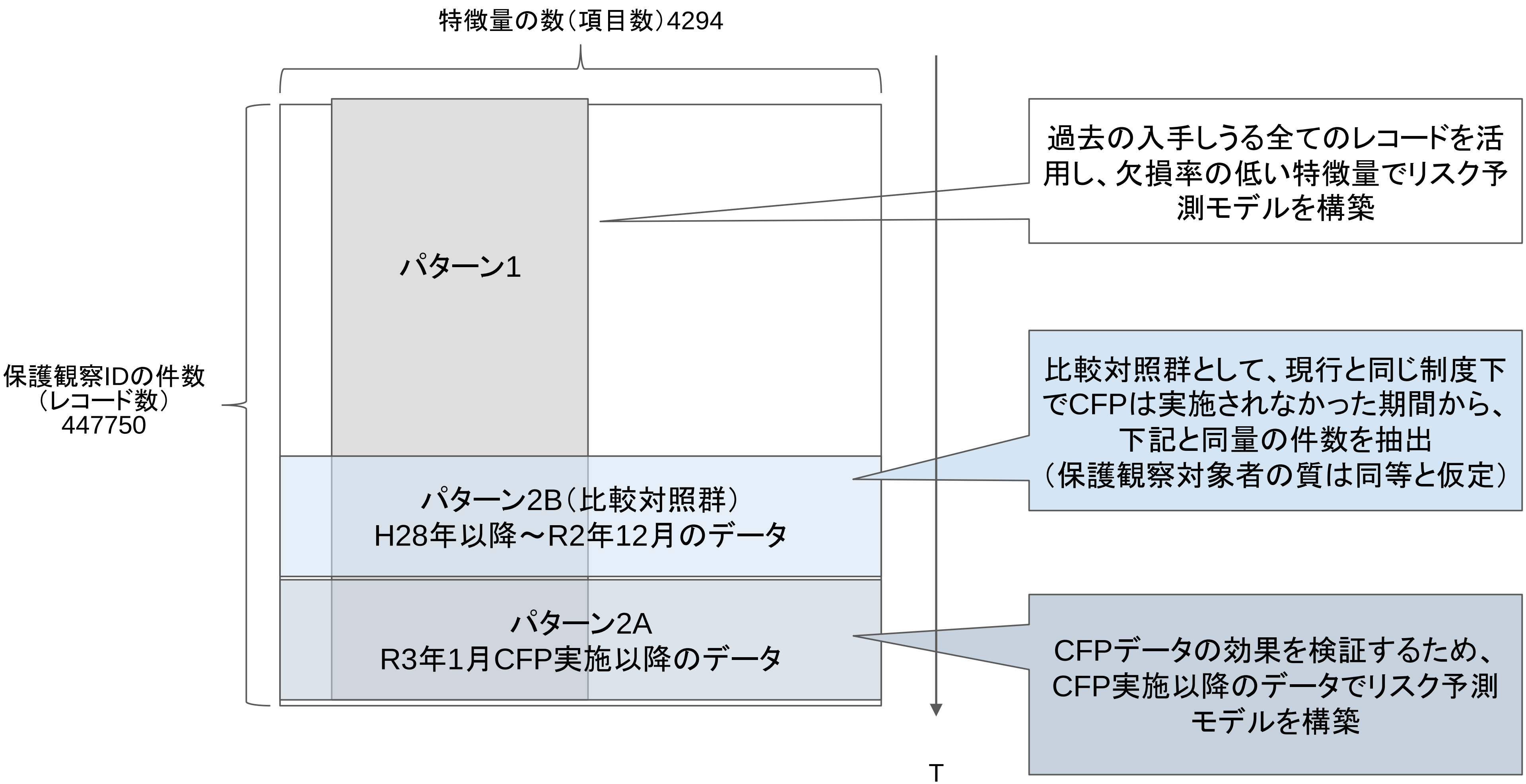
リスク予測モデル構築のパターン（1/2）

前頁のように、欠損率に関する課題があるため、以下2パターンで試行することとした。
加えて、全体、及び、号種ごと（1～4号種の計4種）についても試行した。

1. 項目は少ないが、レコード数（行数）を優先的に確保して学習	■ 欠損値の割合は2割未満の項目に絞り込む（2割を超える特徴量は削除） ■ 欠損値は、0-1データは0で、それ以外の数値データはその中央値で補間 ■ CFPテキストデータは前述のサブカテゴリデータに加工 ■ CFPテキストデータ以外の文字データは、欠損値かどうかの0-1データに変換 ■ 5年再犯率の予測
2. レコードは少ないが、項目数（列数）を優先的に確保して学習	■ CFP情報（含むテキストデータ）や時間的要因を考慮した項目は残るよう、CFPを制度として実施したR3年1月以降のデータを活用（パターン2A） ■ パターン2Aの比較対照群として、H28年以降～R2年12月のデータを活用（パターン2B） ■ 欠損値の扱い等は上記と同様 ■ データの期間を考慮し、1.5年再犯率の予測

リスク予測モデル構築のパターン（2/2）

前頁の図解



分析結果(1/2)

- 分析結果の詳細は別添のExcelファイルにまとめました。構成は以下の通り
  - パターン1の号種1～4、パターン2の全体と号種1～4それぞれで、リスク予測モデルを構築
  - パターン2は、CFPデータがある情報に絞ったA群と、それ以前の期間から同量をサンプリングしたB群の2つで実施

Excelファイル名	概要
chapter2-3分析経過パターン1.xlsx	■ パターン1（全期間）
chapter2-3分析経過パターン2A.xlsx	■ パターン2A ■ R3年1月CFP実施以降で、CFPデータがあるもの
chapter2-3分析経過パターン2B.xlsx	■ パターン2B ■ H28年以降～R2年12月のデータ（CFP未実施） ■ パターン2Aの比較対照群、データ量も上記に合わせてランダムサンプリング

どのパターンでも、「欠損値2割以下」「リスク比が有意（3つ組の全てが1以上or1未満）」でデータを絞り込んでいる。

×

シート名	概要
処理時間	■ アルゴリズムごと号種ごとの処理時間 ■ マシンスペックなどのメモ
欠損率	■ 準備したデータの欠損率
リスク比	■ 準備したデータで算出したリスク比
再犯率	■ 全体・号種ごとの再犯率やレコード数 ■ 欠損率に関する可視化グラフ掲載
評価指標	■ 全体・号種ごとに構築したリスク予測モデルの性能評価指標の一覧 ■ 令和4年度調査研究に準拠
重要度〇〇	■ Lightgbmアルゴリズムで算出された説明変数の重要度、上位20件グラフ

⇒ **以降では、一部の分析結果を抜粋して示しつつ、分析結果の概要を説明する**

分析結果(2/2)

- 次頁から**号種1～4のリスク予測モデルについて以下を掲載**
  - **予測性能**の一覧表
  - リスク予測モデルにおける**変数重要度（上位20件、**LightGBMアルゴリズムで構築したモデルによるもの）

- **利用した計算機のスペックについて**
  - CPU：Intel Xeon E-2278G CPU @ 3.40GHz
  - メモリ：62.6Gbit + 同サイズのSWAP領域
  - GPU：なし
  - ディスク：SSD 1TB
  - OS：Ubuntu 20.04.6 LTS

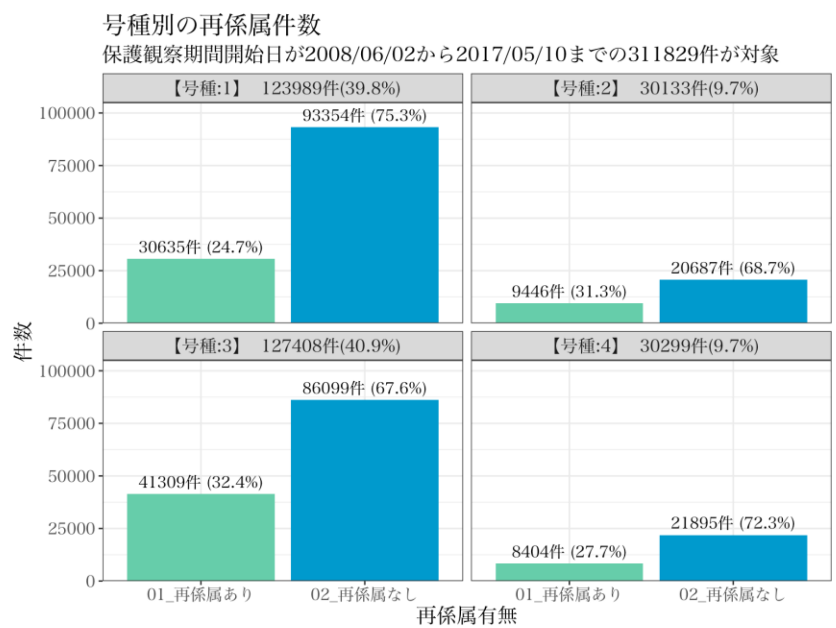
- **今回の複数の機械学習アルゴリズムの比較における検討条件**
  - 計算時間を最大2日間と設定
    - 実運用におけるコスト等の制約として閾値を仮置き。一定以上時間の掛かるアルゴリズムは除外
  - 説明変数のグルーピング・縮約は実施しない
    - 予測性能の向上は期待できるが、元の説明変数がわかりづらくなるため本調査研究では「説明可能性」を優先
  - 多重共線性（説明変数間に相関が高いものが存在する）を避ける分析・処理は実施しない
    - 予測性能の向上は期待できるが、データ項目の操作等にかかる作業工数を考慮

リスク予測モデル構築（パターン1）

- 号種別の再犯率（下表）
  - データ量を確保するため、集計時点で5年を経過していないデータを含む。
  - 項目数は、元々の4294項目から、「欠損値2割以下」「リスク比が有意」で絞り込んだもの

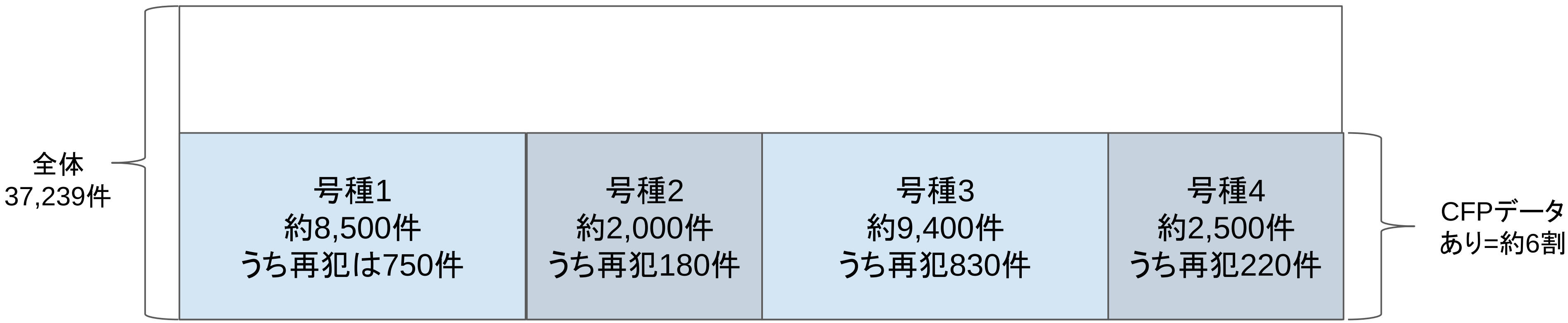
	全体	1号	2号	3号	4号
再犯率	26.54%	23.95%	29.50%	28.67%	24.83%
レコード数	447,750	168,972	40,724	188,532	49,518
項目数	784	786	655	1,126	709

- 令和4年度調査研究における再犯率（右図：抜粋）
- 経験的な再犯率の水準と大きな乖離が無いことを確認



リスク予測モデル構築（パターン2）

- パターン2A（R3年1月～R4年9末のCFP実施期間）における1.5年以内再犯のデータ概算（下図）
  - 全体で37,239件、そのうちCFPデータがあるのは60～65%、再犯率は約8.84%
- パターン2B（対照群：H28年以降～R2年12月のCFP未実施期間、ランダムサンプリングでA群とデータ量は同じ）では、再犯率が全体で約15.8%



- [法務省：再犯防止推進白書 (moj.go.jp)](moj.go.jp) における「出所受刑者の2年以内再入率」で妥当性を検討
  - H28～R2の再犯率が15～17%、パターン2Bはその水準として妥当
  - R3年は14.1%であり、再犯率は減少傾向とはいえ、**パターン2Aはやや乖離**がある
    - 提供データはR5.3.31までの再係属状況しかカバーしていないため、R3.10.1以降の事件は、実質的には1年半より短い期間での再係属状況
    - データ件数も少ないため誤差もより大きく出やすく、本来想定される再犯率よりも低くなる原因になっていると推測

パターン1考察

パターン1. 項目は少ないが、レコード数（行数）を優先的に確保して学習	■ 欠損値の割合は2割未満の項目に絞り込む（2割を超える特徴量は削除） ■ 欠損値は、0-1データは0で、それ以外の数値データはその中央値で補間 ■ CFPテキストデータは前述のサブカテゴリデータに加工 ■ CFPテキストデータ以外の文字データは、欠損値かどうかの0-1データに変換 ■ 5年再犯率の予測

- **得られた評価指標の値から、実務上は十分な予測性能があるリスク予測モデルを構築できた**
  - 令和4年度調査研究の報告書にもある複数の指標（ROC、Recall、Specificity、Precision、Negative Predictive Value）において、令和4年度調査研究の結果を上回っている
  - 特に、前述で指摘した「再犯率」を下回る（≒予測する意味がない）指標はなかった

- **パターン1では、機械学習アルゴリズムの選定としては LightGBMが合理的と考えられる**
  - 予測性能：　LightGBMが最も良かった
    - ただし、例えば号種1～2では、勾配ブースティング（xgbTreeやxgbLinear）などのアルゴリズムも同等の予測性能を示すなど、予測性能では必ずしもLightGBMだけがよいとは言えない
  - 処理時間：　LightGBMが圧倒的に短かった
    - 例えば号種1～2であれば、LightGBMであれば5分以内、勾配ブースティングでは1.5～4時間
    - 号種3では、LightGBMであれば10分程度で予測モデル構築できたが、勾配ブースティングでは2日経っても構築できなかった
  - 処理時間は、予測モデル構築に必要なマシンスペック（≒予算）、運用の難易度などの選定基準に影響する重要な指標

- **総合的にも、LightGBMはベストな選択肢と言える**
  - 今後もレコード件数や項目数が増え、処理時間・マシンスペックもより必要になることを考慮すると、処理時間が圧倒的に短く、良い予測性能を示すLightGBMは、この後しばらく優先的に活用すべきアルゴリズムの一つと言える

パターン2考察（1/2）

パターン2. **レコードは少ないが、項目数（列数）を優先的に確保して学習**	■ CFP情報（含むテキストデータ）や時間的要因を考慮した項目は残るよう、CFPを制度として実施したR3年1月以降のデータを活用（パターン2A） ■ パターン2Aの比較対照群として、H28年以降～R2年12月のデータを活用（パターン2B） ■ 欠損値の扱い等は上記と同様 ■ データの期間を考慮し、1.5年再犯率の予測

- **得られた評価指標の値から、実務上は十分な予測性能があるリスク予測モデルを構築できたと言える**
  - A群、B群ともに令和4年度調査研究の報告書にもある複数の指標（ROC、Recall、Specificity、Precision、Negative Predictive Value）において、令和4年度調査研究の結果を上回っている
  - 特に、前述で指摘した「再犯率」を下回る（≒予測する意味がない）指標はなかった。Negative Predictive Valueは低調だが、令和4年度調査研究におけるprecisionに比べて有用な水準と言える。

- **パターン2のA群は、同B群と比べると、指標によって濃淡はあるものの、概ね予測性能が高くなるものが多い**
  - 例えば、データ全体で分析したときのglmnetアルゴリズムを活用したリスク予測モデルで比較すると、AUC-ROC指標で0.08程度の精度向上が認められる
  - 現時点では、この予測性能の向上の要因が全てCFPによる効果だけとは言えないが、その可能性を示唆

- **パターン2のA群は、パターン1と同等の予測性能を示した**
  - 予測対象とする再犯率の期間（1.5年と5年）などに差異があるため単純な比較はできないが、
    予測性能の向上にとって、CFPアセスメント項目をリスク予測モデル構築に活用することは、
    大量の過去データを活用することと同等の効果を得られる可能性が示唆される

パターン2考察（2/2）

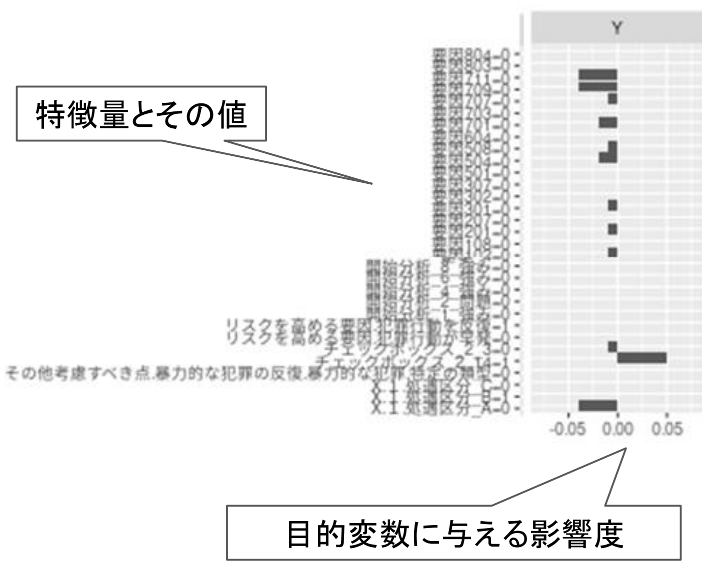
- **パターン2でもLightGBMは総合的に良かったが、他のアルゴリズムも十分選択肢として残る結果になった**
  - 予測性能について
    - A群では、LightGBMも予測性能の高いアルゴリズムの1つだったが、勾配ブースティング（xgbTreeやxgbLinear）やglmnet、rf（ランダムフォレスト）といった機械学習アルゴリズムも同等の予測性能を示した
    - B群では、むしろ他のアルゴリズムがLightGBMを上回った
  - 処理時間について
    - LightGBMの他に、glmnet、rf（ランダムフォレスト）などが選択肢として残る
      - LightGBMは30秒未満で処理完了
      - glmnetやrf（ランダムフォレスト）は数分で処理完了するため、実務的には十分に許容範囲内
      - 勾配ブースティング（xgbTreeやxgbLinear）は1.5～2.5時間であり、採用に際しては、利用する計算機のスペックを検討することが必要

- パターン2（実務ではA群）の予測モデル構築をするのであれば、LightGBMの他に、glmnet、rf（ランダムフォレスト）といった複数のアルゴリズムも候補となる
  - 対象とするデータに対してこれらのアルゴリズムを比較・評価した上で、ベストな予測性能を示すリスク予測モデルを構築・活用するのが、現状では最も望ましい運用方法と考えられる
  - 但し、今後CFPの適用実績が蓄積され、レコード件数や項目数が増えると、処理時間・マシンスペックもより必要になり、パターン1のようにglmnetやランダムフォレストではマシンスペックが足りなくなるといった状況も想定される
  - 今後、新たな機械学習アルゴリズムが出現する可能性もあり、本調査研究で実施したような比較評価を通して、適切なアルゴリズムを選定する必要がある

補論：保護観察対象者ごとに特徴量の影響度を見る方法（SHAP値）

- モデルの「説明可能性」について
  - ここまで、リスク予測モデル全体における特徴量の影響度（大域的な説明）をpermutation importanceと呼ばれる指標で評価してきた
  - 一方、個々の保護観察対象者に着目し、その再犯リスクに影響する特徴量を説明する（局所的な説明）もあり、その代表的な手法がSHAP(SHapley Additive exPlanations)と呼ばれる

- 本調査研究ではSHAP値と呼ばれる手法での実現も検証した（イメージは下図）
  - 全てのリスク予測モデルで出力しても確認しきれないと思慮し、パターン2のA群、号種2、機械学習アルゴリズムはglmnetを活用したリスク予測モデルで1名分のみ出力（160名分で約90MB）。
  - 詳細は別添のExcelファイル「chapter2-3特徴量の局所的な説明.xlsx」を参照



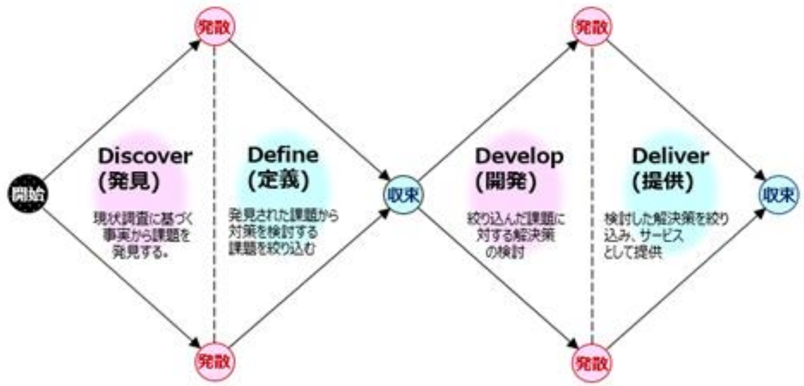
- SHAP等の（局所的）説明を運用する上の課題について
  - 影響度が大きい特徴量がわかっても、それがコントロール可能な要因でなければ、リスク抑制に使えない。
    ⇒　コントロール可能かどうかのフラグ付与など整備が必要
  - 現状では予測モデルに使われた特徴量（つまりシステムで管理されているデータ項目）の名称で表示されるので、そのままでは保護観察官など現場の方には理解できない
    ⇒　各データ項目と現場向けの説明情報を付与する整備が必要

参考：https://qiita.com/kenkenvw/items/dda72d982faf2ec8d2fa

Ⅲ. 試験用画面サンプルの作成とユーザーからの意見聴取

試験用画面サンプルの作成とユーザーからの意見聴取の進め方

- 「試験用画面サンプルの作成とユーザーからの意見聴取」は、2018年に内閣官房 情報通信技術（IT）総合戦略室から示された「サービスデザイン実践ガイドブック（β版）」に則って、サービスデザインの典型的なプロセスである「ダブルダイヤモンド」であらわされる、「発見」「定義」「開発」「提供」のフェーズに分けて進める



（出所）内閣官房 情報通信技術（IT）総合戦略室　サービスデザイン実践ガイドブック（β版）より

- 本調査研究の過程において構築・実装された、再犯リスク予測の初期モデルを活用し、保護観察のプロセスにおいて、どのような価値を、どのような画面で提示するのかを、サービスデザインのプロセスに則って検討し、試験的に作成する

試験用画面サンプルの作成の進め方（初期案）

- サービスデザイン実践ガイドブックに記載の「サービスデザインのプロセス」に則って、以下のような進め方を想定
- デザイン手法については、ユーザーの環境等に合わせて一定のカスタマイズを予定して進める

発見

- 仕様書p4(3)アにより、アセスメントに係る現場業務フローのヒアリングを実施
- ヒアリングでは得られない情報を補足する取り組みを実施
  - 保護観察官・対象者だけでなく、保護司等を含む、ステークホルダーを整理
    # ステークホルダーマップ
  - 情報が不足する場合、ユーザ（保護観察官）のニーズをインタビュー等を通じて、現在の業務を把握
    # デプスインタビュー
- ユーザーが、本業務で試作するシステムにおいて、どのような価値を求めているのか導出し、ユーザー体験の可視化によって、現状の課題を把握
  # KA法・カスタマージャーニーマップ(As-Is)・ペルソナ（想定ユーザー像）

定義

- KA法で作成した価値マップに基づき、AI-CFPに求められるニーズの定義
  - ペルソナの視点で、重要度や優先度を定義
  - 戦略を策定し、シナリオを作成
- 分析結果の提示方法の検討（中間報告）

開発

- プロトタイプの開発
  - 既存データを用いたサンプル（ユーザーにおける入力画面、事案単位の解釈補助を含む出力画面等）
- デジタル庁のデザインシステム（後述、実施におけるIT環境等）を活用し、効率的にプロトタイピングを行う

提供

- 定めた価値の提供を実現し、ユーザーの課題が解決できているか評価を実施
  - 可能であれば、保護観察官に参加いただいて、ユーザビリティテストを実施
    # 思考発話法、行動観察
  - 保護観察官によるユーザビリティテストが不可能な場合、エキスパートレビュー（UXの専門家が実施する評価）
- ここで洗い出された課題、改善のポイントについて、次の開発サイクルにつなげるためのまとめ

※青字は利用した手法

1. 発見～定義フェーズ

発見フェーズの実施事項（4月後半～6月前半）

令和5年度　本調達に係る調査研究　（以下「本調査研究」という）における実施ポイント

（A）既存のデータから適切なアセスメント項目の設計を進める
（B）再犯リスク予測の初期モデルを構築・実装する
（詳細は仕様書 第3の2(1)）

（C）画面上で動作するサンプルを試験的に作成する
（D）運用上の課題を洗い出し、システム実装に向けた検討を進める
（詳細は仕様書 第3の2(2)）

■ 仕様書p4(3)アにより、アセスメントに係る現場業務フローのヒアリングを実施

■ ヒアリングでは得られない情報を補足する取り組みを実施

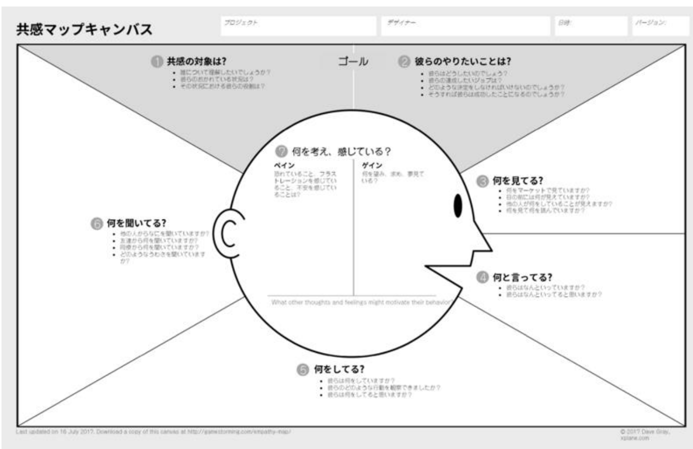
- 保護観察官・対象者だけでなく、保護司等を含む、ステークホルダーを整理
  # ステークホルダーマップ
- 情報が不足する場合、ユーザ（保護観察官）のニーズをインタビュー等を通じて、現在の業務を把握
  # デプスインタビュー

■ ユーザーが、本業務で試作するシステムにおいて、どのような価値を求めているのか導出し、ユーザー体験の可視化によって、現状の課題を把握
KA法・カスタマージャーニーマップ(As-Is)・ペルソナ（想定ユーザー像）

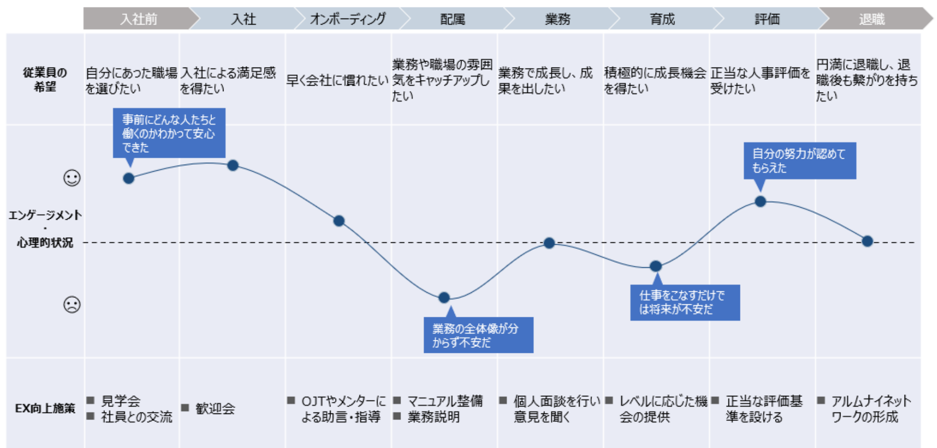
※青字は利用した手法

発見フェーズの実施事項

- 仕様書p4(3)アにより、アセスメントに係る現場業務フローのヒアリングを実施
- ヒアリングでは得られない情報を補足する取り組みを実施
  - 保護観察官・対象者だけでなく、保護司等を含む、ステークホルダーを整理
    # ステークホルダーマップ
  - 情報が不足する場合、ユーザ（保護観察官）のニーズをインタビュー等を通じて、現在の業務を把握
    # デプスインタビュー
- ユーザーが、本業務で試作するシステムにおいて、どのような価値を求めているのか導出し、ユーザー体験の可視化によって、現状の課題を把握
  # KA法・カスタマージャーニーマップ(As-Is)・ペルソナ（想定ユーザー像）



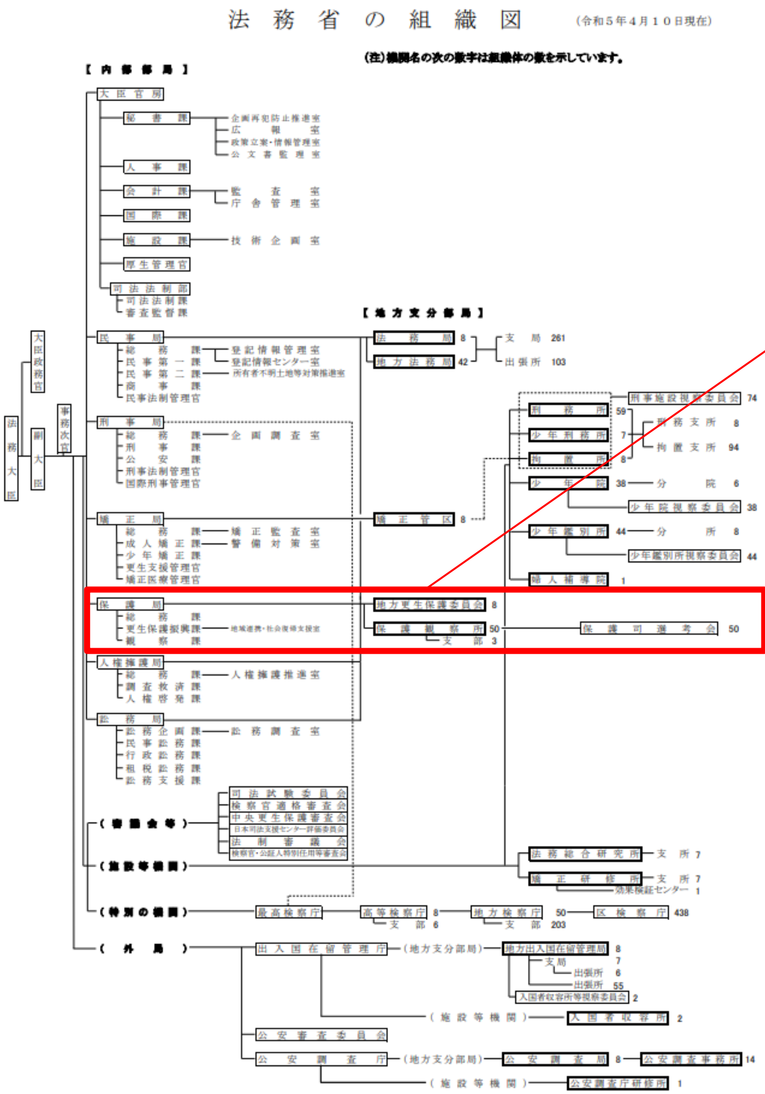
インタビュー⇒共感マップ

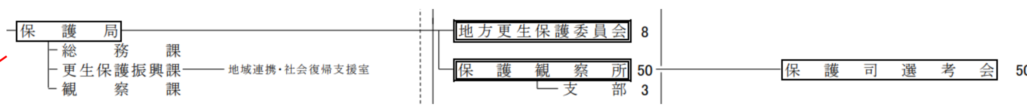


カスタマージャーニーマップ

アウトプット（画面）からの逆算の前に、ユーザーの体験視点での設計のためのリサーチ

ステークホルダーの整理（1）



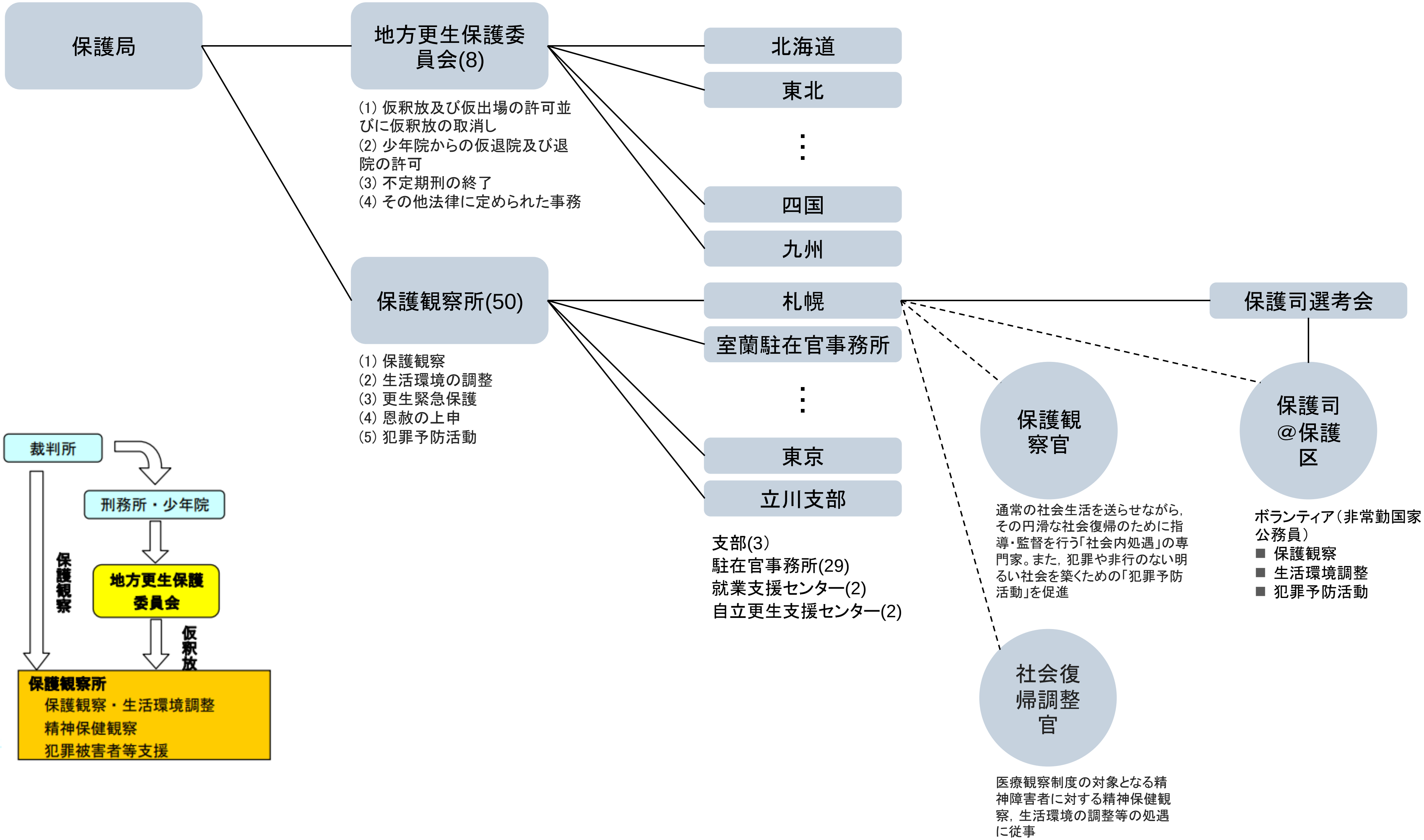


CFPは、

- ■「保護観察所の長」が実施するものとされている。
- ■ 令和5年度から、「地方更生保護委員会」は、その業務を実施するに当たり、CFPによる分析等に有用と考えられる情報を収集したときは、保護観察所に引き継ぐこととされた。

法務省サイトより：
https://www.moj.go.jp/content/001296193.pdf

ステークホルダーの整理（2）



価値の導出

(できごと)	(心の声)	(価値)
少年簿を借りに行く	データ連携できていればなぁ	事件立件にすぐにとりかかれる価値
事件立件をする（生年月日・氏名・ふりがな・国籍・性別が必須）	基本情報なのでミスなく登録しよう	正確に入力できる価値
事件立件をしている最中でエラーが出る	ブラウザーがだめなのか？	どんなブラウザでも利用できる価値
事件立件をする（生年月日・氏名・ふりがな・国籍・性別が必須）	接続に時間がかかりすぎる	事件立件にすぐにとりかかれる価値
事件立件をする（生年月日・氏名・ふりがな・国籍・性別が必須）	保存できてなかったらどうしよう	確実に保存できる価値
面談時、各庁で用意しているフォーマットに記録をする	CFPの充実や実施計画に役立つ記述をしよう	聞き取り項目がCFPや実施計画にそのまま役立つ価値
面談が立て込んだので終了後にまとめて記録を作成する	面談直後に記録ができなかったなぁ	時間をかけずに記録ができる価値
事件調査票の作成	画面がスリープになって記録が消えてしまった	確実に保存できる価値
管理表の作成	作成画面にたどり着くまでのステップが多すぎるなぁ	ログイン後の画面遷移が少ない価値
CFP分析の実施	保護観察期間が始まっているので早く保護司に引き継がなくては	CFP分析が短時間でできる価値
実施計画の作成	CFPが短時間できれば実施計画に時間がかけられるのに	CFP分析が短時間でできる価値
CFP分析の実施	少年簿から情報収集をしないと情報が不足している	少年簿のデータが連携している価値
CFPの作成	作成画面にたどり着くまでのステップが多すぎるなぁ	ログイン後の画面遷移が少ない価値
CFPの作成	過去の事例を参照して書きたいなぁ	類似の事例が短時間で検索できる価値
庁内検討	印刷画面にたどり着くまでのステップが長いので手元で作ろう	ログイン後の画面遷移が少ない価値

フォーカスポイント

- CJM及び一部23年度実施のデジタル技術を活用した更生保護行政最終報告書を参考に、提供価値の導出を実施
  - 23年度のデジタル技術～報告書については詳細を読み込み、課題と提供価値の導出に参照
  - AI-CFPと、その周辺の環境等に切り分け、AI-CFPの提供価値について検討

（できごと）実施計画の作成		（できごと）面談時、各庁で用意しているフォーマットに記録をする		（できごと）実施計画の作成		（できごと）CFP分析の実施	
（心の声）CFPが短時間できれば実施計画に時間がかけられるのに	（価値）CFP分析が短時間でできる価値	（心の声）CFPの充実や実施計画に役立つ記述をしよう	（価値）聞き取り項目がCFPや実施計画にそのまま役立つ価値	（心の声）CFPが短時間できれば実施計画に時間がかけられるのに	（価値）CFP分析が短時間でできる価値	（心の声）保護観察期間が始まっているので早く保護司に引き継がなくては	（価値）CFP分析が短時間でできる価値

- CFP分析が短時間でできる価値
  - CFPや実施計画を作成するために必要な情報を用意できているか、CFP以前に確認できる価値
  - CFP分析の結果を見て、実施計画が短時間で作成できる価値
  - 分析結果が妥当かどうか確認するため、過去の事例と照らし合わせることができる価値（作成者及び上司の承認プロセスにおいて、受容性を高めることで、判断の材料になる）
  - 要因分析をもとに、今後の処遇の参考になるよう、本人が変えられることと、変えられないこと（過去の経験）を区別し、変えられる部分について実施計画に反映できる価値

データ分析と合わせて、画面の構成要素を決め、UIに反映する

AI-CFPに対するニーズを定義

- CFP分析が短時間でできる価値
  - CFPや実施計画を作成するために必要な情報を用意できているか、CFP以前に確認できる価値
  - CFP分析の結果を見て、実施計画が短時間で作成できる価値
  - 分析結果が妥当かどうか確認するため、過去の事例と照らし合わせることができる価値（作成者及び上司の承認プロセスにおいて、受容性を高めることで、判断の材料になる）
  - 開始時要因分析をもとに、今後の処遇の参考になるよう、本人が変えられることと、変えられないこと（過去の経験）を区別し、変えられる部分について実施計画に反映できる価値

利用状況の把握・ヒアリングと、当初の課題設定のギャップ

①現状CFPにより判定している長期再犯リスクについて、保護観察官ごとの判断のばらつきを減少させ、情報の見落とし等を防ぎ、より適切な分析を支援するためのAI技術開発と実装・運用を行い、**保護観察官のアセスメントの質の向上と書類精査等に係る負担軽減を図る**。

②急性再犯リスクについて、短期的な再犯リスクを判断するモデルを確立し、急性再犯リスクの高まりの検知・発報が期待できるAI技術開発と実装・運用を行い、**保護観察官による適時適切な介入の判断を支援する**。

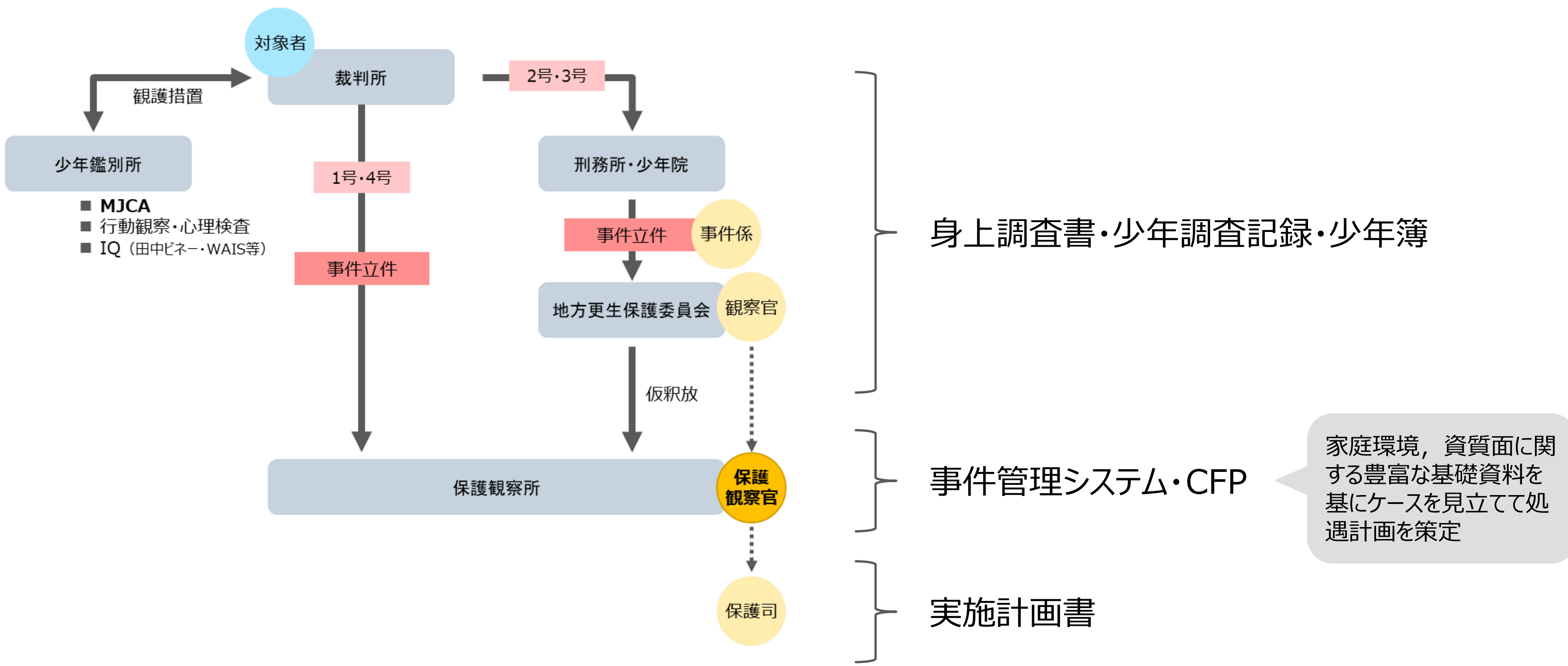
追加リサーチのご提案

統括保護観察官インタビュー

- 概要
  - 参加者
    - 保護局観察課　アセスメントや事件事務等を担当する係の職員
    - 保護観察所　統括保護観察官　2名（A庁、B庁から1名ずつ）
    - MURC　黒田、小原
  - 実施日
    - 11月22日13:30～
  - 受領資料
    - A庁保護観察所調査資料票（交通1・4号用）　様式
    - A庁予備調査票（更生緊急保護）様式
    - B庁保護観察所調査票（1・4号）様式
  - 冒頭PJの概要を説明後、インタビュー・ディスカッションを実施

2. 定義～開発フェーズ

課題①：情報の分散


対象者
裁判所
観護措置
少年鑑別所
■ MJCA
■ 行動観察・心理検査
■ IQ（田中ビネー・WAIS等）
1号・4号
事件立件
2号・3号
刑務所・少年院
事件立件
事件係
地方更生保護委員会
観察官
仮釈放
保護観察所
保護観察官
保護司
身上調査書・少年調査記録・少年簿
事件管理システム・CFP
家庭環境，資質面に関する豊富な基礎資料を基にケースを見立てて処遇計画を策定
実施計画書

AI-CFPの活用の価値の再確認（発見～定義フェーズの振り返り）

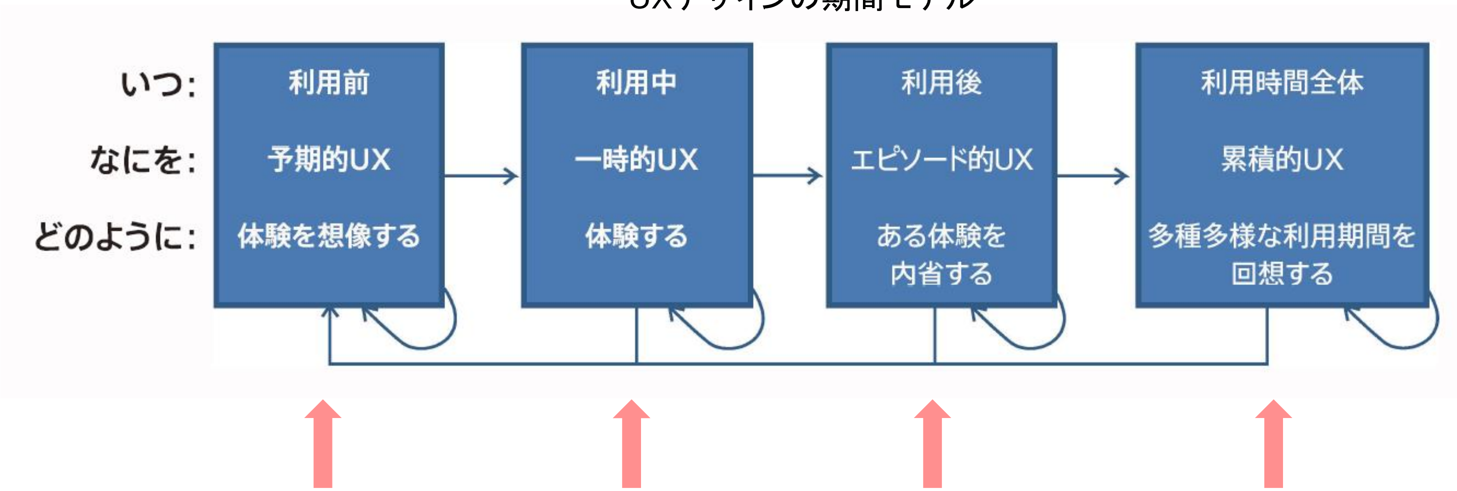
- CFP分析が短時間でできる価値
  - CFPや実施計画を作成するために必要な情報を用意できているか、CFP以前に確認できる価値
  - CFP分析の結果を見て、実施計画が短時間で作成できる価値
  - 分析結果が妥当かどうか確認するため、過去の事例と照らし合わせることができる価値（作成者及び上司の承認プロセスにおいて、受容性を高めることで、判断の材料になる）
  - 開始時要因分析をもとに、今後の処遇の参考になるよう、本人が変えられることと、変えられないこと（過去の経験）を区別し、変えられる部分について実施計画に反映できる価値

3. 提供フェーズ

エキスパートレビュー

- UXの専門家が、インターフェースデザインについて、ユーザー操作の問題点になりやすいポイントについて、ガイドラインに基づいて評価を行う。開発の途中段階で実施することが多く、作りこみの前に実施されることが多い。
- 初期モデルの活用方法を決めたうえで、アプリケーションの開発⇒ユーザビリティと進める。
- UXプロセスの中で、利用中に限定せず広い範囲で評価が可能

UXデザインの期間モデル



- 当日面接でも、短時間で準備ができる
- 物理的な情報のやり取りがない

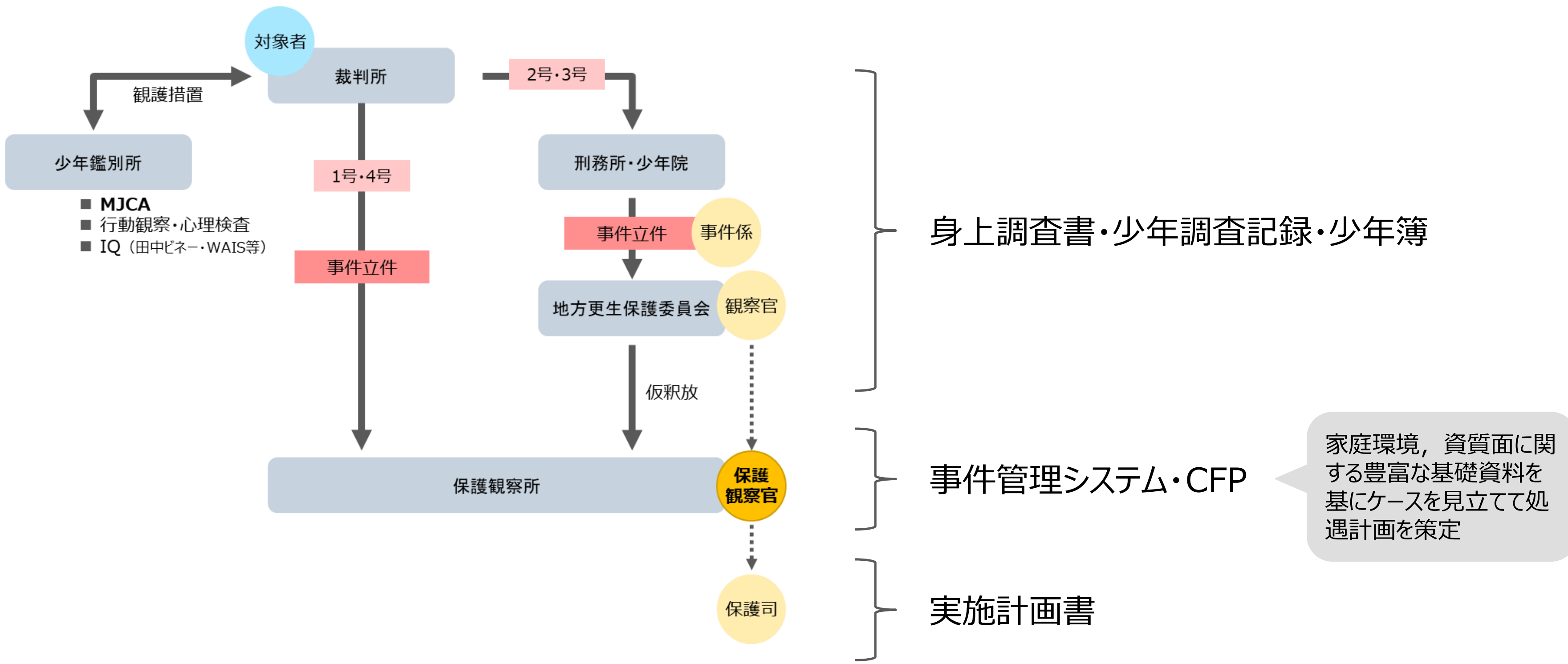
- 対象者が到着した時に、面接の準備ができている
- 対象者に情報を正確に入力してもらえる
- 経験の差によるアセスメントにばらつきがない
- 短時間でCFPを用いたアセスメントができる
- 提出書類が簡単に揃えられる
- 確実に記録できる

- 過去の事例を簡単に参照できたので、適切なアセスメントができた
- 焦らず面接に臨め、しっかり話が聞けた

- 自分の過去の事例にあてはまるものがなく悩んだが他の人の事例が参考になった

ワークフローに関する課題

■ **情報が分散しており、一部デジタル化されているものの、書類の受け渡しが発生している**



保護観察にとどまらず、対象者のフローに合わせたシステムが必要

面談にPCをもちこめないなどの状況があり、データ入力のハードルが高い

ユーザビリティに関する課題

初回の行動観察及びヒアリングのリサーチにより以下の価値を導出

- **CFP分析が短時間でできる価値**
  - CFPや実施計画を作成するために必要な情報が用意できているかをCFP以前に確認できる価値
  - CFP分析の結果を見て、実施計画が短時間で作成できる価値
  - 分析結果が妥当かどうか確認するため、過去の事例と照らし合わせることができる価値
（作成者及び上司の承認プロセスにおいて、受容性を高めることで、判断の材料になる）
  - 開始時要因分析をもとに、今後の処遇の参考になるよう、本人が変えられることと、変えられないこと
（過去の経験）を区別し、変えられる部分について実施計画に反映できる価値

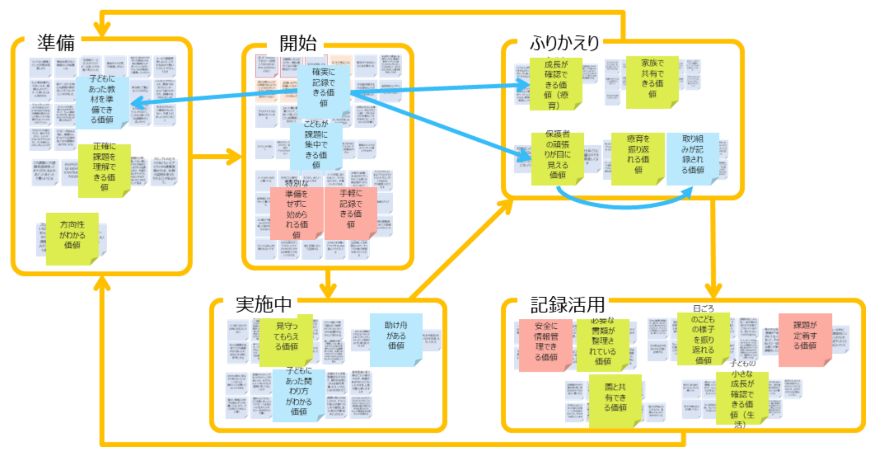
基本的なユーザー体験の設計が未実施と想定される
ユーザー体験の設計を実施して改善を行うことで、大幅な改善が見込まれる

(ご参考)事例

■ 目的
- データ収集を推進するためのアプリケーションの在り方を検討する
- ユーザーの利用状況を把握し、よりデータを収集しやすい仕組みをデザインする

利用状況の把握 ▶ 価値・ニーズの抽出 ▶ アプリケーションの改善 ▶ データの収集・蓄積・活用

デプスインタビュー　KA法・価値マップ



データの保存がうまくいかないケースが頻発⇒
最初に実施した改善は、保存ボタンの改善

4.　まとめ

試験用画面サンプルの作成とユーザーからの意見聴取

本調査研究において、サービスデザインのプロセスであるダブルダイアモンドの方式に則って、以下の事項を実施した。

1. **発見フェーズ**
   - 保護観察官へのヒアリングを通じて、業務フローの理解及びステークホルダーマップを作成
   - 行動観察とインタビューを通じてカスタマージャーニーマップを用いて、現状の整理を実施

2. **定義フェーズ**
   - リスク予測モデルの活用による、保護観察官への提供価値の導出のため、統括保護観察官へのインタビューを実施
   - 予測モデルの活用として、以下の2点にフォーカスを定めた
     - 長期再犯リスクを、パス図と要因分析を同時に実施しながらデータとして蓄積し活用する
     - 対象者ごとの再犯のサイクルに気づき、急性再犯リスクの高まりを検知し、適時適切な介入の判断を支援する

3. **開発フェーズ**
   - 下記のポイントにフォーカスしながら、3度のワイヤーフレーム及び画面サンプルを試作
     - ①CFP分析が短時間でできる価値の提供
     - ②要因分析を時系列に記録し、再犯パターンに気づけるような仕組みを作る
     - ③長期再犯リスク及び、短期的な再犯リスクについて気付きを得る

4. **提供フェーズ**
   - エキスパートレビューを実施し、評価及び今後改善すべきポイントについてレポートを作成
     （最終的な画面の修正案は、次頁以降に記載）

Ⅳ. 来年度以降に向けた活動の整理

本作業項目における検討内容

提案書で提案したタスクを基に、これまでの検討経緯を踏まえ、以下の手順で検討した

1. 業務に実装・運用するための課題の整理

- AI-CFPシステムを業務に実装し、運用する際の工夫や留意事項を「デジタル・ガバメント推進標準ガイドライン」8～9章を基に網羅的に抽出（システム設計・開発は本調査研究の対象外）

2. AI-CFPの運用に固有の課題の整理

- 試験用画面サンプル作成におけるヒアリング、及び、定例打合せ等で伺った、AI-CFPの運用に固有の課題を整理

3. AI-CFPの実現に向けたロードマップ案の策定

- 左記の2つを統合した上で、法務省における今後の活動方針を踏まえてロードマップ案として整理

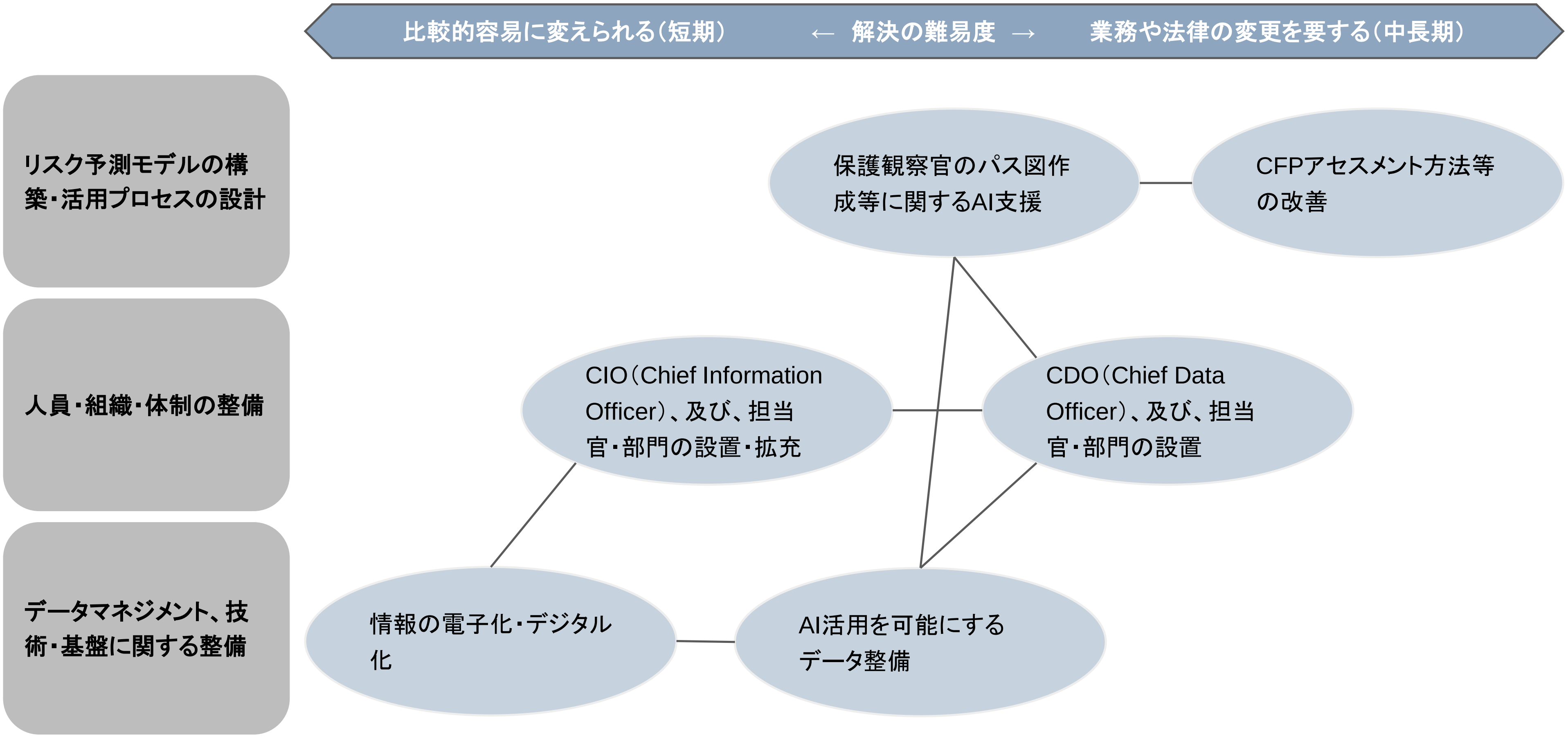
1. 業務に実装・運用するための課題の整理

- 「デジタル・ガバメント推進標準ガイドライン」8～9章を基に網羅的・機械的に、リスク予測モデルを業務に実装・運用する際の工夫や留意事項を抽出した
- 得られた活動を整理（下図）。詳細は別添のExcelファイル「chap4.xlsx」にある

	計画フェーズ	実証フェーズ	運用フェーズ
リスク予測モデルの構築・活用プロセスの設計	■ **必要なタスクやプロセス、そのための役割分担・責任範囲を定義** ■ 構築・活用プロセスにおけるKPIやリスクの抽出	■ 構築・活用プロセスの修正、役割やKPI・リスクの改訂・確定 ■ 実行に向けた準備、内外への周知・広報など	■ KPIデータ収集による業務モニタリング ■ 内部の組織・体制やシステム等の変更に応じたプロセスの見直しなど **運用を通じた定期的な見直し**
人員・組織・体制の整備	■ 上記プロセスの実行に必要となるスキルの抽出、現行の人員・組織・体制とのギャップ分析 ■ 検証のためのリハーサル実施計画の策定	■ **検証のためのリハーサル実行** ■ 修正した構築・活用プロセスも踏まえた人員・組織・体制の案、内部の教育・訓練または外部からの人材調達の計画策定	■ 内部の教育・訓練または外部からの人材調達の実行や文書化 ■ **運用を通じた定期的な見直し**
データマネジメント、技術・基盤に関する整備	■ 上記プロセスと既存データやシステムとの整合性を確認・整備 ■ 関与する関係者が理解し活用できるデータの品質や水準の整備、必要に応じて是正など	■ 活用するデータの収集・加工プロセスを確認 ■ 既存システムへのフィードバック必要に応じてシステム改修、データ加工プロセス修正など	■ データ項目・定義や収集・加工プロセスの改廃に応じた対応など **運用を通じた定期的な見直し**

2. AI-CFPの運用に固有の課題の整理

- これまでヒアリングや定例打合せで伺った、AI-CFPの運用に固有の課題を整理（下図）
- 各ノードの詳細は次頁以降に記載



2. AI-CFPの運用に固有の課題の整理（詳細：1/2）

データマネジメント、技術・基盤に関する整備

情報の電子化・デジタル化

- 現状では保護観察官に情報が紙で来てしまう。また内部手続きにおいても紙で印鑑をもらう業務フローである。このため、**システムが無くても業務ができてしまう**状況にある
- 加えて**システムが使いづらいために使われず**、データはwordで保存するのみといった状況も発生
- とにかくデジタル化を推進することが必要である

AI活用を可能にするデータ整備

- 現状は自由記述項目が多く、かつ、入力者ごとに表現が異なったり、短文・略記や空欄が多かったりする。一部の自由記述項目は作業負荷低減の観点からコード化・定型文化が進められているが、それらコード等には抜け漏れが少なからず見受けられる。このため、そのままでは**AIに活用し難い**
- データ項目や記入ルールの整理、コードの体系化など、**AIの基盤としてのデータ整備に不可欠**
- 本調査研究の仕様書で挙げられていた「急性リスク」という観点についても、そもそも時系列で扱えるデータが少なかった。そのようなデータ項目の構成だけでなく、常に上書きされて最新データしか残らない等の**運用も問題もあった**
- 現状を把握せずに「急性リスク」などの観点を挙げることは**データを管理・把握していない体制**面の問題とも見受けられる。本調査研究におけるデータ提供の大幅な遅延や混乱もその問題の一端である

人員・組織・体制の整備

CIO（Chief Information Officer）、及び、担当官・部門の設置・拡充

- 上記の課題に継続的に対応するには、**複数部門・システムを組織横断的に統括し、ITリソースの最適化を図る責任者、及び、これを実行する担当官・部門の設置**が不可欠である
- 特に保護局と支所、社内外の関連部門の連携に責任を持つ担当官・部門を別途設置して状況改善を図ることが望ましい。これは、保護観察官や保護局職員の個々の頑張りだけでは上記の課題の改善には手が回らないものと推察されるためである

2. AI-CFPの運用に固有の課題の整理（詳細：2/2）

人員・組織・体制の整備（つづき）

CDO（Chief Data Officer）、及び、担当官・部門の設置	■ 個別に開発・運用されるITリソースの横断的な管理だけでなく、**データマネジメントについてもシステム・部門を横断して対応する責任者、及び、これを実行する担当官・部門の設置が不可欠** ■ データが機械で扱える形になっていないことは**政府方針とのズレとしても問題**であり対策が必要 参考：統計表における機械判読可能なデータの表記方法の統一ルールの策定 (soumu.go.jp)

リスク予測モデルの構築・活用プロセスの設計

保護観察官のパス図作成等に関するAI支援	■ 1号観察で〇〇という条件ならこういうパス図である、というパターンをAIが提示できるようになれば、経験の浅い保護観察官が何もないところから書くよりも便利になる、との法務省からの意見があった。これは技術的には提案書で例示したベイジアンネットワークなどの手法で実現できる可能性がある。 ■ しかし、その実現には、上述したデータ面での困難さがある。加えて、本調査研究において、現場担当者へのヒアリングに慎重な法務省の姿勢などから、現場担当者を交えたチューニングが困難であることも示唆されるため、実装プロセスにも課題がある。 ■ パス図は、保護司や保護観察対象者に説明するストーリーとして便利であり、実施計画の作成に必要と現状では考えられている。一方で、保護観察対象者をフラットに見るには、パターン化されたパス図はバイアスの元になるとも思慮される。AI導入にはメリットもあるが、CFP等の**位置づけや目的を再考・整理し**、メリデメなど総合的に考慮する必要がある
CFPアセスメント方法等の改善	■ CFPは大学教授の理論に沿って開発され、これに沿ってデータの整備・蓄積がされている。しかし、ヒアリング等を通じて、実態として保護観察官に使われていない可能性が見受けられた。実際に当初想定よりもCFPによる項目はリスク予測モデルにおいて影響度は大きくなかった ■ 保護観察官を巻き込んで**現場の気づきを取り込み、現場で使えるアセスメント方法とすることで、CFPがより有用になる。加えてCFPへの納得感が醸成されて使われるようになることで、得られるデータの精度が上がり、リスク予測モデルの性能も向上する、AI-CFPが実用的なものになる**と考えられる ■ CFPアセスメント項目の活用については当初より法務省からサブカテゴリ等を必要とするコメントがあった。本調査研究では、データ起点のみでは改善のための知見を得られないことが示唆されており、理論調査や実践を通じた継続的な改善が求められる

3. AI-CFPの実現に向けたロードマップ案の策定

- 令和4年度調査研究報告書5.3節にある「ロードマップ案の概念図」は、ITシステムの開発を目的としたスケジュール案に近い。令和4年度調査研究の時点で見えていた課題からは、単一の「AI-CFPシステム」を構築すれば十分と考えられたためと推察される
- しかし前述の通り、AI-CFPに関連する既存のシステムやデータ管理、更には人員・組織などの課題が絡み合っており、単純にAI-CFPシステムのみ構築すればよい状況ではない

⇒ AI-CFPシステムに限らず、AI-CFPによるアセスメント高度化の実現を最終目標とした活動全体の視点で整理が必要

- これまでに挙げた課題を解決するためのロードマップ案（次頁）
  - AI-CFPシステムとしてのMVP（Minimum Viable Product：最小限の実行可能な製品）を開発し、MVPの運用を通じて関係者とのイメージ共有や動機付けを図りながら、現場のニーズを捉えたり、データ等を整備したりすることで、AI-CFPを継続的に改善していく方針も考えられる
  - しかし本稿では、法務省における今後の周辺活動などの方針を受け、情報のデジタル化やデータ等の整備から着手するロードマップ案として整理した。なおタイムラインの記載は本ロードマップ案では割愛した

- ロードマップ案のPhase1～2を具体化するアクションプランの策定、特にITシステムの設計・開発については令和4年度調査研究報告書における「ロードマップ案の概念図」が、業務面については前述「業務に実装・運用するための課題の整理」が有用である

AI-CFPの実現に向けたロードマップ案

	Phase1 AI-CFPの業務フローやITシステム・データ、組織の基盤整備	Phase2 リスク予測モデルの再構築とAI-CFPの業務実装	Phase3 AIを効果的に活用するための制度改善や各種環境構築
リスク予測モデルの構築・活用プロセスの設計	■ リスク予測モデル関連の**業務フローやKPIの定義**、リスク要因の抽出 ■ 省内外への周知・広報やリハーサルなど実行に向けた準備	■ 整備されたデータによる**リスク予測モデルの再構築** ■ 安定的な運用（新たなデータによるリスク予測モデルの更新を含む）を通じた課題抽出	■ 運用を通じた業務フローやCFPアセスメントの見直し ■ パス図作成など、リスク予測モデルの**他にAIで実現可能なCFP支援策の企画検討～実施**
人員・組織・体制の整備	■ 業務フローの実現に必要な人員・組織・体制の検討 ■ 内部の教育・訓練または外部からの人材調達の**計画策定**	■ 内部の教育・訓練または外部からの人材調達の**計画実行** ■ 特に**リスク予測モデルの開発・運用にかかる人材を優先的に**整備	■ 関連する複数のITシステムや部門を横断的に統括する**CIO的な担当官・部門の設置・拡充** ■ 同じくデータマネジメント面で統括する**CDO的な的な担当官・部門の設置・拡充** ■ CIO・CDO的な担当官・部門の主導によるITシステムやデータ収集・加工プロセス、データマネジメントに関する**組織横断的な改善**
データマネジメントやIT基盤に関する整備	■ **情報の電子化・デジタル化の徹底** ■ 業務フローと整合的なデータ収集・加工プロセスの定義と実装 ■ データ項目辞書、機械判読可能 (*)、データ項目や値など**基礎的な整備**	■ 関係者が理解し活用できるデータの品質や水準など、**リスク予測モデル構築に向けたデータ整備**（項目の可読性や分類、システムやテーブルを跨ぐ重複の整理など）	

（*） 総務省｜報道資料｜統計表における機械判読可能なデータの表記方法の統一ルールの策定 (soumu.go.jp) 等の政府ガイドラインに沿うことを指す。

三菱UFJリサーチ&コンサルティング株式会社

www.murc.jp/