

「こどもに関する各種データの連携による支援実証事業（地方  
公共団体におけるデータ連携の実証に係る調査研究）」  
各採択団体における成果報告書

戸田市

PwC コンサルティング合同会社

2023年3月末

本報告書は、デジタル庁との間で締結された令和4年6月10日付の業務委託契約書に基づき、PwCコンサルティング合同会社が作成したものです。PwCコンサルティング合同会社は、本報告書に関連して、デジタル庁以外の第三者に対して、如何なる義務や責任も負いません。なお、PwCコンサルティング合同会社は、本報告書の日付後に発生した事象について、追加で報告をなし又は本報告書に反映させる責任を負うものではありません。

## 目次

1	実証事業の概要.....	1
1.1	背景・目的.....	1
1.2	実証事業の内容.....	2
1.3	システムの概要.....	8
1.4	スケジュール・実施体制.....	9
1.5	システムの開発費用.....	12
2	支援業務の実施手順（システムを利用した運用フロー）.....	13
3	分析に必要なデータ項目の整理.....	14
3.1	取得・共有・分析するデータ項目.....	14
3.2	データレイアウトの検討.....	16
4	データを扱う主体の整理・役割分担.....	18
5	個人情報の適正な取扱いに係る整理.....	19
5.1	個人情報の取扱いに係る法的整理.....	19
5.2	プライバシー・倫理面での整理や対象となる子どもや家庭への周知.....	22
6	システム企画.....	23
6.1	データ連携方式.....	23
6.2	連携に必要な機能.....	26
6.3	情報へのアクセスコントロールの整理.....	29
6.4	安全管理措置.....	30
7	データの準備.....	31
7.1	データの取得.....	31
7.2	アナログ情報のデジタル化.....	33
7.3	名寄せ.....	33
7.4	データ分析のための加工.....	37
7.5	マスキング.....	38
7.6	外字.....	39
8	システムによる判定基準の設計と評価.....	40
8.1	データ分析と分析結果の評価.....	40
8.2	判定基準の構築と評価.....	44
9	事業効果の評価・分析.....	45
9.1	支援業務の試行結果.....	45
9.2	今後の取組.....	51
10	全国的な展開方策の検討.....	52
11	まとめと今後の課題.....	52

## 1 実証事業の概要

### 1.1 背景・目的

戸田市教育委員会では、これまで、教師の経験と勘と気合い（3K）のみに頼る指導からの脱却、客観的なデータに基づく教育施策の推進を掲げ、教師の指導の可視化・言語化・一般化や、各学級の学力伸長の状況把握、非認知的能力（社会情緒的スキル）の可視化とともに、それらのデータを分析し、支援等につなげられるよう取り組んできた。令和元年には、「子供たちを誰一人取り残さず、一人一人が21世紀を主体的に生き抜く力を身につけるため、教育を科学することを通じ、優れた教師の匠の技の言語化・可視化・定量化や個別最適な学びの実現、EBPM<sup>1</sup>の推進に取り組む」ことを主眼に、戸田市教育政策シンクタンクを設置した。令和3年には、経済学や教育学等の専門家を外部アドバイザーとして委嘱したアドバイザーボード第1回を開催し、着々とデータを活用した教育におけるEBPMを推進する体制を整備してきた。また、戸田市こども健やか部においても、家庭の困難度を示すデータをアウトリーチに活用したり、データを用いた事業の効果検証を行ったりするなど、データを活用した取組を進めてきた。

このように、戸田市では、様々なデータを保有、活用しているが、多くの場合、個々のデータの個別の活用にとどまり、教育委員会内のデータの集約や、健康福祉部やこども健やか部等で有するデータとの効果的な連携はできていなかった。

教育におけるデータの活用について、戸田市では、「教育を科学する」というコンセプトの下、「授業を科学する」「生徒指導を科学する」「学級・学校経営を科学する」という三つの方向性を打ち出しているが、このうち、学校現場では、「生徒指導を科学する」という観点で、例えば不登校等のSOSの早期発見・対応を試行することで積極的な生徒指導を補強していくこと等に対する必要感が高いと捉えている。

---

<sup>1</sup> Evidence Based Policy Making の略。

# 教育を科学する当面の方向性

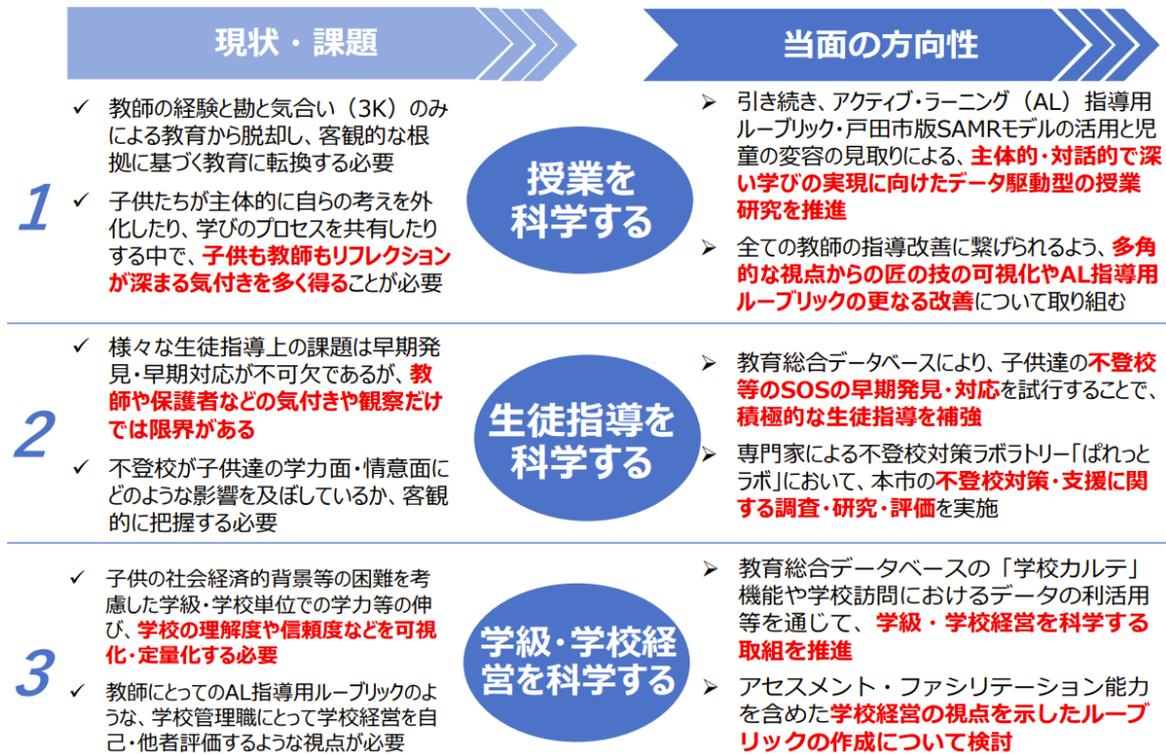


図 1-1 教育を科学する当面の方向性

上記を踏まえ、令和3年度より、教育委員会内におけるデータの整備や部局を横断したデータの連携に向け、どのようなことが可能なのか、また、その実現に向けた課題及びその解決策について検討を進め、本年度、下記を実証コンセプトとして設定し、実証を行うこととなった。

## 実証事業のコンセプト

1. 誰一人取り残されない、子供たち一人一人に応じた支援の実現（子供たちの SOS の早期発見・支援等）
2. EBPM（EIPP<sup>2</sup>）の推進（行政課題特定の精緻化や施策の効果測定等）
3. 新たな知見の創出（匠の技の可視化、学校カルテによる学校現場へのフィードバック等）
4. 関係機関の連携促進（教育委員会と福祉部局等との連携等）

### 1.2 実証事業の内容

今回の事業では、戸田市教育委員会及び首長部局に分散している子供に関わるデータについて、教育分野を軸に教育総合データベースを整備し、併せて、データの標準化やデータフォーマットのオープン化等により、他

<sup>2</sup> Evidence Informed Policy and Practice の略。

自治体においても導入しやすい基盤となることを目指している。実証の主たる目的としては、以下の3点を設定した。

- 不登校等の SOS の早期発見・支援
- 貧困・虐待等の困難を有する子供への支援
- 学校カルテによる現場への継続的改善のためのフィードバック

業務イメージは下図の通りである。基幹系システム、校務系システム、学習系システムのデータを連携させ、教育総合データベースを構築、不登校等のSOSの早期発見や子供への支援等のアクションにつなげていく方針である。

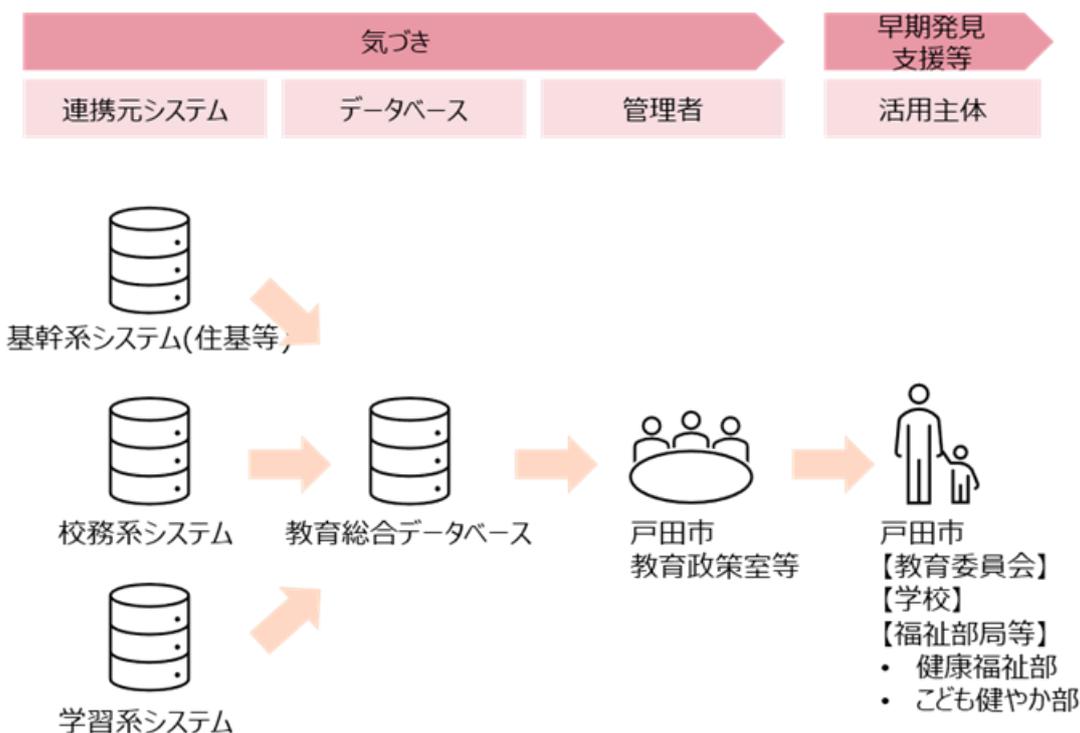


図 1-2 事業概要図

教育総合データベースの具体的な活用イメージとしては、現時点において以下の3つを想定している。

表 1-1 教育総合データベースの活用イメージ

① 子供たちの SOS の早期発見・支援	不登校、いじめに関し、子供たちの SOS が事前に何らかの兆候として現れていないか。それを踏まえ、ニーズに応じた早期支援ができないか。
② 貧困・虐待等の困難を有する子供への支援	上記①のような SOS の兆候が現れた場合に、家庭的な要因に係るデータを市内の関係部局等に共有することにより、貧困・虐待等の困難を有する子供や家庭への支援につなげることができないか。
③ 学校カルテによる現場への継続的改善のためのフィードバック	困難な状況にもかかわらず学力向上等を達成している学校には、共通する特徴があるのではないか。そうした傾向の分析により、継続的改善のためのフィードバックが提供できないか。

これらの活用イメージを踏まえて、教育総合データベースの活用により想定される主なユースケースは下表の

通りである。なお、あくまでも検証の射程として考えられるユースケースについてリストアップを行っているものであり、その全てを早急に検証することは実現可能性に乏しいため、今後、分析結果等を踏まえて関係事業者とも協議を行いつつ、優先順位を定めて随時、一つ一つについて検証の可能性を見定めていくこととしている。

表 1-2 教育総合データベースで想定される主なユースケース

ユースケース	概要
児童生徒ダッシュボード	各児童生徒について、基礎情報（氏名等）や各種調査の結果・回答などが、一覧かつグラフ等の分かりやすい形で表示されるもの また、長期欠席 <sup>3</sup> になっている児童生徒について、その日数の推移や、学校による支援、教育相談等の記述が一覧かつグラフ等の分かりやすい形で表示されるもの
不登校発現リスク判定	ある児童生徒について、学校生活アンケート <sup>4</sup> や埼玉県学力・学習状況調査児童生徒質問紙 <sup>5</sup> 、授業がわかる調査 <sup>6</sup> 等の回答をもとに、その後不登校になるリスクがどの程度高いかを過去データに基づき判定し、高い場合にはアラートとして表示されるもの
不登校深刻度リスク判定	ある児童生徒について、ある月に不登校（長期欠席）になった場合に、当該不登校がどの程度深刻かを過去データに基づき判定し、高い場合にはアラートとして表示されるもの
いじめ深刻度リスク判定	ある児童生徒について、ある時点でいじめとして報告された場合に、当該いじめがどの程度深刻かを過去データに基づき判定し、高い場合にはアラートとして表示されるもの
家庭関係要因判定	ある児童生徒について、学校生活アンケートや埼玉県学力・学習状況調査児童生徒質問紙、授業がわかる調査等の回答をもとに、家庭関係の要因がどの程度強いかを過去データに基づき判定し、高い場合はアラートとして表示されるもの
学習関係要因判定	ある児童生徒について、学校生活アンケートや埼玉県学力・学習状況調査児童生徒質問紙、授業がわかる調査等の回答をもとに、学習関係の要因がどの程度強いかを過去データに基づき判定し、高い場合はアラートとして表示されるもの
教師関係要因判定	ある児童生徒について、学校生活アンケートや埼玉県学力・学習状況調査児童生徒質問紙、授業がわかる調査等の回答をもとに、教師関係の要因がどの程度強いかを過去データに基づき判定し、高い場合はアラートとして表示されるもの
友人関係要因判定	ある児童生徒について、学校生活アンケートや埼玉県学力・学習状況調査児童生徒質問紙、授業がわかる調査等の回答をもとに、友人関係の要因がどの程度強いかを過去データ

<sup>3</sup> 当該年度間に連続又は断続して 30 日以上欠席した場合を指す。以下同様。

<sup>4</sup> 戸田市内の中学校 1 年生、2 年生の全生徒に対して行っている学校生活に関するアンケート。年 1 回、5 月に実施。「7.2 アナログ情報のデジタル化」に詳細を記載。

<sup>5</sup> 埼玉県教育委員会が小学校第 4 学年から中学校第 3 学年までを対象に、年に 1 回実施している調査。子供たちの、「学力の伸び」を測ることができる「教科に関する調査」、学習に対する意欲や学習方法、さらに家庭での生活習慣等に関する「質問紙調査」のほかに、各教育委員会や各学校の取組についての「質問紙調査」も実施。

URL : <https://www.pref.saitama.lg.jp/f2214/gakutyou/20150605.html>

<sup>6</sup> 「授業がわかる調査」は、平成 16 年度から戸田市が独自に行っている、授業に関する児童生徒向けアンケート調査。調査対象は小学校第 4 学年から中学校第 3 学年までで、年に 2 回実施。

	ータに基づき判定し、高い場合はアラートとして表示されるもの
学校基礎情報 可視化	各学校（さらに学年・学級）単位で、児童生徒数や教職員数、県学力・学習状況調査結果、生活保護・就学援助受給児童生徒割合、特別支援教育対象児童生徒割合、日本語指導を必要とする児童生徒割合及びそれらの推移が一覧かつグラフ等の分かりやすい形で表示されるもの
学校カルテ （各種調査結果比較）	各学校（さらに学年・学級）単位で、各種調査やアンケートの結果が、同一集団（学年等）の経年比較やある学年等の過去データとの比較を含め、グラフ等の分かりやすい形で表示されるもの
学校カルテ （伸び分析）	調査実施の前年度の埼玉県学力・学習状況調査結果（又はAi-GROW <sup>7</sup> ）、及び調査実施年度の生活保護・就学援助受給児童生徒割合、特別支援教育対象児童生徒割合、日本語指導を必要とする児童生徒割合を変数として制御 <sup>8</sup> した上で、調査実施年度の県学力・学習状況調査結果（又はAi-GROW）における予測値を算出して、前年度からの伸び等が過去データに基づく予測を上回る・下回るかが学校（さらに学年・学級）単位で、アラートで表示されるもの
学校カルテ （相関分析）	上記の伸び分析を行った上で、過去データに基づき、当該伸びと相関関係が特に高いと考えられるデータを分析し、有意なものがアラートで表示されるもの

上記のユースケースの中から、本実証事業期間においては、「不登校発現リスク判定」の実現に向けた不登校予測モデル構築及び不登校要因分析、「児童生徒ダッシュボード」の実現に向けた検討（協力校での検証を通じた簡易型児童生徒ダッシュボードの構築）、「学校カルテ」の実現に向けた試行（学校訪問時の各種調査分析結果共有とフィードバックの取得）を行った。

今回、リスク判定のうち、特に不登校に焦点を当てて検討を進めたのは、戸田市としての重点施策の対象であったこと、学校現場のニーズが高かった点が挙げられる。前者については、戸田市では、令和4年度の新規の事業として、総合的な不登校施策である「戸田型オルタナティブ・プラン<sup>9</sup>」を開始するなど、重点的に不登校施策に取り組んでいる。後者については、国でも、令和3年度の調査<sup>10</sup>において、小・中学校における長期欠席者のうち不登校児童生徒数は244,940人（前年度196,127人）と9年連続で増加しており、過去最多となるなど、不登校が重要課題となっている中、戸田市でも、令和3年度には同242人（前年度同196人）と不登校のこどもの数は増加傾向にあり、学校現場でも不登校に対する施策のニーズが高い状況である。

不登校発現リスク判定に向けた不登校予測モデル構築については、令和3年度の全小中学校（小学校12

<sup>7</sup> Institution for a Global Society 株式会社が提供する、児童・生徒の資質・能力と各種教育活動の教育効果を定量化する教育機関向け評価ツール。

URL : <https://www.aigrow.jp/>

<sup>8</sup> ここで挙げられている各変数が、県学力・学習状況調査結果（又はAi-GROW）の結果に影響を与えている可能性があるため、統計学的にそれらの影響を取り除く処理を行っている。

<sup>9</sup> URL : <https://www.city.toda.saitama.jp/koho-toda/220901/pdf/P.005.pdf>

<sup>10</sup> 文部科学省「令和3年度児童生徒の問題行動・不登校等生徒指導上の諸課題に関する調査結果」

URL : [https://www.mext.go.jp/content/20221021-mxt\\_jidou02-100002753\\_1.pdf](https://www.mext.go.jp/content/20221021-mxt_jidou02-100002753_1.pdf)

校、中学校6校)のデータを用いて不登校の予測と要因分析を行ったが、プッシュ型支援への活用という実用に耐え得るためには、分析精度を更に高めていく必要があるとの結論となったため、本事業期間中には、リスク判定機能の実装までは至っていない(詳細については8章を参照)。しかし、前述の通り不登校に対する施策のニーズが高い状況であることから、今後のプッシュ型支援の示唆を得るため、戸田市内でデータ利活用を先進的に実践している小学校1校(以降、「協力校」)に協力をお願いし、不登校傾向にある子供についてプッシュ型支援の有効性の検証を行った。具体的には、

- ① 県の学力調査において前年度から学力レベルが不変<sup>11</sup>又は低下している
- ② 戸田市「授業がわかる調査」で回答の肯定度合いが減少している
- ③ 学校独自の「幸せアンケート<sup>12</sup>」で回答の肯定度合いが減少している

以上の3つの基準のいずれかに該当する、長期欠席になっていない児童を抽出しつつ、最終的には協力校との協議を踏まえ、見守り対象の子どもを決定した。対象児童について、当初は市教委がデータの推移を見ようとしたが、そもそも調査未実施の児童も多く、抽出対象から外れていたため、市教委が抽出した児童に限らず、取得できたデータが極端な児童や、不登校傾向がみられる児童など、学校生活での実態を見ながら学校が総合的に判断した。検証期間中に見守りをする中で不登校傾向が見られるかなど、新たな気付きを得るための取組を行った。上記の3つの基準は、学校現場へのヒアリング及び身近で活用可能なデータであるという観点から採用した。

---

<sup>11</sup> 前年度からの学力レベルの比較は、今年度(現在)と前年度同時点(過去)の学力との比較を指す。前年度と今年度で同じ難易度の問題しか解けていなければ(さらに難しい問題を解けるようになっていなければ)不変である。

<sup>12</sup> 今回の協力校独自で実施している、学校生活における Well-being に関連したアンケート。自己肯定感や友人・教師との関係性に関連する質問を小学校第1学年から第6学年までが回答。年に2回実施。

## 令和4年度実証事業におけるプッシュ型の支援について

今年度、デジタル庁実証事業として取り組んでいる「教育総合データベース」の現時点（令和3年度分のデータを基にした暫定的な分析結果）における分析結果や、国内の不登校に係る文献調査においては、以下のことが分かっている。

・最も不登校に強い影響を与える因子は**前月までの欠席数**である。

・**①学業要因・②人間関係・③生活リズムの乱れなどが相互作用で影響を与えている可能性**がある。

令和4年度においては、以下のような形でプッシュ型の支援を試行し、検証を行う。

<b>期間</b>	令和5年2月6日～2月17日
<b>対象</b>	市内でデータ利活用を先進的に実践している小学校1校（以下「協力校」という。）
<b>基準</b>	令和4年度において以下の基準に該当する、長期欠席になっていない児童を抽出 ①県学調において前年度から学力レベルが不変又は低下している ②戸田市「授業がわかる調査」第1回（6月）→第2回（12月）で回答の肯定度合いが減少している ③学校独自の「幸せアンケート」第1回（6月）→第2回（1月）で回答の肯定度合いが減少している ※その他、学校が要支援と判断するに当たり有用な情報も必要に応じて参照
<b>手法</b>	(1) 上記基準のデータも参照しつつ、 <u>学校が要支援として抽出した児童を上記期間において見守り</u> （担任による観察や個別の声掛けなど） (2) その結果、不登校傾向が見られるかなど、 <u>新たに気付いたことなどを記入</u> ※上記①～③の要因に対応するものとして学校独自に保有しているデータがあればそれも参照の上、記入 ※抽出児童について休みが増えていたり人間関係の悩み・生活リズムの乱れなどSOSが出ているか、その他要注意な児童や既に不登校傾向となっている子について以前のデータからSOSが裏付けられていたか、といった観点から確認 (3) 教育政策室に状況を報告 → その後、学校からのヒアリングを実施

13

図 1-3 協力校における取組の概要

また、学校カルテの試行については、埼玉県学力・学習状況調査の児童生徒質問紙より抽出した項目、及び授業がわかる調査などのデータを集計し、学校訪問時に、各校の傾向を管理職に共有し、フィードバックを得た。

# 令和4年度における学校カルテの試行について

埼玉県学力・学習状況調査（県学調）の児童生徒質問紙より抽出した以下の項目、及び授業がわかる調査などのデータを集計し、学校訪問時に、各校の傾向を管理職に共有。

## 埼玉県学調の児童生徒質問紙

- ① 難しいことでも失敗をおそれないで挑戦していますか。
- ② 学校の先生たちは自分のよいところを認めてくれましたか。
- ③ 授業で課題を解決するときに、みんなでいろいろな考えを発表すること（がよくありましたか）。
- ④ 授業の始めに、今日どんな学習をするのかをつかんでから学習に取り組んだこと（がよくありましたか）。
- ⑤ 授業の始めには気が付かなかった疑問が、授業の終わりに、頭に浮かんできたこと（がよくありましたか）。

## 授業がわかる調査

- ① 授業がわかりますか。
- ② 授業が楽しいですか。
- ③ （探究心に関する質問）
- ④ （社会貢献意欲に関する質問）
- ⑤ （協働意識に関する質問）

※ほか、一部の学校で実施しているQ-Uアンケート（学級満足度等に関する調査）も活用

### ○教育委員会から学校へのメッセージ

- ・多角的な視点から、データを捉えてもらいたい
- ・子供目線で、取組を振り返ってもらいたい

### ○学校経営をデータから捉える（学校カルテ）

- ・年度間、学年間の傾向の変化
- ・学力だけでなく、学校の雰囲気や授業の質（学校の理解度・信頼度）

### ○継続的な授業改善のためのシステムづくり



17

図 1-4 学校カルテの試行

## 1.3 システムの概要

システムの全体構成としては下図のように教育総合データベースを構築した。これまで独自で管理されていた基幹系、校務系、インターネット系・学習系のシステムを教育総合データベースに連携することでデータの利活用及びその先のプッシュ型支援を目指した。

## 【システム全体構成図】

子どもに関する情報をハイブリッドに連携する「教育総合データベース」を構築。  
埼玉県学力・学習状況調査、その他学力や非認知的（社会情緒的）能力等に係る各種調査、  
出欠状況、保育園・幼稚園の在園状況等をデータベース化

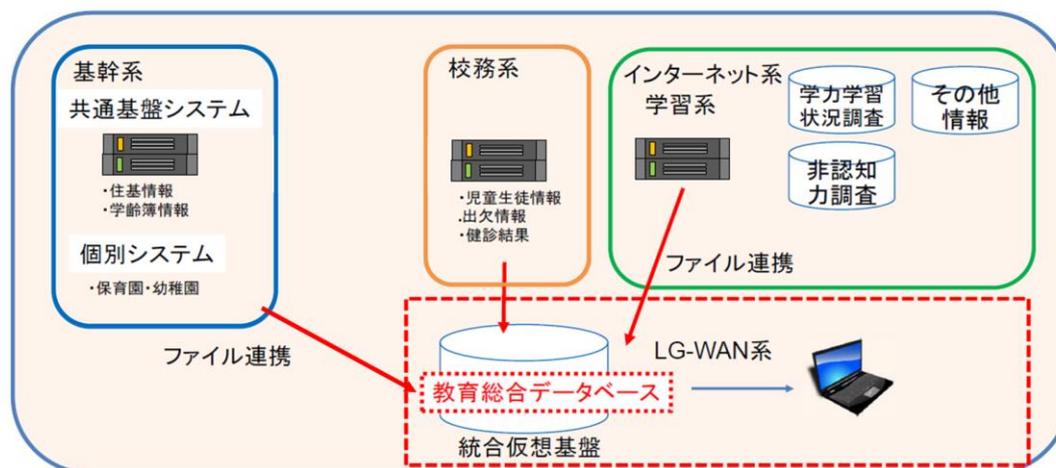


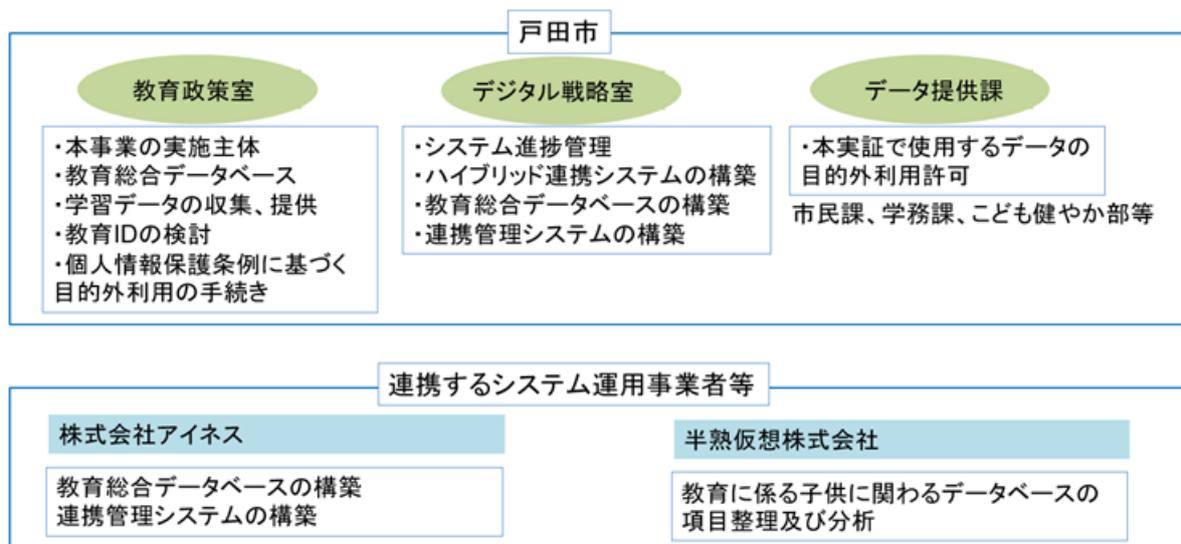
図 1-5 システムの概要図

### 1.4 スケジュール・実施体制

本実証事業の実施期間は令和4年6月から令和5年3月末迄である。スケジュールは大きく以下3つのフェーズに分類できる。①実施体制の整理や法的整理期間、②システム開発期間、③効果検証、それぞれについて記載する。

#### 1) 実施体制の整理や法的整理期間

本実証事業では戸田市教育委員会事務局教育政策室が中心となり、体制の整理をした。戸田市では、子ども健やか部において、貧困や虐待への対応、保育園・幼稚園等を、健康福祉部において乳幼児健診を、教育委員会において小学校・中学校等を、それぞれ所管しているが、本実証事業の中で、データ連携を円滑に進めるための組織改編等は行っていない。体制の整理をした結果、本実証事業でデータ連携する部署、関係機関等、実証事業に参加する関係者の体制、役割等は下図のようになった。



**図 1-6 実施体制**

本実証事業では様々な情報を用いて分析を行い、支援に結び付けていくことから、個人情報の取り扱いについても慎重に検討を進めた。データの保有主体やアクセスコントロール・個人情報の取扱いの整理については、戸田市個人情報保護条例に基づき、令和4年5月に開催された戸田市情報公開・個人情報保護運営審議会への目的外利用・外部提供申請に係る諮問をはじめ、適切に対応するとともに、制度面・運用面での課題の整理を行った。例えば、こども健やか部とのデータ連携においては、実データの共有に限らず、実データの分析結果から得られる子供のSOSの有無の共有等、複数の方策が考えられ、現在の制度において、どのような方策が適切なのか、課題の整理を行った。また、この他、データ利活用に係る基本的な方針や、データガバナンス体制（総括管理主体、保有・管理主体、分析主体、活用主体それぞれの役割分担）や、安全管理措置の実施等について包括的にまとめた「教育データの利活用に関するガイドライン」（以下「ガイドライン」という。）を、アドバイザーボードで複数回議論した上で、令和4年12月に決定・公表した<sup>13</sup>。

## 2) システム開発期間

本実証事業を推進する上で要となる新たなシステム開発には大きく2つのステップで取り組んだ。1つは既存のシステムのデータを教育総合データベースに連携するシステムの開発であり、もう1つは連携した項目を用いて、システム上でリスク判定を実施するための機能の開発である。

事業概要でも説明している通り、教育委員会及び市長部局に分散している子供に関わるデータについて、教育分野を軸にした教育総合データベースを整備することを主眼に置いており、このデータベースの開発を株式会社アイネス（以下「アイネス」という。）と連携して行った。仕様検討、設計を経て、開発を進め、令和5年3月初旬に教育総合データベースの運用が開始された。システム構築においては、契約締結作業の遅れに伴い、実際の作業開始時期が遅れた。当初計画では令和4年6月に仕様検討を開始する予定だったが、結果的に

<sup>13</sup> 戸田市 教育データの利活用に関するガイドライン

URL : [https://www.city.toda.saitama.jp/uploaded/life/126128\\_265228\\_misc.pdf](https://www.city.toda.saitama.jp/uploaded/life/126128_265228_misc.pdf)

は、同年9月末にキックオフ、10月末に仕様確定となり、それ以降の開発スケジュールが後ろ倒しとなった。

システム構築で当初想定よりも工数がかかったのは、各種システムからデータを教育総合データベースに連携するためのデータ加工に係る作業である。具体的には各種システムから取り出したExcelデータや学校及び事業者から取得した各種調査結果の形式を揃え、集約してID付与する作業となる。その後、システムの仕様上、加工したデータをさらに取込用CSVファイルへ変換する必要があったため、これを戸田市側が人力で作業した。そのデータを用いてアイネス側が教育総合データベースへの取り込み作業を実施した。取込時には重複データや表示形式不整合等の対処に必要な加工作業に起因するエラー、及び取込用CSVファイルへの変換作業に起因するエラーが発生した（第7章参照）。

また、教育総合データベース上へのリスク判定機能実装を目指しており、半熟仮想株式会社（以下「半熟仮想」という。）がデータベースのデータ項目を用いて分析を行った。このリスク判定機能開発のための分析については、後述する第8章で詳しく説明しているが、教育現場での運用に耐えるようなリスク予測モデルを構築するために分析を続けている。難易度の高い不登校のモデル構築作業に取り組み、今回は1,000以上のモデルを構築して網羅的に検証することで最適なモデルを模索した。

現状、必要な情報は抽出できるものの、判定機能の搭載、各種調査の結果・回答が例えばグラフ等の分かりやすい形式で表示するためのユーザーインターフェース（以下「UI」という。）に係る機能の搭載などが、実証事業期間では間に合わないなど、実運用に至るまでの課題が残っている状況である。

### 3) 効果検証

上述2)の通り、教育総合データベースの構築が3月までかかり、リスク判定機能も継続して検討したため、これらの効果検証を実証期間内には行っていない。しかし、本実証事業の期間内では協力校での検証を行った。そこから得た示唆をもとに、今後教育総合データベースの活用方法を検討していく予定でいる。



表 1-3 システムの開発費用

No	項目	費用概算
1	情報連携システム構築費用	3,740 万円
	(1 のうち) ソフトウェア開発費	3,408 万円
	(1 のうち) ハードウェア費用	67 万円
	(1 のうち) ミドルウェア費用	265 万円
2	判定ロジック開発費用	909 万円

## 2 支援業務の実施手順（システムを利用した運用フロー）

本年度の事業では、教育総合データベース自体を直接支援業務に活用するまでには至っていないが、以下、不登校を例にとり、現時点での業務手順の想定案を記載する。

不登校における教育総合データベースの各種データの活用イメージは下図の通り。

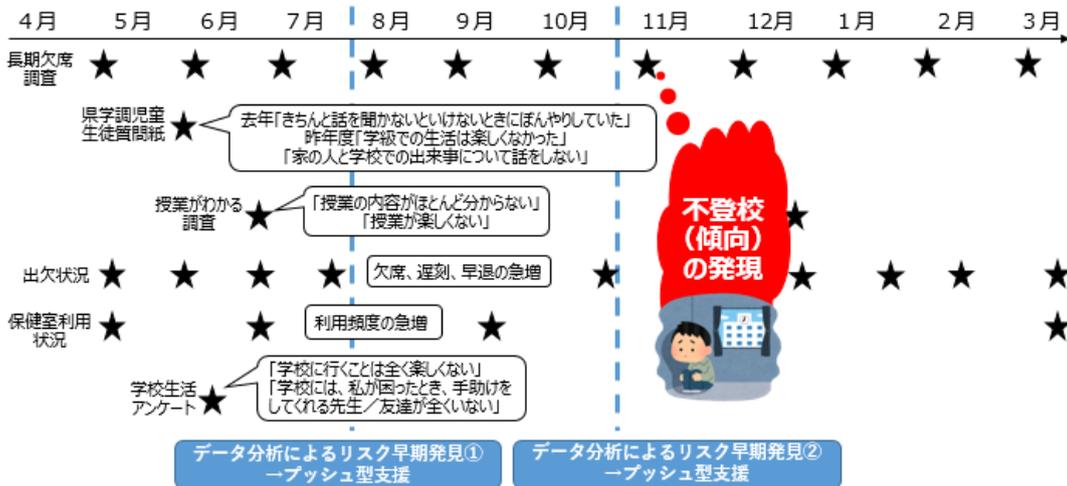
### 具体的な活用イメージ（モデルプラン）の例（不登校）

#### <不登校のSOSの早期発見・支援>

不登校（傾向を含む。）の課題が顕在化する前から、子供たちは困難を感じ、SOSを発出している可能性があるのではないか。そうしたことをデータ連携・分析により早期発見することで、未然防止のための学校等での個別のケア・支援につなげることが出来るのではないか。

#### <イメージ（項目は例）>

★：それぞれのデータを取得しているおおよその頻度を指す



(※) データ項目等はあくまでも例であり、これに限られるものではない。また、データ分析の時期についてもあくまでもイメージを示したもの。

図 2-1 教育総合データベースの具体的な活用イメージ（不登校）

上記のように、各種データを取得したのち、データ分析を行い、プッシュ型支援につなげていく想定である。その場合、今後は以下のような業務フローになることを想定している。まず、教育総合データベースに今後実装予定である不登校発現・不登校深刻度リスク判定機能によりアラートが表示され、そのアラートを教育政策室の不登校支援担当職員が確認する。確認結果を、アラートがあがったこどもの所属校の学校長にアラートの旨を伝え

る。なお、今後のリスク判定機能のUI次第では、学校長が直接アラートを確認できるフローも可能性として考えている。その後、アラートを受けた学校の学校長等が教育総合データベースのダッシュボード機能を活用し、該当する子どもに関する情報を確認する。その情報をもとに、校内でケース会議を開催し、支援方針を決定する。なお、検討の結果見守りの方針になる子供もいれば、声掛けや面談等の直接的な支援も考えらえる。

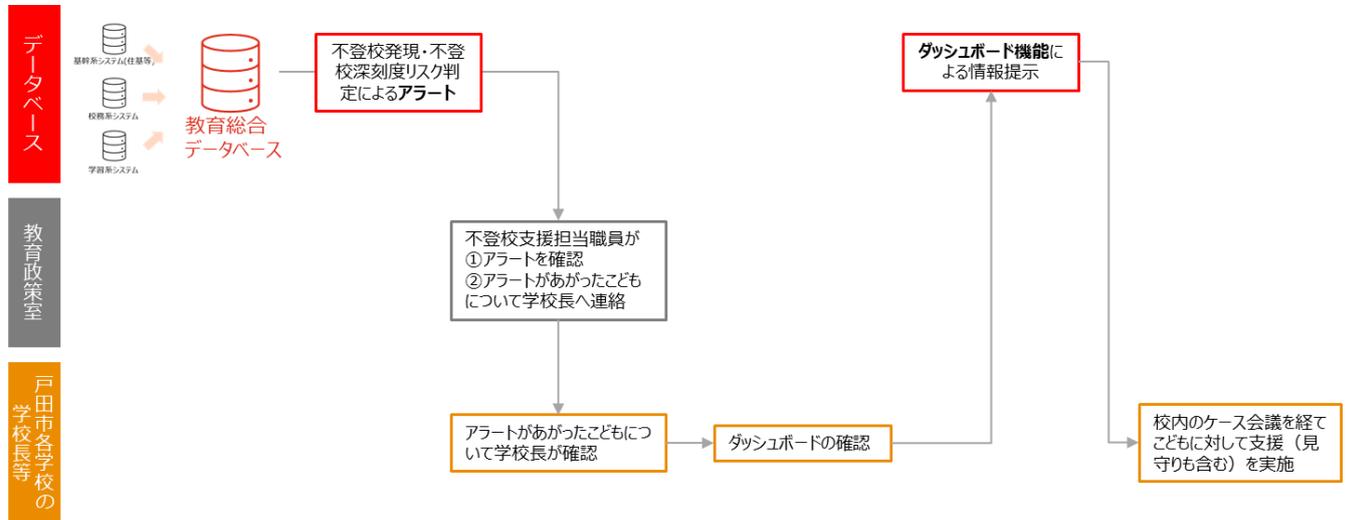


図 2-2 将来的な業務プロセスのイメージ

### 3 分析に必要なデータ項目の整理

#### 3.1 取得・共有・分析するデータ項目

今回の事業で連携した項目一覧は下図の通りである。今回の実証事業では、ユースケースが複数あることから、幅広くデータを取得しつつ、個人情報保護法の観点で、利用目的ごとに、使うデータ、使わないデータを区別している。

基礎情報	生徒指導	学力等	その他
氏名・生年月日・性別等	長期欠席調査	県学力・学習状況調査	出欠・遅刻・早退の状況
在籍学校名・クラス・出席番号	いじめ等に関する記録	県学力・学習状況調査質問紙	授業がわかる調査
埼玉県学力・学習状況調査管理番号	教育相談の利用の有無	Reading Skills Test	学校生活アンケート調査
就学前段階	健康	非認知的能力調査 (AiGROW)	
保育・幼稚園在園時の状況	乳幼児健診結果		
	学校健診結果等		

図 3-1 連携データ項目一覧

データの保存期間としては、データ項目ごとに、当該データを保有する保有・管理主体において、法令に保存

期間の規定があるもの<sup>15</sup>はそれに従い、それ以外のものについては、戸田市文書管理規程（表3-1）に基づき、当該文書の主務課の長（文書管理者）として適正に管理を行う。

データベースに搭載される情報のうち、こうした保有・管理主体から提供を受けたものについては、特段、それが改変されるものではないため、データベースの取組が、その保存期間に特段の影響を与えるものではない。他方、データベースにおける分析結果等については、データベースに固有のものとして生成される新たな情報ということができ、したがって、その保存期間について、総括管理主体である教育政策室として検討を行う必要がある。

この点、ガイドラインにおいて、データベースにおける分析結果等は、それを参考として、活用主体が当該児童生徒に対するアセスメントや支援を行うという意味で、重要な意義を有する情報であり、文書管理規程上1年保存とされている「照会、回答、報告等に関するもので軽易な文書等」又は3年保存とされている「照会、回答、報告等に関するもので5年保存を要しない文書等」「出張票等」「復命書」といった文書と同列に扱うことは適切ではない。他方、上記の指導要録や健康診断票の保存期間が5年であること、及び文書管理規程上10年保存とされている文書が「告示及び広告に関する文書等」「市議会に関する重要な文書等」といった文書であることも鑑みれば、データベースにおける分析結果等の保存期間は5年を基本としつつ、対象となる文書ごとに、個別具体的に検討を行うことが適当と記載されている。

---

<sup>15</sup> 例えば、学校教育法施行規則（昭和二十二年文部省令第十一号）第二十八条では、以下の通り指導要録や学籍に関する保存期間を規定している。

（以下抜粋）

第二十八条 学校において備えなければならない表簿は、概ね次の通りとする。

一 学校に関係のある法令

二 学則、日課表、教科用図書配当表、学校医執務記録簿、学校歯科医執務記録簿、学校薬剤師執務記録簿及び学校日誌

三 職員の名簿、履歴書、出勤簿並びに担任学級、担任の教科又は科目及び時間表

四 指導要録、その写し及び抄本並びに出席簿及び健康診断に関する表簿

五 入学者の選抜及び成績考査に関する表簿

六 資産原簿、出納簿及び経費の予算決算についての帳簿並びに図書機械器具、標本、模型等の教具の目録

七 往復文書処理簿

② 前項の表簿（第二十四条第二項の抄本又は写しを除く。）は、別に定めるもののほか、五年間保存しなければならない。ただし、指導要録及びその写しのうち入学、卒業等の学籍に関する記録については、その保存期間は、二十年間とする。

その他、学校保健安全法施行規則（昭和三十三年文部省令第十八号）では、以下の通り健康診断表の保存期間を規定している。

（以下抜粋）

第八条 学校においては、法第十三条第一項の健康診断を行ったときは、児童生徒等の健康診断票を作成しなければならない。

2 校長は、児童又は生徒が進学した場合においては、その作成に係る当該児童又は生徒の健康診断票を進学先の校長に送付しなければならない。

3 校長は、児童生徒等が転学した場合においては、その作成に係る当該児童生徒等の健康診断票を転学先の校長、保育所の長又は認定こども園の長に送付しなければならない。

4 児童生徒等の健康診断票は、五年間保存しなければならない。ただし、第二項の規定により送付を受けた児童又は生徒の健康診断票は、当該健康診断票に係る児童又は生徒が進学前の学校を卒業した日から五年間とする。

表 3-1 戸田市行政文書管理規則で定められている文書の保存期間

保存年限	保存基準
永年保存	<ol style="list-style-type: none"> <li>1 市の沿革に関する文書等</li> <li>2 市の廃置分合及び行政区画に関する文書等</li> <li>3 条例、規則、訓令等の制定及び改廃に関する文書等</li> <li>4 市議会の会議録、議決書等に関する文書等</li> <li>5 審査請求に関する重要な文書等</li> <li>6 訴訟及び和解に関する文書等</li> <li>7 叙位叙勲及び褒章に関する文書等(市長公室所管のもの)</li> <li>8 職員の任免、賞罰等に関する文書等(人事課所管のもの)</li> <li>9 年金、退職手当、公務災害補償等の決定に関する重要な文書等</li> <li>10 調査報告書、統計書、年報等で重要な文書等</li> <li>11 予算及び決算に関する重要な文書等(財政課所管のもの)</li> <li>12 公有財産の取得及び処分並びにこれらに関する登記関係の文書等</li> <li>13 重要な事業に関する文書等</li> <li>14 工事の設計書等で重要な文書等</li> <li>15 原簿、台帳等で重要な文書等</li> <li>16 許認可等に関するものでその効力を有する期間が 10 年を超える文書等</li> <li>17 その他永年保存する必要があると認められる文書等</li> </ol>
10年保存	<ol style="list-style-type: none"> <li>1 告示及び公告に関する文書等</li> <li>2 市議会に関する重要な文書等(議会事務局所管のもの)</li> <li>3 審査請求に関するもので永年保存を要しない文書等</li> <li>4 諮問、答申等に関する重要な文書等</li> <li>5 表彰に関する重要な文書等</li> <li>6 重要な事業に関するもので永年保存を要しない文書等</li> <li>7 工事の設計書等で永年保存及び 5 年保存を要しない文書等</li> <li>8 契約書、許可証等で重要な文書等</li> <li>9 許認可等に関するものでその効力を有する期間が 5 年を超え、10 年以下の文書等</li> <li>10 その他 10 年保存する必要があると認められる文書等</li> </ol>
5年保存	<ol style="list-style-type: none"> <li>1 照会、回答、報告等に関する重要な文書等</li> <li>2 諮問、答申等に関するもので 10 年保存を要しない文書等</li> <li>3 表彰に関するもので 10 年保存を要しない文書等</li> <li>4 非常勤職員の任免、臨時職員の雇用等に関する文書等</li> <li>5 工事の設計書等で永年保存及び 10 年保存を要しない文書等</li> <li>6 契約書、許可書等で 10 年保存を要しない文書等</li> <li>7 監査に関する文書等</li> <li>8 金銭の出納に関する文書等</li> <li>9 許認可等に関するものでその効力を有する期間が 3 年を超え、5 年以下の文書等</li> <li>10 その他 5 年保存する必要があると認められる文書等</li> </ol>

### 3.2 データレイアウトの検討

今回教育総合データベースに連携されるデータのデータレイアウトについては、既存システムにおける標準化対応の継続適用の観点と、データ連携作業の効率化の2つの観点で検討した。標準化対応の継続適用について、

下表3-2上の情報区分として「学齢簿情報」及び「基幹システム連携情報」にあたるデータは、基幹系システムである共通基盤システムから連携されており、この共有基盤システムは元々地域情報プラットフォーム標準仕様書<sup>16</sup>及び特定個人情報データ標準レイアウト<sup>17</sup>に準拠した形となっているため、今回教育総合データベースに連携される際もこれらのデータレイアウトに準拠した形で連携されている。

それ以外のデータについては、本実証事業の期間内に、実際に必要なデータを抽出できる状態にすることを最優先に、作業効率やコストの面で適切なレイアウトを検討した。例えば、各種Excel形式のデータを手作業でCSV形式に変換した上で教育総合データベースに取り込んだが（詳細は6.1データ連携方式を参照）、その際、変換作業が効率的に行えるよう考慮したデータレイアウトとすることを優先した。例えば、校務情報は、株式会社EDUCOMの校務支援システムである「EDUCOMマネージャー-C4th<sup>18</sup>」からデータを出力したが、校務で用いるための帳票形式の出力にしか対応しておらず、出力されたExcel形式の帳票を手作業でCSVに変換する必要があったため、その作業負担を少なくすることを考慮したデータレイアウトとした。なお、EDUCOMマネージャー-C4thは「地域情報プラットフォーム準拠製品（教育情報アプリケーションユニット）」として登録されているが、データ出力のインターフェースは地域情報プラットフォームに対応していない。対応させるためにはシステムの改修が必要となるが、本実証事業では必要な費用と期間を勘案し、システムの改修は行わず、手作業によるデータの準備、連携を選択した。

---

<sup>16</sup> 市区町村が利用している業務システム間のデータ連携を実現する標準仕様である地域情報プラットフォーム標準仕様書の一仕様。特に教育情報の標準仕様を定義するもの。

<sup>17</sup> 行政手続における「特定の個人を識別するための番号の利用等に関する法律」において定義された「情報提供ネットワークシステム」を介して交換される「特定個人情報」のデータレイアウトを定義する規格。

<sup>18</sup> URL <https://sweb.educom.co.jp/swas/index.php?frame=C4th>

表 3-2 情報区分別の取得データテーブル

情報区分	テーブル名
学齢簿情報	学齢簿基本情報
基幹システム連携情報	保育園入所情報
	幼稚園入所情報
	乳幼児健診情報（3か月・1歳・1歳8か月・3歳6か月・5歳）
	予防接種情報（定義のみ）
校務情報	校務基本情報
	出欠状況（小学校、中学校）
	歯科保健状況
	教育相談一覧
	保健室利用状況
	Ai GROW受検進捗
	学校生活アンケート
	長期欠席調査
	いじめに関する記録
	授業がわかる調査（小学校、中学校）
	RST結果情報
学力学習情報	県学調学力データ
	県学調質問紙
共通テーブル	氏名索引
	全国市町村情報

#### 4 データを扱う主体の整理・役割分担

今回の事業においてはデジタル庁の有識者会議において策定された「こどもに関する各種データの連携に係る留意点（実証事業ガイドライン）」に準拠したデータガバナンス体制を構築するため、下表の体制を構築し、ガイドラインにおいてそれぞれの役割を整理した。

表 4-1 協力校における取組で使用したデータとその保存期間

組織	役割
教育委員会教育政策室	<p>&lt;総括管理主体メイン、保有・管理主体、分析主体、活用主体&gt;</p> <ul style="list-style-type: none"> <li>データの収集や提供、組み合わせ、分析主体との連携等事業全般の総括・管理</li> <li>データ分析を踏まえ、不登校等やその傾向が発現する前にリスクを早期発見し、未然防止のための学校等での個別のケア・支援につなげる</li> <li>また、学校カルテによる学校現場への継続的なフィードバックの実施</li> </ul>
企画財政部デジタル戦略室	<p>&lt;総括管理主体へのサポート&gt;</p> <ul style="list-style-type: none"> <li>教育総合データベースの構築・更改、連携管理システムの構築・更改</li> </ul>

こども健やか部 健康福祉部 教育委員会学務 課	<保有・管理主体、(こども健やか部のみ) 活用主体> ・本実証事業で使用するデータ項目の提供 ・(こども健やか部) 分析結果を踏まえた真に支援を必要とすることもやその家庭等への支援
戸田市立学校の 学校長	<活用主体> ・分析結果を踏まえた真に支援を必要とすることもたちへの支援 ・学校経営・指導改善に向けた学校カルテの活用
半熟仮想	<分析主体> ・データ項目の整理及び分析
アイネス	<総括管理主体へのサポート> ・デジタル戦略室と連携し教育総合データベースの構築、連携管理システムの構築

## 5 個人情報の適正な取扱いに係る整理

### 5.1 個人情報の取扱いに係る法的整理

今回の事業を推進するにあたっては、教育政策室外の部署が保有する個人情報をデータベース構築のために利用することに関し、戸田市個人情報保護条例第8条に基づき、令和4年5月に戸田市情報公開・個人情報保護運営審議会に諮問し、承認を得ている。

- ① 教育総合データベースの構築が新たに個人情報取扱事務を始める場合に該当すること（市個人情報条例第8条）
- ② 担当部署である教育政策室以外の部署が保有する個人情報を取得する場合に該当すること（市個人情報条例第9条）
- ③ データベースの構築に係る事務を外部に委託等する場合に該当すること（市個人情報条例第13条）

### <目的外利用・外部提供に係る個人情報保護審査会への諮問>

教育政策室外の部署が保有する個人情報をDB構築のために利用することに関連し、以下について市個人情報保護条例に基づき、戸田市情報公開・個人情報保護運営審議会への諮問し、承認を得た。

- ①新たに個人情報取扱事務を始める場合（市条例第8条）
- ②担当部署以外の部署が保有する個人情報を取得する場合（同第9条）
- ③当該事務を外部に委託等する場合（同第13条）

【主な取得データ一覧】

	基礎情報	学校生活	学力等	生徒指導
教育委員会 保有データ (教育政策 室保有)	クラス	出欠・遅刻・早退	県学調結果・同調査質問紙	長期欠席調査
	県学調管理番号	学校生活アンケート	授業がわかる調査	いじめ等の記録
		Q-Uアンケート等	Reading Skills Test	教育相談利用有無
			非認知的能力調査（AiGROW）	SC・SSW相談

	基礎情報	健康		就学前段階	健康
教育委員会 保有データ (目的外利用)	氏名・生年月日・性別等	学校定期健診	市長部局 保有データ (外部提供)	保育幼稚園在園時の状況	乳幼児健診
	学校名・学年	保健室利用状況		保育要録	
	宛名コード				

図 5-1. 個人情報保護措置について<sup>19</sup>

#### 【参考】戸田市個人情報保護条例（関係部分抜粋）

（個人情報取扱事務の登録）

第8条 実施機関は、個人情報取扱事務を新たに開始しようとするときは、次に掲げる事項を市長に届け出てその登録を受けなければならない。

- (1)個人情報取扱事務の名称及び目的
  - (2)個人情報取扱事務の所掌する組織の名称
  - (3)個人情報取扱事務の管理責任者
  - (4)個人情報取扱事務の対象となる個人情報に関する事項
  - (5)個人情報取扱事務の対象となる個人情報に要配慮個人情報が含まれるときは、その旨
- (略)
- 3 市長は、(略)届出が提出された場合は、当該届出を審議会に報告しなければならない。

（目的外利用等の制限）

第9条 実施機関は、個人情報取扱事務の目的外のために保有個人情報（略）を利用（以下、「目的外利用」という。）し、又は実施機関以外の者にこれを提供（「外部提供」という。）してはならない。

2 実施機関は、前項の規定にかかわらず、次の各号のいずれかに該当する場合においては、目的外利用又は外部提供（略）をすることができる。

- (1)法令等に定めがあるとき。
  - (2)本人の同意があるとき
  - (3)個人の生命、身体又は財産の安全を守るため、緊急かつやむを得ないと認められるとき。
  - (4)実施機関が、審議会の意見を聴いた上で、公益上特に必要があり、かつ本人の権利利益を侵害するおそれがないと認めるとき。
- (略)

（外部委託の保護措置）

第13条 実施機関は、正当な理由に基づき、個人情報取扱事務を実施機関以外の者に委託する場合は、あらかじめ、審議会の意見を聴いた上で、個人情報を保護するため必要な措置を講じなければならない。

(略)

図 5-2 戸田市個人情報保護条例（関係部分抜粋）<sup>20</sup>

<sup>19</sup> 戸田市，教育政策シンクタンクアドバイザーボード第2回資料「教育総合データベース（デジタル庁実証事業）の検討状況について」

URL: [https://www.city.toda.saitama.jp/uploaded/life/123270\\_258499\\_misc.pdf](https://www.city.toda.saitama.jp/uploaded/life/123270_258499_misc.pdf)

<sup>20</sup> 同上

また、同条例第6条第2項第4号及び第9条第2項第4号に基づき、戸田市情報公開・個人情報保護運営審議会の意見を聞いた上で、公益上特に必要であり、本人の権利利益を侵害するおそれがないと認められたため、取得する際には本人同意や本人への通知は不要と整理することとした。

令和5年度以降については、デジタル社会形成整備法による改正後の個人情報保護法において、地方公共団体の個人情報保護についても、共通的なルールとして国の個人情報保護法が適用されることとなることから、この改正法に基づき、必要となる個人情報の保護措置を実施していくこととする。

## 5.2 プライバシー・倫理面での整理や対象となる子どもや家庭への周知

データ連携にかかる不安感や懸念を払しょくするうえで、検討の内容やプロセスについて幅広く公開し、市民や世論と対話しながら政策を進めていくべきとの考えに基づき、今回の事業推進に当たっては、様々な場で機会を捉えて説明を行ってきた。令和3年度は非公開で開催したアドバイザリーボードについては、令和4年7月から公開で開催し、全国各地から中央省庁、地方自治体、教育関係者、大学・研究機関、民間企業など幅広い傍聴者が参加し、ガイドラインの策定や、本実証事業の進捗状況について課題も含めて公開し、助言をいただいている状況である。

表 5-1 戸田市における情報公開に関する取組<sup>21</sup>

日時	会議等の名称	URL
令和4年5月19日	令和4年第5回戸田市教育委員会定例会	<a href="https://www.city.toda.saitama.jp/soshiki/371/kyo-somukaigi-r04.html">https://www.city.toda.saitama.jp/soshiki/371/kyo-somukaigi-r04.html</a>
令和4年7月20日	戸田市教育政策シンクタンクアドバイザリーボード（第2回）	<a href="https://www.city.toda.saitama.jp/soshiki/373/thinktank-advisoryboard.html">https://www.city.toda.saitama.jp/soshiki/373/thinktank-advisoryboard.html</a>
令和4年7月27日	戸田市教育委員会公式 note を開設し、アドバイザリーボード（第2回）の議題について解説	<a href="https://note.com/toda_boe/n/nf903a94acb9d">https://note.com/toda_boe/n/nf903a94acb9d</a>
令和4年8月1日	戸田市教育委員会公式 note において、データ利活用の意義やデータベースの構築の内容について解説	<a href="https://note.com/toda_boe/n/n72f89ca3d1e6">https://note.com/toda_boe/n/n72f89ca3d1e6</a>
令和4年9月16日	令和4年度第1回戸田市総合教育会議	<a href="https://www.city.toda.saitama.jp/soshiki/111/hisyo-sougoukyouikukaigi202209.html">https://www.city.toda.saitama.jp/soshiki/111/hisyo-sougoukyouikukaigi202209.html</a>
令和4年11月14日	戸田市教育政策シンクタンクアドバイザリーボード（第3回）	<a href="https://www.city.toda.saitama.jp/soshiki/373/kyo-seisaku-thinktank-advisoryboard3.html">https://www.city.toda.saitama.jp/soshiki/373/kyo-seisaku-thinktank-advisoryboard3.html</a>
令和4年11月25日	戸田市教育委員会公式 note において、アドバイザリーボード（第3回）の議題について解説	<a href="https://note.com/toda_boe/n/n6a46edc8f8ec">https://note.com/toda_boe/n/n6a46edc8f8ec</a>
令和4年12月15日	令和4年第12回戸田市教育委	<a href="https://www.city.toda.saitama">https://www.city.toda.saitama</a>

<sup>21</sup> 戸田市，教育データの利活用に関するガイドライン，令和4年12月

URL: [https://www.city.toda.saitama.jp/uploaded/life/126128\\_265228\\_misc.pdf](https://www.city.toda.saitama.jp/uploaded/life/126128_265228_misc.pdf)

	員会定例会	<a href="http://www.city.toda.saitama.jp/soshiki/371/kyo-somukaigi-r04.html">.jp/soshiki/371/kyo-somukaigi-r04.html</a>
令和 5 年 2 月 28 日	戸田市教育委員会公式 note において、ガイドライン策定に係る議論経過を解説	<a href="https://note.com/toda_boe/n/n1d01e978dbe8">https://note.com/toda_boe/n/n1d01e978dbe8</a>
令和 5 年 3 月 1 日	戸田市教育政策シンクタンクアドバイザリーボード（第 4 回）	<a href="https://www.city.toda.saitama.jp/soshiki/373/kyo-seisaku-thinktank-advisoryboard4.html">https://www.city.toda.saitama.jp/soshiki/373/kyo-seisaku-thinktank-advisoryboard4.html</a>

今後とも、戸田市の会議や国の会議も含め、様々な機会を捉えて、データベースの進捗状況やその成果・課題等について積極的に情報発信を行っていく方針である。

## 6 システム企画

### 6.1 データ連携方式

本実証事業におけるシステム概要図は下図の通りである。

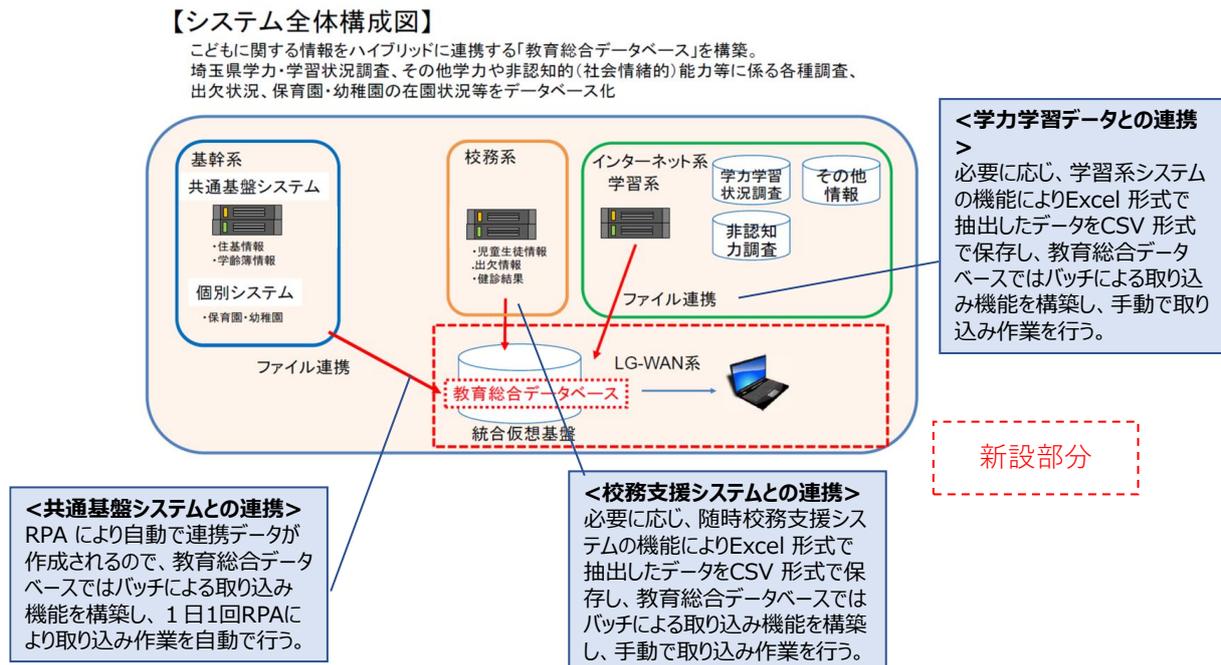


図 6-1 システムの概要図

今回開発した教育総合データベースは、LG-WANセグメントに設置され、自治体内の各システムと連携する。連携元の各システムが設置されているネットワーク上の場所及び外部との接続は下表に整理した通りである。

表 6-1 連携元・連携先のシステムのネットワーク上の場所及び外部との接続

端末を設置している場所	戸田市教育政策室内
-------------	-----------

		ネットワーク上の場所	外部との接続
連携元システム	共通基盤システム	個人番号利用事務系	閉域ネットワーク
	校務支援システム	校務系	閉域ネットワーク
	学習系システム	インターネット接続系	オープンネットワーク
連携先システム	教育総合データベース	LGWAN 接続系	閉域ネットワーク

データの連携方法は、①学齢簿のデータと②学齢簿以外のデータで連携方法が異なる。

学齢簿のデータは、共通基盤システムから共通基盤連携サーバを介し、RPA<sup>22</sup>で自動的に教育総合データベースサーバの所定のフォルダに格納される方式である。一連の連携はFTPS<sup>23</sup>による特定通信とRPAによる自動化で実現している。具体的には以下のプロセスである。

- ① 共通基盤連携サーバで学齢簿連携データが作成される。
- ② RPAで特定通信によりT-NET<sup>24</sup>連携サーバに格納される。
- ③ RPAで特定ユーザによりファイルがコピーされる。
- ④ 教育総合データベースサーバの所定のフォルダに格納される。

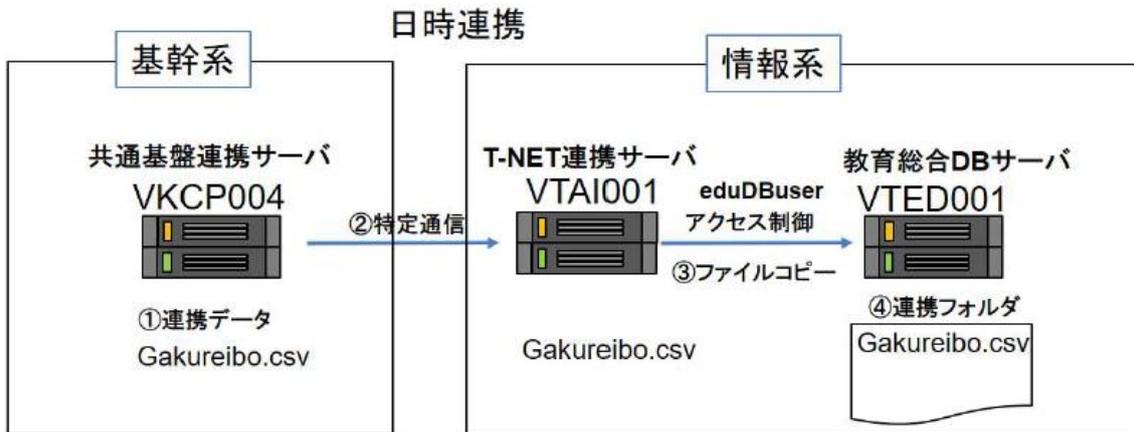


図 6-2 学齢簿のデータ連携

学齢簿以外のデータは、教育政策室のファイルサーバから連携サーバを介し、教育総合データベースサーバの所定のフォルダに格納される。こちらもRPAにより自動化されている。

<sup>22</sup> Robotic Process Automation の略。PC 上で人間が行う作業を、処理手順を登録することによって自動化するソフトウェア技術。

<sup>23</sup> File Transfer Protocol over SSL/TLS の略。通信を暗号化してセキュアな接続でファイル転送を行う仕組み。

<sup>24</sup> 戸田市の庁内向けネットワーク環境。個人番号関係事務系、LG-WAN 接続系のネットワークともいわれるもの。

- ① 教育政策室のファイルサーバに連携データが格納される。
- ② RPAでT-NET連携サーバに格納される。
- ③ RPAで特定ユーザによりファイルがコピーされる。
- ④ 教育総合データベースサーバの所定のフォルダに格納される。
- ⑤ 日時処理結果が教育政策室のファイルサーバに格納される。

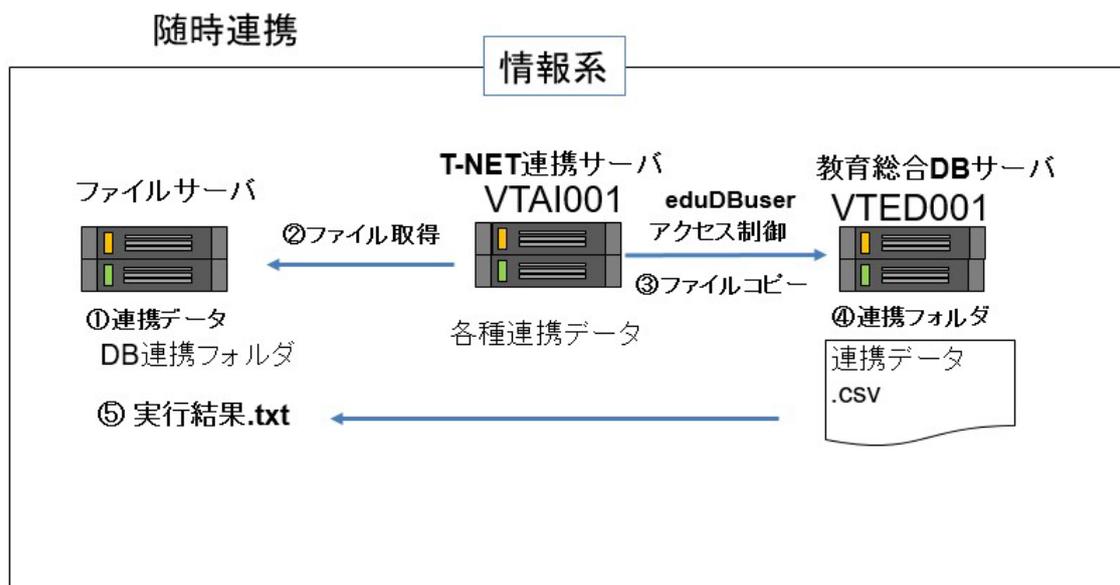


図 6-3 学齢簿以外のデータ連携

以上のプロセスで教育総合データベースサーバに格納されたデータは、日時のバッチ処理によって、教育総合データベースに取り込まれる。

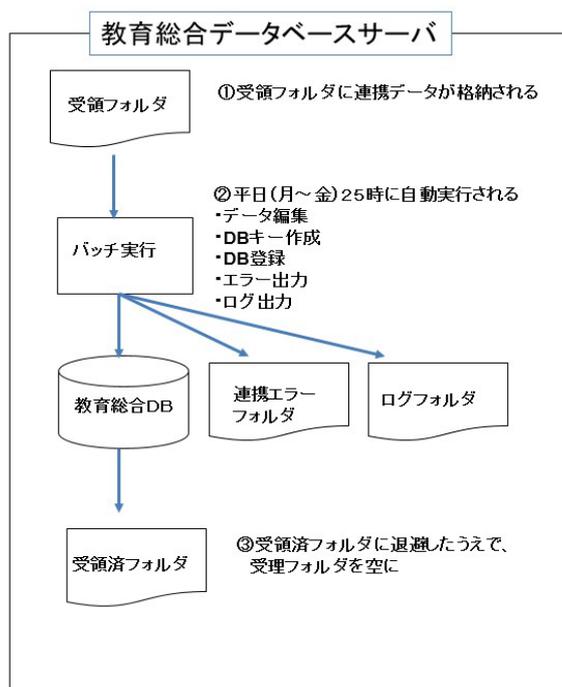


図 6-4 日次のバッチ処理

運用フローとしては、まず、教育政策室担当が連携データを格納し、RPAにより連携サーバを介して受領フォルダに格納する。翌日、実行結果をログデータから確認し、もしログに取り込みの異常がみられた場合には、保守事業者に連絡、適宜修正を図っていく流れである。

それぞれの連携元システムについては今回改修をしておらず、教育総合データベースについてはデータ管理の機能はWebRingsのパッケージを利用したものの、それぞれのシステムからデータ連携をする機能については新規開発した機能である。またデータ分析については、半熟仮想へ業務委託することで実施した。

## 6.2 連携に必要な機能

教育総合データベースは、学齢簿基本情報、住民基本台帳の4情報（氏名、住所、性別、生年月日）、住民登録外の4情報（氏名、住所、性別、生年月日）、児童生徒情報、学力学習情報を、IDキーを紐づけることで人に紐づいたデータ取得とデータ分析を行うデータベースである。

表 6-2 教育総合データベースが保有するテーブル

情報区分	テーブル名
学齢簿情報	学齢簿基本情報
基幹システム連携情報	保育園入所情報
	幼稚園入所情報
	乳幼児健診情報（3か月・1歳・1歳8か月・3歳6か月・5歳）
	予防接種情報（定義のみ）
校務情報	校務基本情報
	出欠状況（小学校、中学校）
	歯科保健状況
	教育相談一覧
	保健室利用状況
	Ai GROW受検進捗
	学校生活アンケート
	長期欠席調査
	いじめに関する記録
	授業がわかる調査（小学校、中学校）
	RST結果情報
学力学習情報	県学調学力データ
	県学調質問紙
共通テーブル	氏名索引
	全国市町村情報

表 6-3 教育総合データベースが保有するテーブル

テーブル名	内容
学校	学校コード、学校名
季節	季節コード、季節名
県学調調査名	調査コード、調査名
就学校区分	就学校コード、就学校
住所コード	住所コード、町名
性別	性別コード、性別
操作員	操作員コード、操作員
入学区分	区分、入学区分
退学区分	区分、退学区分
猶予区分	区分、猶予区分

教育総合データベースの機能については下表の通りである。このうち、パッケージにはないカスタム開発となった機能はNo.14-16のバッチ処理機能であり、この機能によって連携元システムからデータを取り込んでいる。

表 6-4 機能一覧

No.	カテゴリ	要件
1	認証	ユーザ ID、パスワードでログインできること。
2		パスワードの変更ができること。
3		ログアウトできること。
4	児童生徒検索、 登録・修正	児童生徒の最新情報を在籍校名、学年、組などの各種検索条件を指定して検索できること。
5		住民基本台帳氏名（学齢簿情報）のほか、学校で使用している氏名（校務基本情報）から検索ができること。
6		連携 ID の登録・訂正ができること。
7		各種情報の訂正ができること。
8		個人に対してメモ入力ができること。
9		児童生徒名簿一覧の学校ごとの取り出しができること。 帳票名：学校カルテ
10	異動の管理	クラス編成に関する管理ができること。
11		転校に関する管理ができること。
12		進級・卒業の管理ができること。
13		履歴情報の管理ができること。
14	バッチ処理	バッチ処理でデータの登録ができること。 帳票名：登録者一覧・登録エラーリスト
15		バッチ処理で氏名索引の作成ができること。
16		バッチ処理で連携 ID の作成ができること。 帳票名：連携 ID エラーリスト
17	操作ログ	操作ログが個人単位で取得できること。
18		操作ログが個人単位で検索できること。
19	アクセス権	アクセス権限が個人単位で設定できること。
20		アクセス権限が項目単位でできること。
21	コード辞書	学校、住所、季節、県学調調査名等の修正ができること。
22	EUC 機能	データベースに格納されたデータの抽出ができること。
23		連携 ID に紐づく関連したデータの抽出ができること。
24		検索条件入力時の指定で、AND,OR 条件の設定、データの並べ替え、集計ができること。
25	検索	児童生徒や学校等を検索／選択することで、当該児童生徒や学校等に係るデータが表示されること。
26	リンク	分析結果や表示されているアラートをクリックすると、それに関連する元データが DB 内で表示されること。
27	校務支援システムとの連携	校務支援システムで更新（新規登録・追加・修正・削除）されたデータが DB にも自動連携されること。

### 6.3 情報へのアクセスコントロールの整理

アクセスコントロール機能は、ID、パスワードにより個人単位で参照権限を制御する方式で、相談情報等の機微な情報を取り扱う場合は、手のひら静脈による生体認証等の2要素認証を行う。



図 6-5 ログインのイメージ

教育総合データベースの権限設定は下記の通り。ガイドラインに基づき、「付与する権限は必要最小限」との考え方の下、職種や所属、業務分掌に応じてアクセスコントロール（権限管理）を設定している。アクセス権の変更権限等の管理者機能を有するのは、事業全体総括職員とデジタル戦略室職員のみである。

表 6-5 権限設定の想定

**【権限管理マッピング】**

●:権限あり、▲:一部のみ、×:権限なし

役職・業務等	管理者機能	個人検索	調査結果参照	調査結果編集
教育政策室長	×	●	●	×
教育政策室担当課長	×	●	●	×
教育センター所長	×	●	●	×
事業全体総括職員	●	●	●	●
データ整備・分析職員	×	●	●	●
庁内他部局調整職員	×	▲ 他部局連携データのみ	▲(同左)	×
不登校支援担当職員	×	▲ 不登校関連データのみ	▲(同左)	×
いじめ対策担当職員	×	▲ いじめ関連データのみ	▲(同左)	×
デジタル戦略室職員	●	×	×	×

教育政策室  
(総括管理主体) 所属

## 6.4 安全管理措置

組織的安全管理措置については、戸田市情報セキュリティポリシー<sup>25</sup>に準拠して対応している。また、人的安全管理措置として、安全管理措置に関する研修を行った。物理的安全管理措置としては、以下の観点で管理している。

- 入退室記録：職員証を IC カードとして入退室管理をしている。市役所の中で管理区画と市民が自由に入れるフリー区画があり、管理区画には職員証がないと入れない。更に、職員の中でも一般職員と情報システム担当職員では入れる区画が異なり、例えばサーバールーム等には一般職員は入室することができないようになっている。外部ベンダーについては、必要な作業に応じて許可を出しており、全て記録を残している。
- 外部記録媒体：統合管理ツールを使って許可されたものしか使用できないようになっており、許可されていない外部記録媒体を PC に指しても認識しない。利用する外部記録媒体は暗号化機能を有しているものに限り、利用時には、利用記録を残すことが定められている。
- 監査：全部署において、3年に1回内部監査及び外部監査を行って、戸田市情報セキュリティポリシーに従って、正しい運用がなされていることを確認している。

さらに、不正侵入や災害等に備えた措置については、庁舎内への出入りについて、業務時間外においては、常時警備員が待機しており、部外者の立ち入りは不可としている。

技術的安全管理措置として、今回のデータベースについては、システムを介しアクセスすることとし、データベースに搭載されている個人情報等へのアクセスに際しては、データベースにユーザーの認証機能を実装した。具体的には、ガイドラインに基づき、ID及びパスワードによる個人単位で、職種や所属等の必要な区分に基づいた権限管理（アクセスコントロール）を定めるとともに、例えば機微性の高い情報を取り扱う場合には、その性質を踏まえ、データ項目単位で生体認証等の2要素認証を行う（下図参照）など、適切な管理を行うこととした。このように、「付与する権限は必要最小限にする」との基本的な考え方の下で、アクセスコントロールのための措置を講ずることとした。

---

<sup>25</sup> 戸田市 情報セキュリティへの取組み URL：<https://www.city.toda.saitama.jp/soshiki/154/joho-security-top.html>



## 1. 2要素認証の構築イメージ

教育総合DBへの2要素認証の構築イメージとなります。

業務端末へのログインは従来通り①ID/PASSで認証し、教育総合DBへの認証を②生体認証としています。



図 6-6 2要素認証のイメージ

また、アクセスログ機能についてもデータベースに実装した上で、ログイン時刻やアクセス時間、ログイン中の操作内容等が特定できるようにし、アクセスログの確認を定期的に行うこととした。さらに、アクセスログへの不当な削除・改ざん・追加等を防止する措置やファイアウォールの設置など、外部からの不正アクセスを防止するための必要な措置についても講ずることとした。

併せて、データの出力機能をデータベースに実装する場合には、データベースから出力された情報について可搬媒体や情報機器による持ち出しは原則行わないこととし、その用途等について出力を行った者から報告を求めることとした。また、不要となった個人データについては、廃棄・消去を適切に行うこととした。

加えて、データベースの構築並びにデータベースの項目整理及び分析を担当する事業者においても、実証事業の実施の範囲内で、必要かつ適切な安全管理措置が行われるよう、本市は定期的な確認等を行うこととした。

## 7 データの準備

### 7.1 データの取得

教育総合データベースがデータを取得する対象は、戸田市立小中学校に在籍する児童生徒であり、主に令和3年度及び4年度のデータ等12,000件程度を取得している。データの連携方式は下図の通り。

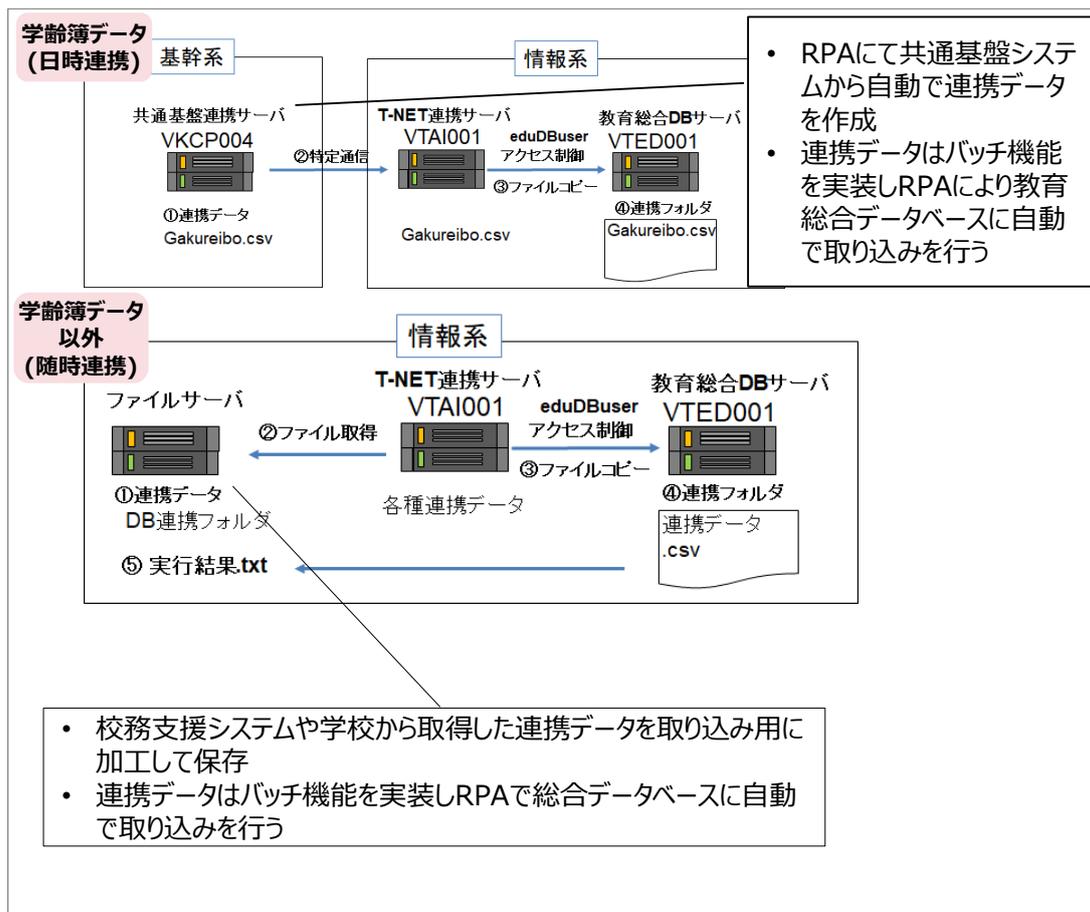


図 7-1 データ取得のイメージ

今後、データの連携頻度については検討すべき課題が残っている。例えば、今回の不登校の分析については、できる限り早期にデータベースに反映され、精緻な分析に繋げていくことが望ましい。しかし、実態としては、中学校においては欠席数が学期に1度の頻度、小学校でも月に1度しかデータとして出力できていない。将来的には、週次、できれば日次で出力できることが望ましいと考えられる。このようなデータ連携頻度に関する課題の発生要因としては、下表の内容が挙げられる。

表 7-1 データ連携頻度に関する課題

課題	具体例
①データの連携元の人による入力・記録作成の課題	日次でデータを作成していない（月次でデータを作成している）。
②連携元システムの仕様上の課題	系統的に、月次や学期単位でしかデータを抽出できない仕様になっている。
③データの取得にかかる課題	各種調査結果が届くのに時間がかかる場合がある。 ※調査によっては、Google Forms を活用した調査を行うことで、結果の取得期間の短縮が可能になっている。
④教育総合データベースへ連携させるデータを準備する工数の課題	データ準備に人手の工数がかかっているため、大量のデータを日次で連携しようとする、作業工数上、実現が難しくなる可能性がある。

## 7.2 アナログ情報のデジタル化

教育総合データベースが利用するデータのうち、Excelレベルではデータ化されているものがほとんどであったが、一部、アナログのものも実証事業開始時には存在した。

例えば、戸田市で独自に中学生を対象として実施している「学校生活アンケート」については、令和3年度までは紙で調査を行い、心理カウンセラーが目視で結果を見ながら対応が必要な生徒を抽出していた結果、心理カウンセラーの作業負担が大きく、調査の開始から最後の集計結果が出るまで約2カ月も時間を要していた。これを令和4年度からGoogle Formsで行ったことにより、アンケート終了後から日で対応の緊急性が高い生徒の結果を学校に返却できるなど、集計の簡素化、教師や心理カウンセラーの負担軽減につながったとともに、個別シートを基にした学校の生徒対応が早期に可能となったところである。

こうしたデジタルの効用について、下図にあるように、前後比較や関係者の視点からのメリットを含め、分かりやすい形で引き続き情報発信を行うことにより、データ活用に対する理解を醸成していく予定である。

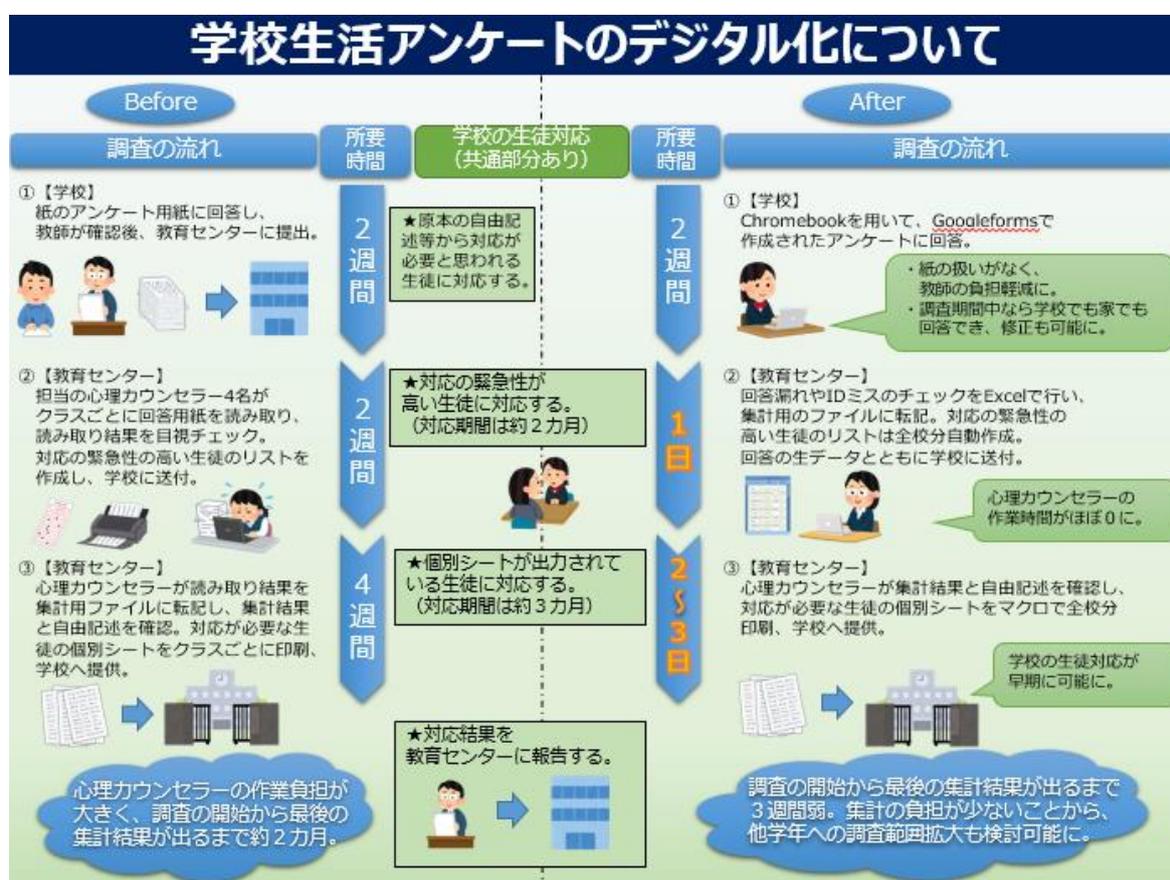


図 7-2 学校生活アンケートのデジタル化

## 7.3 名寄せ

名寄せとは、異なるシステム間で共通のユニークなキーが存在していない中、濁音の清音化等を行い、利用できるデータに整理した後、片方のシステムから別のシステムにユニークなキーを用いてマッチングさせた後、データを追加する作業である。



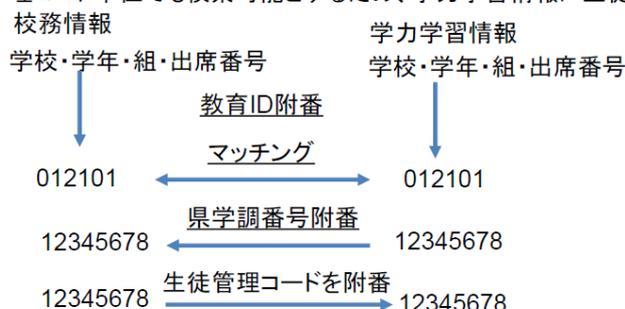
実証期間中は、去年度の校務支援データを対象に名寄せを行ったが、学齢簿データは本年度のデータしか抽出することができないような仕様となっており、学齢簿データと校務支援データの対象年度がずれていたため、学年の変更や転校によってマッチングできないケースが多く、名寄せができなかった件数が1,394件ほどあった。この対応として、一つ一つ手動で名寄せをするのではなく、年度がずれてマッチングできなかった卒業生分のデータを住基システムから抽出した上で、もう一度名寄せ処理を行うと、システム上のマッチングができなかった件数は100件程度に減少することが見込まれる。しかし、本実証事業期間では期間的に対応が難しく、この処理は行っていない。残りの100件程度のマッチングができない要因としては、以下が想定される。

- 校務支援システムに正式名が入っていない（外国人の通称名使用、旧姓使用等）
- 外国人のふりがなが正しくない
- 生年月日に誤りがある

次に、校務情報と学力学習情報の名寄せを行ったが、対象年度がずれていたためにマッチングができなかった1,394件については名寄せを見送った。これはシステム設計上、学齢簿情報と校務支援システム情報の名寄せが正しく行われないと他のデータとの連携が正しく行われなくなる仕様になっているためである。校務情報と学力学習情報の名寄せは、学校コード・学年・組・出席番号をもとにそれぞれに教育IDを敷設し、教育IDで校務情報と学力調査データをマッチングすることで、学力学習情報には生徒管理コードを、校務情報には埼玉県学力・学習状況調査番号を敷設する処理を行った（下図参照）。

#### 校務情報と学力学習情報との紐づけ

校務情報と学力学習情報との紐づけは、県学調番号で行いますが、校務情報に県学調番号を保有させるために教育IDを附番し、学力学習情報にも教育IDを附番することによりマッチングを行います。また生徒管理コード単位でも検索可能とするため、学力学習情報に生徒管理コードを附番します。



#### 【留意事項】

- ・出席番号が正しく入っていない場合がある。
- ・特別支援学級の組がひらがな等で設定されており、コード化が必要な場合がある。
- ・特別支援学級の出席番号が正しくない場合がある。
- ・学力学習情報は学校コードが入っていない場合があるため、あらかじめ学校コードのデータを追加する必要がある。

図 7-5 学齢簿と校務情報との紐づけの留意点

下表に、名寄せ及びデータ分析のための加工のプロセスを示す。

表 7-2 名寄せ及びデータ分析のためのデータ加工について

項目		内容	工数（人日）	
			ツール	手動
名寄せを行った理由		校務支援情報に学齢簿情報の宛名番号及び学習情報の埼玉県学力・学習状況調査番号を付与するため	-	-
名寄せを行った順番と件数		①清音化カナ氏名・生年月日で校務情報と学齢簿情報を突合し、校務情報に宛名番号を敷設	0.5	-
		②上記①の手順で生じたアンマッチに手動で宛名番号を敷設	-	0.2
		③学校コード・学年・組・出席番号により教育 ID を敷設	0.5	-
		④教育 ID（学校コード・学年・組・出席番号）で校務情報と学力調査データを突合し、生徒管理コードと埼玉県学力・学習状況調査番号を敷設	0.5	-
		⑤上記③④の手順で生じたアンマッチに手動で学校コード・学年・組・出席番号を修正	-	-
		⑥③, ④の作業を繰り返す。	-	-
名寄せに使った項目		清音化カナ氏名 生年月日	-	-
		学校コード・学年・組・出席番号	-	-
データ連携のための加工	連携先のデータフォーマットや表記に合わせるための加工	連携元データを Excel 形式で出力の上、手動で CSV 形式に変換	-	5
	欠損値や不整合データへの対応	重複データや表示形式の不整合について確認の上、手動で修正	-	5
データ分析のための加工	欠損値や不整合データへの対応	欠損を取り扱える機械学習モデルについては特段処理はなし。取り扱えないモデルは欠損値に平均値を埋めた。	-	-
利用できなかった情報の規則性や対応に時間がかかったこと		校務支援データと学齢簿データの対象年度がずれていた。対象年度をそろえたデータを抽出できるよう連携元データを改修することでアンマッチの件数が減る見込み。	-	-

## 7.4 データ分析のための加工

データ分析を行う上で、欠損が生じているケース（例えば、第3章の連携データ項目一覧に示したAi-GROWのあるデータ項目に対して、特定の子供のデータが欠損しているなど）があり<sup>26</sup>、これらについては予測モデルの分析の際に、欠損を取り扱える機械学習モデルはそのまま、取り扱えないものは欠損値に平均値を埋めて分析した。また、今回、データ粒度の統一については、取り扱った情報の状況から特に対応は不要であった。要因分析に関してはデータの標準化（全体の数値の平均値を0とし、標準偏差を取る）を行っている。

また、データをシステムに取り込む際に追加のデータ加工作業が発生した。大きく分けて、取り込み用CSVファイルへの変換、不整合データへの対応、そして名寄せエラーへの対応を行った（下図）。下図③の名寄せエラーについては前節で触れた通りである。①の作業については、変換マニュアルを参照して作業をしたが、作業者のデータを取り扱うスキルに依る部分もあり、例えば見た目は同じ数字でも表示形式が文字列なのか数字なのか見分ける必要があったりと、CSVファイルの取り込み時エラーが多数生じた。②の作業では、重複データについては重複したデータをピックアップして、原因の特定とそれに応じた他のデータと照合することで、重複データひとつずつ見ながらつづした。複数の重複が発生したら、正しい方法でやり直すといったことを行った。また、表示形式の不整合については、フィルタをかけて半角全角で違うことを確認したり、文字数を数式でカウントする等、不整合の形式によって適切な方法で判別して修正を行った。この作業は目視で行ったためかなりの工数がかかった。

---

<sup>26</sup> データの欠損が起きたケースとしては、以下のようなケースがある。

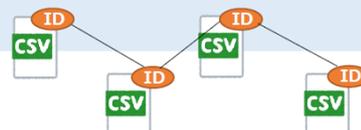
- ・調査日に欠席していてデータが存在していないケース
- ・特別支援学級の児童で、調査対象に含めなかったケース
- ・データ入力の時点で、出席番号等の入力ミスがあり、結果として重複データになってしまい、個人を特定できなかつたため、分析の対象外としたケース

## データベース構築（システム面）の進捗について③

### 【教育総合データベースへのデータのセットアップ】

システムにデータリストをセットアップするために、以下の作業等を実施

- ①Excel ファイルを取込用CSVファイルへ変換
- ②データ形式整備（重複データ削除、表示形式との不整合対応等）
- ③名寄せエラー対応（約 1 割のデータにエラー）



#### ①取込用csvファイルへの変換

- ・Excelのデータリストを、取込用のCSVファイルへ変換
- ・**全41ファイルを手作業**で実施したことにより、**取込時多数のエラー**が発生

#### ②データクレンジング

- ・データリストの作成段階で一通り確認したものの、システム取込の段階で、**細かいエラーをつぶし切れていなかった**ことが発覚  
例) 重複データ、表示形式の不整合（年月日の表示形式や、数値の桁数など）等

#### ③名寄せエラー対応

- ・本システムは、宛名番号をベースに児童生徒のデータ連携を実施
- ・名寄せできなかったエラーで対応可能なものは**個別に対応**

8

図 7-6 データベースへのデータ取り込みにあたっての加工

### 7.5 マスキング

分析等を行う際の情報の取扱には、元データである個人が特定されないようにするための対応が必要である。氏名や生年月日等、個人を特定できる情報はもとより、他の情報との組み合わせにより個人を特定できる情報についても保護する必要がある。

戸田市においては、事務担当者が各利用データに、氏名にまつわる情報（漢字表記やフリガナ等）が含まれているかどうかを目視し、含まれている場合には、その情報が含まれている列を指定して暗号化するマクロを利用し、暗号化した。漢字やフリガナ表記に加え、テキストベースで個人名が出てくる箇所については、項目ごと削除している。どこまでのデータを分析に使うのかを判断するのがポイントであり、今回でいえば、組や番号等、判断が難しい情報も含まれていた。

マクロは、これら個人を特定できる情報を記載した列を指定して、SHA256関数を用いて暗号化するものである。下図の黄色ハイライトセルに、個人を特定できる情報を含むファイル名（パス含む）、シート名、列番号、開始行番号を記載し、「実行」ボタンを押すことで、暗号化が完了する。

項目	記入欄	記入例	備考
バス・ファイル名	C:\Users\orang\OneDrive\Hanjuku\Toda\20220731_個人情報暗号化マクロ_チェック用サンプルデータ - Copy.xlsx	C:\test.xlsx	元ファイルの値を変更しないよう、コピー版ファイルのバス・ファイル名を記載する
暗号化対象シート名	Sheet1	Sheet1	
暗号化対象列番号	1	1	ExcelのA列の場合1, B列の場合2, ...
暗号化対象開始行番号	2	1	Excelの2行目から始まっている場合2

図 7-7 個人情報暗号化マクロ

分析主体である半熟仮想にデータを受け渡す際は、戸田市として利用しているオンラインストレージサービスを利用し、マスキングした上でパスワードをかけて送付した。

## 7.6 外字

教育総合データベース上の連携において、氏名は文字情報基盤<sup>27</sup>を採用した（詳細は下図参照）。住民情報システムは外字となるため、共有基盤に格納の際に文字コードの変換を行った。戸田市では、以前から、外字をすべて同定し文字情報基盤との対応を整理するとともに、文字情報基盤をマスターとする文字管理システムを構築、既存文字と文字情報基盤との対応テーブルを維持してきた。また、変換テーブルに基づき、既存の文字を文字情報基盤に変換できるツールも構築したが、これまでは、既存ベンダーが文字情報基盤に対応することが出来なかったため、実装することが出来なかった。今回は新規開発のシステムであったため、文字情報基盤を最初から実装したシステムを構築できたという経緯がある。現在、一部の文字が文字情報基盤に存在していない文字のため、既存文字と文字情報基盤との対応テーブルを包摂しているが、今後はデジタル庁が整備するMJ+と同定し、MJ+に変更する予定である。

<sup>27</sup> 行政で用いられる人名漢字等約 6 万文字の漢字を整備したもの。住民基本台帳ネットワークシステム統一文字、JIS、常用漢字に含まれないなどの、標準がなかった漢字の標準化を推進。文字情報基盤により、これまで外字でしか表現できなかった人名、地名を標準規格に沿って表現できるようになる。

## 5 文字情報について

教育総合データベースで使用する氏名は文字情報基盤を採用する。

- ・文字集合は UNICODE (ISO/IEC 10646) とする。
- ・IVD は MOJL\_JOHO コレクションとする
- ・使用するフォントは IPAMJ 明朝とする。

住民情報システムが FUJ 明朝 (90JIS) ベースなので共通基盤に格納する際、文字コード変換を行っている。

氏名のみ IVS を含む文字情報基盤に対応しているが、IPAMJ 明朝はデフォルトグリフ表示機能により、文字情報基盤に対応していない漢字項目についても文字化けせず表示が可能。

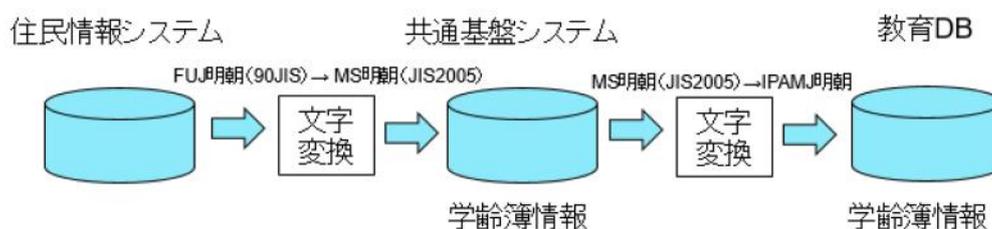


図 7-8 文字情報の外字対応について

## 8 システムによる判定基準の設計と評価

教育総合データベースのデータをもとに、不登校等のリスクについてアラートを表示する機能を検討してきた（具体的なユースケースの想定は2章参照）。本実証事業期間中は、目的①子供たちのSOSの早期発見・支援に即して、特に不登校になるリスクについて判定する機能の構築を検討した。具体的には、ある児童生徒について、教育総合データベース上のデータ（学校生活アンケート、県の学力調査、学習状況調査児童生徒質問し、授業がわかる調査などの回答）をもとに、将来不登校になるリスクがどの程度高いかを判定し、高い場合にアラートを表示することを検討してきた。ただ、実証期間中には現場での実用に耐える判定ロジックの構築が実現しなかった。本節では、その判定ロジックの検討過程と今後の展望について記載する。

### 8.1 データ分析と分析結果の評価

不登校のリスク判定機能を構築するために、戸田市では①ショートリスト分析と②ロングリスト分析の2段階の

分析を行った<sup>28</sup>。①のショートリスト分析はデータの取得対象及び利用するデータ項目を限定して行った分析であり、データ数を増やした分析の際に参考となる傾向を見出すことを試みた。それを踏まえ②のロングリスト分析では、扱うデータ件数及びデータ項目も増やし、①で見られた傾向等も踏まえた分析を進めた。分析手法の検討に当たっては、文献として「統計的機械学習の基礎-データマイニング・推論・予測<sup>29</sup>」やBelloniらの論文「Inference on Treatment Effects after Selection among High-Dimensional Controls<sup>30</sup>」などを参考にしている。

ショートリスト分析では、一つの中学校における単年度のデータを用いて分析を行った。不登校の指標として長期欠席調査のデータを用い、それを予測する変数として出欠情報や保健室利用データ、学校生活や学習習慣に関する県学力調査質問紙調査や人間関係に関する学校生活アンケート、また学力の指標として県の学力調査の学力分析データを用いた（下表）。

表 8-1 ショートリスト分析で用いたデータ

年度	令和3年度		
対象生徒	市内中学校（1校）の1年生及び2年生（計約700人、長期欠席の生徒68人）		
利用データ	不登校（長期欠席）	長期欠席調査 ※教師による手書きコメントについては未利用	
	生活状況	客観的	・出欠情報 ・保健室利用データ
		主観的	・埼玉県学調質問紙調査（学校生活や学習習慣に関するアンケート） ・学校生活アンケート（人間関係等に関するアンケート）
学力	県学調学力分析データ ※非認知能力については未利用		

ショートリスト分析では、前学期の欠席率が高い場合、次学期の長期欠席になる傾向が高いという結果が得られた。その他、保健室の利用回数が多い生徒や学力が低い生徒、また友人関係が良好ではない生徒に欠席の傾向は見られたが、いずれも統計的に有意な結果は得られなかった。課題としては、サンプルサイズが小さかったことや欠席に関するデータの情報量が少ない（学期ごとの欠席数のデータしかない）ことなどが挙げられた。ロングリスト分析では、ショートリスト分析を拡張し、サンプルサイズを大きくして、複数の予測項目を組み合わせて分析の深掘りをすることを目指した。（図8-1、8-2参照）

<sup>28</sup> 今回は必要なデータが出そろうまでに時間を要したため、一旦入手可能なデータで、今後のデータ分析の目安をつけるための初期分析を行った。この分析をショートリスト分析と呼ぶ。

<sup>29</sup> 「統計的学習の基礎：データマイニング・推論・予測」Trevor Hastie, Robert Tibshirani, Jerome Friedman 著；井尻善久 [ほか] 訳，共立出版，2014.6

<sup>30</sup> Belloni, A., V. Chernozhukov, and C. Hansen, "Inference on Treatment Effects after Selection among High-Dimensional Controls", The Review of Economic Studies, Volume 81, Issue 2, April 2014, Pages 608-650

今回の分析の結果、最も長期欠席との相関が高い項目は、**前学期の欠席率**であった。その他にも相関が見られる項目は存在するが、相関の程度は小さく、**現時点で断定的な傾向として捉えることは難しい**。

利用データ種類		分析結果（概要）
生活状況	客観的	<ul style="list-style-type: none"> <li>・ <b>前学期の欠席率が高い生徒</b>は、長期欠席になる傾向が非常に高い。</li> <li>・ <b>保健室を利用している生徒</b>は、欠席率は高いが長期欠席になる可能性は高いとは言えない。</li> </ul>
	主観的	<ul style="list-style-type: none"> <li>・ <b>友人関係が良好でない生徒</b>は、長期欠席になりやすい傾向がある。一方で、家庭内及び教師との関係は、長期欠席にはあまり影響を与えていない可能性がある。</li> <li>・ 学校の楽しさ、部活参加の有無、相談希望の有無、を含むその他の質問項目には、長期欠席との関係性が現時点ではあまり見られなかった。</li> </ul>
学力		<b>学力が低い / 授業が分からないと感じている生徒</b> は、欠席率が高い傾向がある。

図 8-1 ショートリスト分析の結果概要

#### 4. 現在の課題と今後の対応策

現在の課題	今後の対応策
<b>データ数が少なく、分析結果の解釈が難しい</b>	<ul style="list-style-type: none"> <li>・ データ不足故に、全体の平均傾向がうまく読み取れず、<b>個々のケースに結果が左右されてしまう</b></li> <li>・ また、AIモデルを構築するには圧倒的に正解データが少なく、<b>過学習をしてしまう</b></li> </ul>
<b>最重要項目である欠席率についての情報量が少ない</b>	<ul style="list-style-type: none"> <li>・ (1)満足できる予測のために<b>最低限必要なデータの項目と分量の条件</b>を洗い出す</li> <li>・ (2)(1)で出した条件を満たす<b>データを整備するか、予測は諦めるかの意思決定</b>を行う</li> </ul>
<b>主観的な回答の解釈が難しい</b>	<ul style="list-style-type: none"> <li>・ 中学生より<b>記録幅の細かい小学生の出欠記録</b>を用いて、より詳しい欠席パターンが抽出できるかを検証する</li> <li>・ また、<b>複数年度のデータ</b>が使える場合は、長期的な欠席傾向の変化も分析に組み込む</li> </ul>
<b>データ数が少なく、分析結果の解釈が難しい</b>	<ul style="list-style-type: none"> <li>・ 生徒が人間関係などの辛さを<b>アンケートで正直に答えているとは限らず</b>、真の問題が見えづらい</li> <li>・ <b>複数項目の組み合わせ</b>を工夫する</li> <li>・ 県学調査問紙調査やAiGrowなどでスコア化されている、<b>個人の非認知能力を組み合わせる</b></li> </ul>

図 8-2 ショートリスト分析の課題と対応策

ロングリスト分析では、ショートリスト分析を受け、令和3年度の単年度データではあるが、市内全域に取得範囲を広げ、利用データ項目も増やし、欠席数予測と長欠率予測の主に2つのパターンを用いてアプローチを行い、パラメータなどの詳細設定を変えることで、約1,000種類のモデルを構築・検証した（詳細は巻末の別添資料を参照のこと）。

モデル構築では、月初に入手可能なデータを用いて当月に不登校となる可能性の高い生徒を特定することを目指している。また、モデル検証では、各モデルについて、毎月そのモデルが算出する不登校可能性に基づいて

対応を行った場合、どれだけの不登校児童へアプローチできるかの性能を評価する。分析では、当月に新たに長欠となる「新規長欠生徒」を予測の対象として設定し、年内に長欠を1度以上経験している生徒については分析対象からは除外した（利用可能データが1年分に限られていたため、年度をまたいだ長欠経験については考慮していない）。「新規長欠生徒」は①年度初めから当月初日までに一度も長欠を経験していないこと、②当月初日までの総欠席数が30より少ないこと、③当月に長欠とみなされること、以上の基準を満たす生徒のこととした。

結果、長欠リスクが一番高いと予測された児童5人（学校あたり）にアプローチすると、新規長欠生徒の50%をカバーできるとがわかり、このモデルでは、欠席情報（累計欠席数と前月欠席率）が重要であることが示唆された。ただし、現状では、まだ教育現場での活用には精度が不足していると考えている。今後、長欠リスクの学校での活用方法、機械学習の精度向上に取り組んでいく必要がある。

分析内容	<ul style="list-style-type: none"> <li>予測モデルの構築：現在利用できるデータをできるだけ網羅的に用いて、<b>各月の欠席数</b>（病欠あるいは事故欠）と<b>長欠認定の発生</b>を予測するモデルを約1000通り構築。</li> <li>モデルの精度検証：大量のモデル候補の性能を網羅的に比較することで、最適なモデルを学習し、モデルごとの精度の違いを検証した。</li> </ul>
分析結果の要点	<ul style="list-style-type: none"> <li>現時点での最適なモデルで<b>長欠リスクが一番高いと予測された児童5人（学校あたり）にアプローチすると、新規長欠生徒の50%をカバー</b>できるとがわかった。</li> <li>このモデルでは、<b>欠席情報</b>（累計欠席数と前月欠席率）が<b>大きな影響力</b>を持ち、その他の変数はモデルの精度向上にはあまり大きく寄与していないことが示唆された。</li> <li>またそれとは別に、過去の総欠席数順をそのまま長欠リスクとする単純なモデルも高い性能を発揮した。</li> </ul>
今後の課題	<ol style="list-style-type: none"> <li>長欠リスクを学校現場にどのように伝えて活用すべきか・すべきでないか検討する。</li> <li>サンプル数と使用できる変数の数を増やすことで、データの拡張を目指す。 <ul style="list-style-type: none"> <li>現状では小学校/中学校のみにしか存在しないデータが存在する。</li> <li>学年ごとに質問番号が異なるデータなど、結合の難しいデータが存在する。</li> </ul> </li> <li>より質と量が高まったデータで複雑な機械学習モデルの性能を再検証する。</li> </ol>

図 8-3 今回不登校予測モデルの分析結果

今後は、分析の精度を高めながら、長欠リスクをどのように学校現場に伝え、活用すべきかを検討していく。分析の精度を高める観点では、今あるデータを2年度分、3年度分と拡張していくことで、詳細な経年分析を行ったり、環境や季節性による傾向を見つけてモデルに組み込むようなことも可能になるだろう。より、質や量の観点でデータを充実させながら、より分析の精度が高くなるような、複雑な機械学習モデルの性能を再検証していくことが課題である。

また、不登校に影響を与える因子と、因子間の交互作用<sup>31</sup>を検証した。結果として、学力、生活リズムの乱れ、遅刻率、早退率、保健室利用率、BMIといったデータにおいては欠席率と正の相関がみられたが、不登校に影響を与える因子として断定的な結果は得られておらず、今後も検証していくことが必要との結論となった。

<sup>31</sup> ある因子に影響を与える因子があったかどうか、それが欠席率低下の抑制要因となりうるか否か、などを検証した。

分析手法	<ul style="list-style-type: none"> <li>• <b>単回帰</b>：長欠/欠席率と、各因子候補の2変数間の相関関係を測る手法</li> <li>• <b>ダブルラッソ回帰</b>：データ駆動で長欠/欠席率と関連度の高い因子、またはそれらの因子と関連度の高い変数のみを選び、各因子の長欠/欠席率との関連度合いを測ることのできる手法</li> </ul>
分析結果の要点	<ol style="list-style-type: none"> <li>① 単回帰分析では、<b>学力、生活リズムの乱れ、遅刻率、早退率、保健室利用率、BMI</b>に欠席率との正の相関が見られた。</li> <li>② それらのうちいくつかの項目には、抑制要因になりうる可能性がある因子も見つかったが、現在のデータ量では断定的な結果は得られていない。</li> </ol>
今後の課題	<ol style="list-style-type: none"> <li>① 不登校因子になりうる追加項目の調査を学校とも連携して進め、より詳細で包括的なデータの収集を目指す。 <ul style="list-style-type: none"> <li>➢ 現状のデータでは潜在的かつ重要な特徴量の抜け落ちの可能性が高く、因子ごとの要因を正確に予想することが難しい。</li> </ul> </li> <li>② 不登校生徒の回答の抜け落ちに対応する。 <ul style="list-style-type: none"> <li>➢ 現データでは不登校生徒の回答が欠落しているケースが多い。</li> </ul> </li> </ol>

図 8-4 不登校の要因分析結果

## 8.2 判定基準の構築と評価

8.1節に記載の通り、現場で判定基準としても活用するには必要な精度が得られておらず、今回の事業期間では、教育総合データベース上に判定機能を構築していない。他方で、前述の通り、協力校にて、プッシュ型支援の検証を行うこととした。学校現場との意見交換や活用可能なデータの観点から、下記の判定基準（いずれかに該当すれば抽出）を抽出しつつ、最終的には協力校との協議を踏まえ、見守り対象の子どもを決定し、プッシュ型支援の試行を行った。

- ① 埼玉県学力・学習状況調査において前年度から学力レベルが不変又は低下している
- ② 戸田市「授業がわかる調査」第1回（6月）→第2回（12月）で回答の肯定度合いが減少している
- ③ 学校独自の「幸せアンケート」第1回（6月）→第2回（1月）で回答の肯定度合いが減少している

本取組においては、市の教育委員会にて、上記の①～③の結果や推移を簡易的にダッシュボード化し、見守り対象の子どもを抽出する際に参照できるようにした。学校や学年や学級ごと、また個人ごとのデータがカテゴリの選択状況に応じて集計される仕組みである。実際に活用した教師からは、「それぞれバラバラに、かつ定点で見ていたデータを一覧や推移で見ることができ、アンケートの数値が下がっている児童がいたこと、そしてその児童のことをあまり気にかけてあげられていなかったことに気付いた」「学校にとって比較的身近なデータが集計されていたので、データを確認するハードルが低かった」といったコメントがあった。

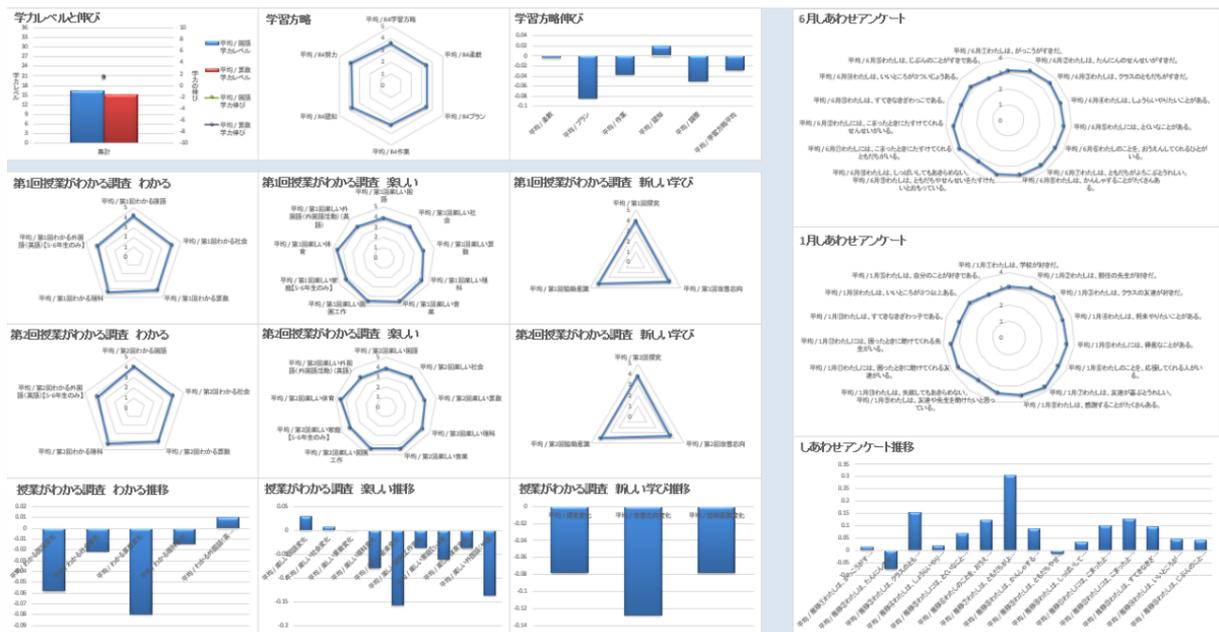


図 8-5 今回の事業で構築した簡易ダッシュボード

当初は市教育委員会がデータの推移を見ようとしたが、そもそも調査未実施の児童も多く、抽出対象から外れていたため、最終的には市教育委員会が抽出した児童に限らず、取得できたデータが極端な児童や、不登校傾向がみられる児童など、学校生活での実態を見ながら学校が総合的に判断し、学校全体で20名弱の児童を抽出した。

## 9 事業効果の評価・分析

### 9.1 支援業務の試行結果

教育総合データベースを用いた早期のリスク判定機能については、今後の実装予定となるため、それによるプッシュ型支援は今後、本格的に実行していくこととなる。本節では、協力校における取組や学校カルテを用いた独自分析から得られた示唆をもとに、今後教育総合データベースを用いたプッシュ型支援の展望について記載する。

協力校における取組では、前述の児童に対して、プッシュ型支援の有効性を検証した。抽出した児童について、学校で利用しているケース会議用の資料を用いて、教師が対象児童の2月初め時点までの学校生活の様子を見取った情報と、2週間後の変容を入力してもらった。この情報は蓄積され、教育相談や児童支援の記録と紐づけて支援のための情報として利用していく予定である。

The form is titled 'ケース会議用フォーマット' (Case Conference Format). It is divided into three main sections: A (Specific support methods), B (Wishes/Actions), and C (Confirmation/Methods). Below these are sections for '児童の実態' (Child's Status), '考えられる背景' (Possible Background), and '2週間後の実容' (Status after 2 weeks). The bottom bar contains buttons for '新規登録' (New Registration), '上書き保存' (Overwrite Save), '入力情報の一括クリア' (Clear All Input Information), and '削除' (Delete). Navigation arrows point to 'ダッシュボードへ' (To Dashboard) and 'ケース会議記録一覧へ' (To Case Conference Record List).

ケース会議用フォーマット

The dashboard is titled '2022 年度 サポート ダッシュボード' (2022 Annual Support Dashboard). It shows '児童生徒情報' (Student Information) and a table of '各種記録について' (About Various Records). The table has columns for '記録の種類' (Record Type), '全件数' (Total Count), '2022年度件数' (2022 Annual Count), '2021年度件数' (2021 Annual Count), and '連続対応' (Continuous Response). Below the table are buttons for '記録の入力へ' (To Record Input) and '記録の入力へ' (To Record Input).

記録の種類	全件数	2022年度件数	2021年度件数	連続対応
D記録	1	1	0	0
E記録	0	0	0	0
F記録	0	0	0	0
G記録	3	3	0	3
H記録	0	0	0	0

相談系、支援系のダッシュボード  
(各種記録の件数や有無を確認)

他の記録と紐づけて確認

図 9-1 支援内容の記録イメージ

今回抽出された支援対象の児童の特徴は以下の通りである。

- 自己肯定感を中心に数値の低い項目が多い
- 児童間トラブルがあった
- 家庭との連絡が取りづらい
- 学習への不安を抱えている
- 自己表現などが苦手で、教師や友人との関係が比較的希薄

人によるアセスメントの観点では、多くはデータと実態の整合性が高く、データで抽出される児童は教師が気に掛ける児童と一致する場合が多いが、データを見て気付いたケースもあった。実態把握にはQ-U<sup>32</sup>アンケートも活用したが、人間関係も見える点で有効だった。以下が、今回の要支援児童に対して行った、実際の支援内容とその効果の例である。

<sup>32</sup> QUESTIONNAIRE-UTILITES の略。学校生活意欲や学級満足度、ソーシャルスキルをアンケートにより調査する。

表 9-1 協力校での検証における支援内容と効果の例

支援内容	効果
ノーマークだった児童への担任からの声掛けを意識的に増加	授業で発表したことがなかった児童が発表するように
ばれっとルーム <sup>33</sup> の活用を促す	学習に不安を感じる児童、自己肯定感が低めの児童のセーフティーネットに
保護者と児童の最新の状態について共有する	保護者が不登校の視点で児童に目をかけるように

また、学校カルテの試行においては、埼玉県学力・学習状況調査の児童生徒質問紙から抽出した項目および授業がわかる調査の分析を学校別に行い、その結果について学校の管理職にフィードバックを行った。

- 「難しいことでも失敗をおそれないで挑戦していますか。」といった、戸田市として重視する資質・能力に関連するデータの傾向（学年別傾向）
- 授業がわかる調査における、学級ごとの、第1回（6月）から第2回（12月）にかけての肯定的な回答の推移の傾向（組別傾向）

以上について、学校・学年・学級ごとに傾向の差が見られた。また、教師の授業の工夫が子ども達から見たデータでも肯定的に裏付けられているという事例も確認された（下図参照）。

<sup>33</sup> 不登校（傾向）の児童への居場所としての校内サポートルームで、当該協力校には令和4年11月から設置。

## 学校訪問での活用資料の例①（県学調質問紙の分析）

「難しいことでも失敗をおそれないで挑戦していますか。」については、肯定的な回答の割合が小4は県・市平均を上回る。小5・小6は、昨年度（小4・小5時）から引き続き、特に、強い肯定の回答の割合が県・市平均よりも低い。

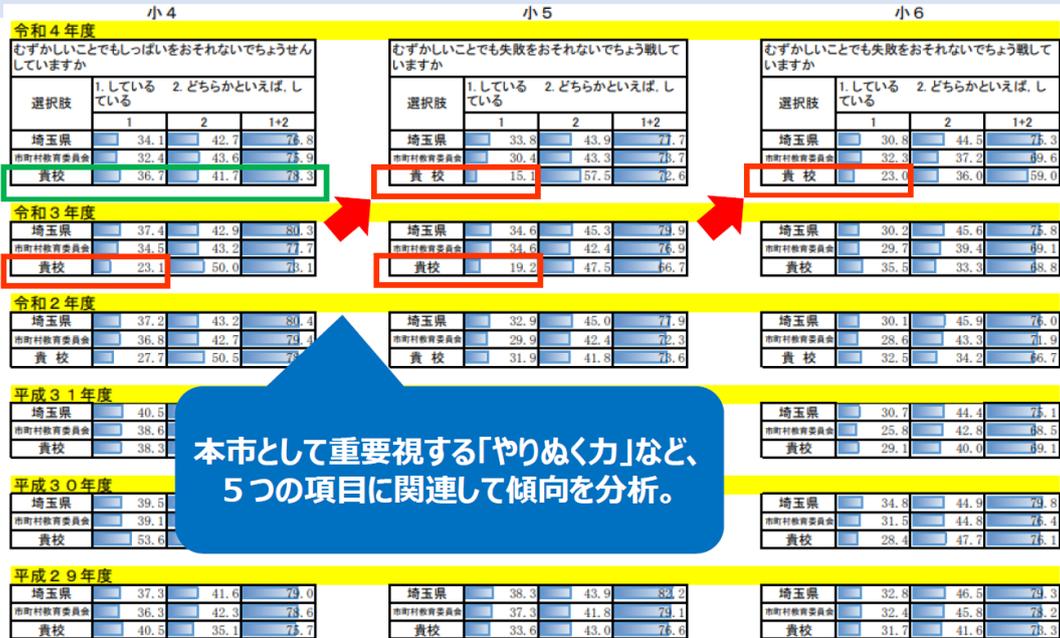


図 9-2 学校カルテの試行（埼玉県学力・学習状況調査の分析例）

## 学校訪問での活用資料の例②（授業がわかる調査の分析）

<学級別の傾向（小5）>

全体的には1組が肯定的な回答の割合が増加傾向にある教科等が比較的多く、2組が減少傾向にある教科等が比較的多い。3組は②「授業が楽しいですか」の家庭・外国語、探究心及び協働意識が肯定的な回答の割合が減少傾向にある。

令和4年度第1回（1組）

令和4年度第2回（1組）

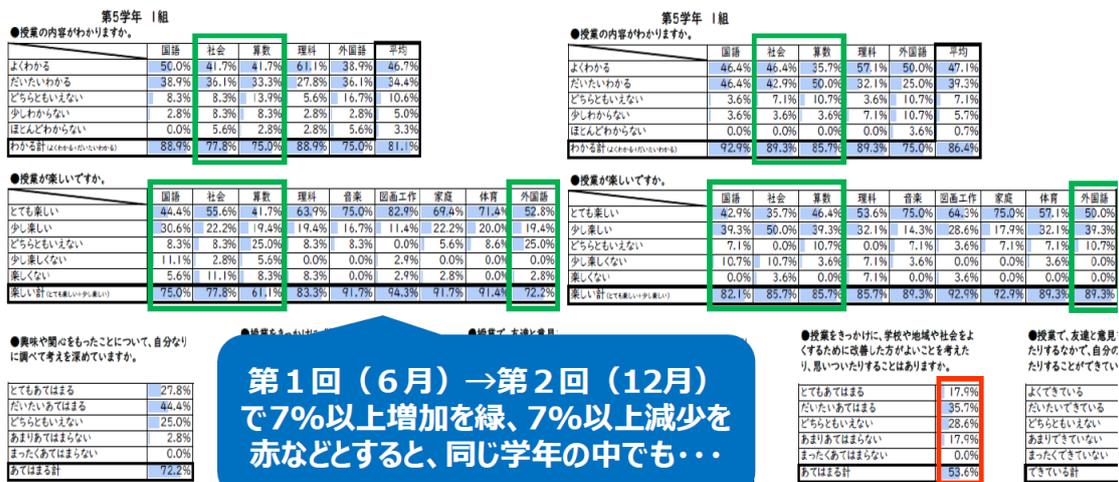


図 9-3 学校カルテの試行（授業がわかる調査の分析例①）

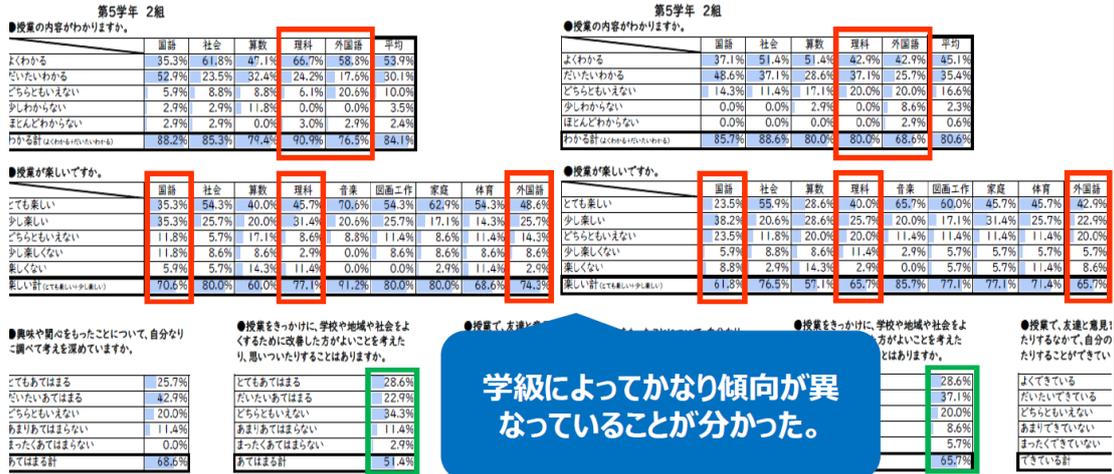
## 学校訪問での活用資料の例③（授業がわかる調査の分析）

＜学級別の傾向（小5）＞

全体的には**1組が肯定的な回答の割合が増加傾向にある教科等が比較的多く、2組が減少傾向にある教科等が比較的多い**。3組は②「授業が楽しいですか」の**家庭・外国語、探究心及び協働意識が肯定的な回答の割合が減少傾向**にある。

令和4年度第1回（2組）

令和4年度第2回（2組）



20

図 9-4 学校カルテの試行（授業がわかる調査の分析例②）

今回、学校カルテの試行結果について、学校長へのアンケートを行ったところ、今回提示した資料については、全員から「非常に」又は「まあまあ」有用であるとの回答が得られている。

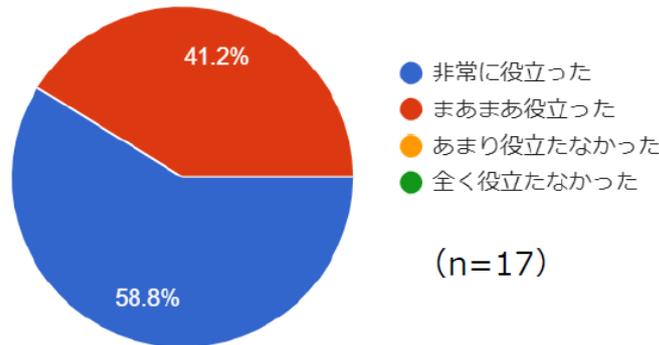


図 9-5 学校カルテの試行に対するアンケート結果（役立ち度）

回答理由に関しての主なコメントは以下の通り。

- 静的な平均値だけではなく、様々な粒度に分解すると言うことで、データの見方が変わり、活用の方向性が定まりやすくなった。
- 面談は終わりましたが、関係の職員にはいただいた資料を参考にして、個別に話をしたいと思いました。
- 人事評価の期末面談で活用でき、本人が授業・学級経営について振り返る機会、材料となった。また、

埼玉県学力・学習状況調査の質問紙分析は、研修部で共有を図り、今後の研修の方向を探る材料として活用できた。

また、具体的な活用方法については、データ利活用の考え方の共有や管理職内での共有、成果や課題が感じられた学級又は教科等のフィードバック等に回答が集まった。

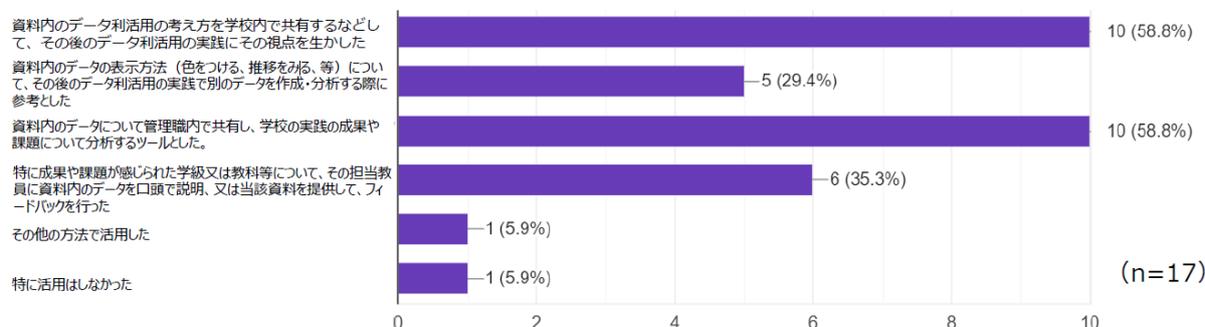


図 9-6 学校カルテの試行に対するアンケート結果（活用方法）

今後、学校でデータ利活用の実践を進めていくに当たっての課題について聞いたところ、教職員のデータリテラシーやデータの散在、授業改善等に生かす実践の不足等の回答が多かった。

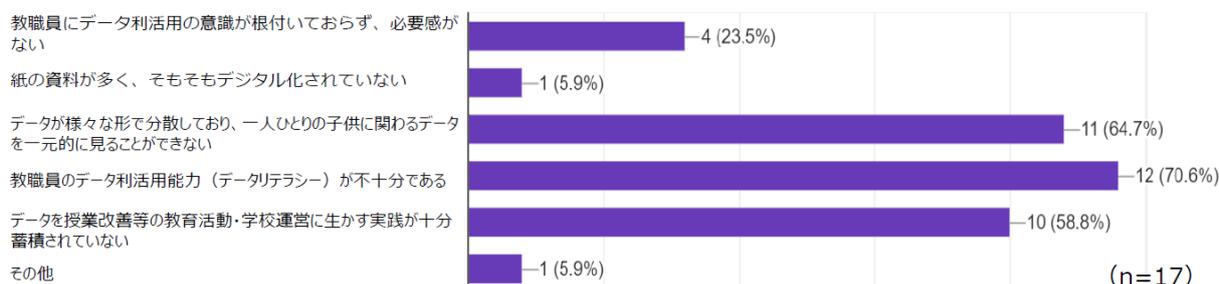


図 9-7 学校カルテの試行に対するアンケート結果（今後の課題）

回答理由に関する主なコメントは以下の通り。

- 管理職が、このデータを見て、ひとりひとりの職員にどのようにフィードバックしていくか。良いところを示しながら、さらに伸ばしていく、良さを広げていくという全面展開の仕方。課題を示しても、職員のやる気をそぐだけであると思われるので。
- データ利活用について、数値的な分析からもう一步踏み込んで、「現状の考えられる原因」「改善策の例示」などが示されると有り難い。それは、学校がやることであるとも考えられるが、分析等の時間の確保、誰が行うのか、周知して自分の指導等に生かすようになるまでの手間、さらに出てきた課題を教師で共有する等の時間を捻出するのが難しい。
- 学校が所有している数多くの調査・学力データを児童個人と教師に紐付けして6年間蓄積させ、必要事案に応じて管理職や教師が必要なデータを組み合わせて表示・分析等の活用することができるシステムの構築。学校の職員が調査データを追加できるとフレキシブルな運用が可能となり、活用ハードルが下がると思う。
- データ量が多く、校内で整理・分析するのに時間がかかってしまう。効率的なデータ整理、分析の方法等

について担当者研修やデータ活用の教師向けの研修などのサポートがあるとよいと思います。データ利活用のよさを一般の教師に実感させる研修が必要。

- いただいた資料であるが、平均値のみがクローズアップされており、集団が理想的な正規分布であればある程度は集団の傾向はつかめるだろうが、現実の集団分布は様々であり平均値だけで集団の傾向を捉えることには無理がある。現在、本校では市教委と連携して、学年や学級を平均値のみで一括りにするのではなく、各学級に集団を（中略）分け、集団ごとに各教科、学習方略、伸び等の数値から有意差検定を行い、特性や要支援児童を洗い出し、学習指導に生かしている。今後も市教委と連携して、宝の山である埼玉県学力・学習状況調査のデータを有効活用していきたい。

## 9.2 今後の取組

現時点での課題と今後の方向性を下図に示す。本実証事業期間においては、不登校発現リスク判定に向けた不登校予測モデル構築及び不登校要因分析を行ったが、プッシュ型支援への活用という実用に耐え得るためには、分析精度を更に高めていく必要があるとの結論となり、本実証事業期間中には、リスク判定機能の実装までは至っていない。今後、新たに取得するデータの項目を追加したり、既存のデータ項目についてはデータ量を増やしたりすることで、精度の向上を図っていくとともに、データの取得方法を見直して、データベースに容易に搭載できるようにすることも考えられる。

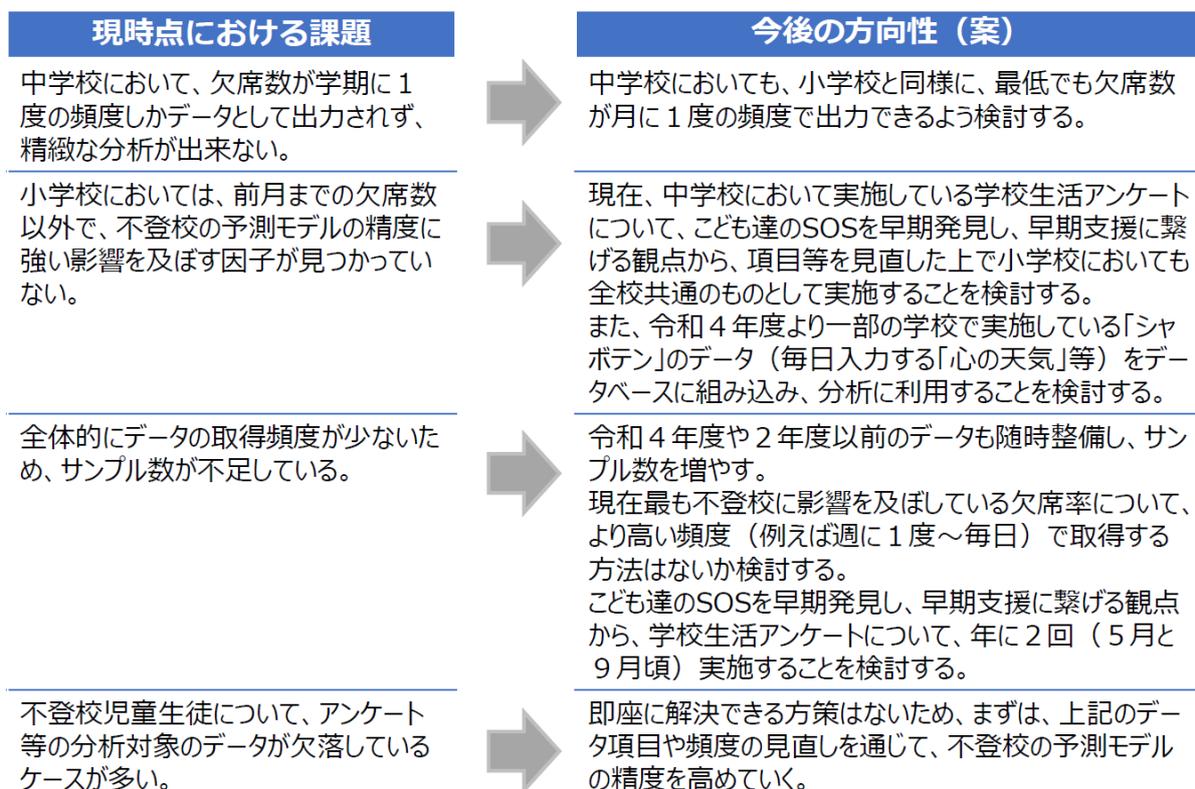


図 9-8 現時点での課題と今後の方向性案

今回の協力校におけるプッシュ型支援試行においては、個々の児童のデータを学校内で一元的に、かつ短時間で見られるようになったことで、改めてデータを見ながら分析し、こどものことを話す時間を確保できた点が効果

であるが、一方で、不登校（傾向）になる前の初期対応が必要かつ重要だが、担任が気付いても情報が学校全体として共有されにくいという点も課題として挙げられた。発見と初期対応（ケース会議をしてスクールカウンセラーに繋ぐ等）の仕組みを組織的に確立させる必要があり、断続的な休みが見られ始めたときに要因を特定し、支援方法を決定するためにもデータ利活用が求められる。このように、複数段階でみられる課題を解決するために、データを利活用した改善サイクルをどのように回していくかを検討していく必要があると考えられるといった意見が協力校からは挙げられた。

## 10 全国的な展開方策の検討

他自治体からの転入等で情報を連携することが有用になる場合も考えられるが、そもそもどのような情報が自治体間で共有可能かについては、法令上定めのあるもの以外は、個人情報の必要最小限度の原則に照らせば、一自治体としては、あくまでの自治体内の住民に対して行政サービスを行う立場であり、他自治体から情報を共有してほしいというニーズがあったとしても、簡単に渡せるような状況にない。技術的な観点でシステム間の相互運用性という観点もあるが、現状での最大の壁はこのような制度ではないかと考えられる。

また、システム的な観点で他団体への同様の取組の普及を考えると、標準化の観点は重要であろう。今後は、戸田市が全国に先駆けて新たな課題に取り組むファーストペンギンとして、セカンドペンギンがこの取組に挑戦する際のハードルが下がるよう、データの標準化やデータフォーマットのオープン化についても取り組んでいく方針である。また、機械学習等での予測モデルを構築するには1自治体のサンプルサイズが小さい。本市においても複数年度のデータを整備したり、新たなアンケート調査を検討していくが、同様の取組に挑戦する自治体同士のプラットフォームを国主導で形成していただくことも一考の余地があるのではないかと考える。

## 11 まとめと今後の課題

本実証事業では教育総合データベースを構築し、データ分析・プッシュ型支援につながる判定基準の設計に取り組んだが、各種課題も見えてきた状況である。

今後は、当初より掲げていた教育総合データベースの活用イメージ（①子供たちのSOSの早期発見・支援、②貧困・虐待等の困難を有する子供への支援、③学校カルテによる現場への継続的改善のためのフィードバック）の実現に向けて、検討を進めていく。当面の具体的な取組としては、児童生徒ダッシュボード（各児童生徒の基礎情報（氏名等）や各種調査の結果・回答などが、一覧目グラフ等の分かりやすい形で表示可能）、不登校等の発見・深刻度等のリスク判定機能、学校カルテ（各学校の学年・学級単位で各種調査やアンケート結果が学年等の集団単位で経年比較等できる可能）の機能を搭載し、本市教育委員会職員だけでなく、小中学校長等も容易に理解、活用可能なUIを備えたデータベースの構築に向けて取り組んでいきたい。また、上記活用イメージの②貧困・虐待等の困難を有する子供への支援については、令和4年度においては検証する時間的な余裕がなかったため、令和5年度において本格的に着手する予定である。

教育総合データベースの構築については、今回の限られた事業期間で完遂するプロジェクトではなく、中長期的視点が重要である。諸外国の取組でも構築～軌道に乗るまでに10年程度を要しており、すぐ目に見える結果だけに飛びつくことなく、「急がば回れ」の精神で取り組む必要があると考える。直近では、データのサンプルサイ

ズが大きくなることで機械学習の性能の向上が見込まれるので、サンプルサイズの拡大の検討及びより複雑な分析手法の検討を行っていく。また、今回のデータ分析で得られた結果を現場職員に結果を伝えていく際の方法も検討していく。

データベースという「器」が目的ではなく、目指すのはプッシュ型（アウトリーチ型）の支援に活用することである。そのためには、実際に支援に当たる者が、データの意味や分析結果を専門的な知見を持たずとも理解できるためのUI（ダッシュボードやアラート機能）が必要である。また同時に、データベースの取組以前から現場においてデータ利活用の文化と実践が根付いている必要もあると考えており、前述の学校カルテの取組のほか、現場の教職員としてデータ利活用を実践している者に、教育データ利活用アンバサダーを委嘱し、研修でのサポートをいただいているところである。

(別添資料)

令和3年度戸田市実証事業  
不登校予測分析で用いたモデルの構築方法  
及びその精度検証プロセスについて

2023/03/03  
半熟仮想株式会社

# 目次

1 概要	2
2 モデル性能検証プロセス	2
(1) 実験データの作成	2
(2) 予測性能評価の反復	3
(3) 評価値の集計	4
3 予測モデルの詳細	4
(1) 欠席数予測モデル	4
(2) 長欠確率予測モデル	5
4 特徴量の詳細	6
5 要約と今後の指針	8

# 1 概要

本分析では、多変量解析や機械学習を含むさまざまな予測手法を用いて、戸田市の小学生の不登校予測を行った。また、さまざまな予測手法の精度を網羅的に比較することで、予測の精度をできるだけ高めることを目指した。

具体的な分析プロセスは以下の2つに分けられる。

- モデル構築：月初に入手可能なデータを用いて当月に不登校となる可能性の高い生徒を特定することを目指す
- モデル検証：各モデルについて、毎月そのモデルが算出する不登校可能性に基づいて対応を行った場合、どれだけの不登校生徒へアプローチできるかの性能を評価する

なおここで言うモデルとは、あるひとつの予測手法およびそれぞれの手法に必要な詳細設定の組み合わせを意味し、全体では約 1000 種類のモデルの予測精度を検証した（モデルの例：Ridge 回帰を採用し、その正則化パラメータを 1 と設定するモデル、LightGBM を採用し、その学習回数を 100 と設定するモデル）。

本資料の構成は以下の通りである。はじめに各モデルの予測性能の評価方法の説明を行う。次いで評価を行った予測モデル群の詳細説明、最後にモデルに入力した特徴量の説明を行う。

## 2 モデル性能検証プロセス

分析では、当月に新たに長欠となる「新規長欠生徒」を予測の対象として設定し、年内に長欠を 1 度以上経験している生徒については分析対象からは除外した（利用可能データが 1 年分に限られていたため、年度をまたいだ長欠経験については考慮していない）。「新規長欠生徒」は以下の基準を満たす生徒のこととする：

- 年度初めから当月初日までに一度も長欠を経験していないこと
- 当月初日までの総欠席数が 30 より少ないこと
- 当月に長欠とみなされること

モデル性能の具体的な検証プロセスは、(1) 実験データの作成、(2) 予測性能評価の反復、(3) 評価値の集計に分けられる。本分析においては月ごとの予測を行ったため、この 3 つの手順をすべての分析対象月に繰り返すことで各モデルの性能を検証した。以下では各段階の詳細を説明する。

### (1) 実験データの作成

実験データの準備段階では、以下のデータを月ごとに作成した。

- テストデータ（モデルの予測性能を評価するためのデータ）：  
当月の初めに入手できる生徒情報を特徴量 ( $X_{test} \in R^{N \times d}$ ) とし、当月に新たに長欠となった生徒か否かをラベル ( $Y_{test} \in R^N$ ) とする。
- 学習データ（モデル構築に使用するデータ）：  
年度はじめから当月の前月までの全ての月について、テストデータと同様に特

微量とラベルを作成し、それら全てを結合したデータをそれぞれ、特徴量 ( $X_{train} \in R^{N' \times d}$ )、ラベル ( $Y_{train} \in R^{N'}$ ) とする。

ここで、 $N$ は全生徒数、 $N'$ は全生徒数×年度はじめから前月までの月数、 $d$ は特徴量の種類数を表す（使用した特徴量の詳細については、4章に記載）。また、8・9月はコロナ影響のノイズにより除外したため、10月の1ヶ月前は7月とみなす。

## (2) 予測性能評価の反復

精度評価の段階では、a データの抽出、b 学習、c 予測、d 評価のステップ、を繰り返すことで各月のモデルの性能を検証した。単純に一度のみ性能を評価するだけでなく、データをサンプリングして複数回評価を行った理由は、性能の不確実性やばらつきもあわせて評価するためである。また、a~dの反復回数においては、現実的な時間で実行できる最大限の数として、10回を採用した。

### a. データ抽出

訓練データ ( $X_{train}, Y_{train}$ ) からランダムに90%復元なしの抽出をおこなった擬似訓練データ ( $X'_{train}, Y'_{train}$ ) を作成する。

(不確実性を評価する一般的な方法であるブートストラップ法では通常100%の復元あり抽出を行うが、今回は学習アルゴリズムの都合上、重複データの存在が有害となる可能性が高いため、この手順をとった。)

### b. 学習

( $X'_{train}, Y'_{train}$ ) を評価対象のモデルに入力し、以下の計算手順で予測器  $f: R^{N \times d} \rightarrow R^N$  を取得する。

- ① 訓練データ ( $X'_{train}, Y'_{train}$ ) を Stratified K-Fold(\*1)を用いて ( $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ ) に3分割する。
- ② ( $X_1, Y_1), (X_2, Y_2)$  を評価対象のモデルに入力し学習。予測モデル  $f_{1,2}$  を作成する (ただし、early\_stopping(\*2)に対応したモデルの場合、( $X_3, Y_3$ ) を検証データとする)。
- ③ ( $X_1, Y_1), (X_3, Y_3)$  でも同様に予測モデル  $f_{1,3}$  を作成。
- ④ ( $X_2, Y_2), (X_3, Y_3)$  でも同様に予測モデル  $f_{2,3}$  を作成。
- ⑤ モデル  $f$  を  $f_{1,2}, f_{1,3}, f_{2,3}$  の算術平均とする。

(\*1) [Stratified K-Fold](#) : [K-fold](#) (Cross Validation) を拡張したデータ分割の方法。ラベルの分布 (ここでは長欠フラグがついている生徒の数) が各 Fold で等しくなることを制約としてランダムにデータを分割する方法。これにより、偏りのある訓練データが生成され、学習が不安定になることを防ぐことができる。

(\*2) [early stopping](#) : 機械学習モデルの学習を複数回繰り返す場合に適切な回数で停止させる方法。学習の過程で参照しない検証データに対する損失が下がるまでイテレーションを続け、損失が下がらなくなった時点で学習を停止させる。これにより、学習不足・過学習を一定程度防ぐことが可能になる。テストデータと

は別に検証データを用意する必要があることから、学習データが小さくなるという欠点があるため、学習データが少量な場合は K-fold CV の形式で複数回繰り返し行った結果をアンサンブルすることが一般的である。

### c. 予測

テストデータの  $X_{test}$  を  $f$  に入力し、当月の予測値 (=長欠可能性スコア)  $Y_{pred}$  を取得する。

### d. 評価

以下の手順に従い、 $Y_{test}$ ,  $Y_{pred}$  を用いて各種評価指標を計算する。

- ①  $Y_{pred}$  を参照して長欠可能性スコアが高い上位 K 人をリストアップする。
- ② リストアップした K 人の生徒について、 $Y_{test}$  と照らし合わせ、実際にその月に新たに長欠となっていた生徒数 (=一致数) を算出する。
- ③ リストアップした生徒数 (=K) に対する一致数の割合 (%) と実際に新たに長欠になった生徒数に対する一致数の割合 (%) を算出して評価値とする。

## (3) 評価値の集計

モデル性能検証プロセスの最終段階では、(2) で算出した評価値について、以下の値を算出・集計し、モデルの性能評価とする：

- 平均的な予測性能の指標として、10 回分の評価値平均値を算出する
- 予測性能の不確実性・ばらつきの指標として 10 回分の評価値の分散を算出する

## 3 予測モデルの詳細

次いで本節では、2 節で説明した手順に従って性能検証を行ったモデルの詳細について記述する。モデルは、欠席数予測と長欠確率予測の 2 つに大きく分類され、それぞれについて、構築可能な約 500 種類のモデルを網羅的に検証することを目指した。

### (1) 欠席数予測モデル

欠席数予測モデルは、当月の欠席数を予測し、当月までの総欠席数と照らし合わせて長欠可能性スコアを算出するモデルである。長欠可能性スコアについては、各生徒について以下の 2 つの値のうち小さい方を採用した：

- a. 30 - 当月までの総欠席数 - 当月の予測欠席数
- b. 10 - 当月の予測欠席数

上記のスコア定義は「長欠と判断されるまでの残り欠席数 (の予測値)」と解釈できるため、この値が小さいほど長欠可能性が大きいと判断した。

欠席数予測モデルについて試行した予測手法と設定の組み合わせは以下の表の通りである。1つの「予測手法」に対して「設定」の欄の各項目をリストから1つ設定した組み合わせを1つのモデルとし、計485種のモデルを構築・検証した。

(ハイパーパラメータの詳細は本節末に記載した。)

表1：欠席数予測で使用したモデル一覧

予測手法	ハイパーパラメータ
OLS	<ul style="list-style-type: none"> <li>定数項：[あり, なし]</li> <li>係数非負制約：[あり, なし]</li> </ul>
Lasso	<ul style="list-style-type: none"> <li>正則化強度：[1e-2, 1e-1, 1, 10]</li> </ul>
Ridge	<ul style="list-style-type: none"> <li>正則化強度：[1e-2, 1e-1, 1, 10],</li> </ul>
Elastic Net	<ul style="list-style-type: none"> <li>正則化強度：[1e-2, 1e-1, 1, 10],</li> <li>正則化 L1 比率：[0.1, 0.3, 0.5, 0.7, 0.9]</li> </ul>
Poisson 回帰	<ul style="list-style-type: none"> <li>正則化強度：[1e-2, 1e-1, 1, 10]</li> </ul>
Tweedie 回帰	<ul style="list-style-type: none"> <li>正則化強度：[1e-2, 1e-1, 1, 10]</li> </ul>
Random Forest	<ul style="list-style-type: none"> <li>目的関数：[MSE 損失, Poisson 損失, Tweedie 損失]</li> <li>葉の枚数：[5, 10, 15, 20, 31]</li> <li>バギング頻度：[1, 5, 10] ・ バギング比率：[0.1, 0.5, 0.9]</li> </ul>
Light GBM (*)	<ul style="list-style-type: none"> <li>損失関数：[MSE 損失, Poisson 損失, Tweedie 損失]</li> <li>葉の枚数：[5, 10, 15, 20, 31]</li> <li>バギング頻度：[1, 5, 10]</li> <li>バギング比率：[0.1, 0.5, 0.9]</li> <li>ブースティング形式：[GBDT, dart]</li> </ul>

(\*) [LightGBM](#)：勾配ブースティング木 (Gradient Boosting Decision Tree, GBDT) の一種。一般的な線形モデルと比較して表現力・予測性能が高く、かつ計算も高速であるために多くの実務者から好まれる。ビジネスの現場や Kaggle などのコンペティションにおいても広く用いられ、構造化データを用いた機械学習手法のデファクトスタンダードとなっている。(参考：[「初手LightGBM」をする7つの理由](#))

## (2) 長欠確率予測モデル

長欠確率予測では、当月に新たに長欠となるか否かを予測し、その確率の大きさをリスクスコアとして扱った。そのため、長欠可能性リスクスコアが大きいほど、長欠になる可能性が高いと解釈できる。

試行した予測手法と設定の組み合わせは以下の表の通りである。欠席数予測モデルと同様に、1つの「予測手法」に対して「設定」の欄の各項目をリストから1つ設定した組み合わせを1つのモデルとし、計521種のモデルを構築・検証した。

表2：長欠確率予測で使用したモデル一覧

予測手法	ハイパーパラメータ
Logistic Regression	<ul style="list-style-type: none"> <li>定数項：[あり, なし]</li> <li>正則化強度：[1e-10, 1, 10]</li> <li>正則化 L1 比率：[0, 0.1, 0.3, 0.5, 0.7, 0.9, 1]</li> <li>ラベル不均衡補正：[あり, なし]</li> </ul>

Ridge Classifier	<ul style="list-style-type: none"> <li>定数項：[あり, なし]</li> <li>正則化強度：[1e-10, 1, 10]</li> <li>ラベル不均衡補正：[あり, なし]</li> </ul>
Random Forest	<ul style="list-style-type: none"> <li>目的関数：["cross_entropy", "MSE"]</li> <li>葉の枚数：[5, 10, 15, 20, 31]</li> <li>バギング頻度：[1, 5, 10]</li> <li>バギング比率：[0.1, 0.5, 0.9]</li> </ul>
Light GBM	<ul style="list-style-type: none"> <li>目的関数：["cross_entropy", "MSE"]</li> <li>葉の枚数：[5, 10, 15, 20, 31]</li> <li>バギング頻度：[1, 5, 10]</li> <li>バギング比率：[0.1, 0.5, 0.9]</li> <li>ブースティング形式：[gbdt, dart]</li> </ul>

(参考)ハイパーパラメータの詳細

- 線形
  - 定数項：線形回帰に定数項を設ける
  - 係数非負制約：線形回帰の係数が必ず非負になる
  - 正則化強度：L1, L2 正則化の強さを示すパラメータ
  - 正則化 L1 比率：Elastic Net における正則化のうち、L1 が占める割合
  - ラベル不均衡補正：正例（長欠）と負例（長欠でない）の数に偏りがあるとき、それを補正する方法
- 決定木
  - 目的関数：決定木が最小化する目的関数の設定。MSE は通常の二乗誤差、Poisson はカウントデータに対する誤差、Tweedie はゼロ過剰カウントデータに対する誤差
  - 葉の枚数：決定木が作り出す葉の枚数
  - バギング頻度：何イテレーションごとにバギング（データの一部をサンプリングして損失を計算する）を行うか
  - バギング比率：バギングにおいてデータ全体の何割をサンプリングするか
  - ブースティング形式：ブースティング（逐次的に決定木を作成していく方法）の設定。GBDT はデフォルトの設定、dart はニューラルネットの dropout の発想を適用して過学習防止を狙う方法

## 4 特徴量の詳細

モデルの構築にあたり、各生徒の特徴量として様々なデータを使用した。それら特徴量についての説明一覧は以下の表のとおりである。なお、本分析においては、モデルの構築時に特徴量を人手で選定することは行わず、現状入手可能なデータをできる限り多く使用する方針で選択したが、今後より高度なモデルを作るためには、特徴量の取捨選択や新たな特徴量を作成することも検討する必要がある。

表3：モデルで使用した特徴量一覧

データ区分	特徴量名	特徴量定義	数値例
欠席データ	総欠席数	年度内から全月末までの欠席回数	12
	前月病欠率	前月の病欠数 / 登校数	0.1
	前月遅刻率	前月の遅刻数 / 登校数	0.1
	前月早退率	前月の早退数 / 登校数	0.1
	前月事故欠率	前月の事故欠数 / 登校数	0.1
	前月出席停止率	前月の出席停止数 / 登校数	0.1
	前月忌引き率	前月の忌引き数 / 登校数	0.1
保健室データ	前月保健室利用回数	前月の保健室利用回数(内科)	3
学力データ	国語学力	5月に実施される国語のテストの点数	30
	算数学力	5月に実施される算数のテストの点数	30
健診データ	身長	年度はじめに計測した身長(を正規化した値)	5.1
	肥満度	年度はじめに計測した肥満度	-1.1
授業がわかる調査データ (理解度)	国語理解	「授業(じゅぎょう)の内容がわかりますか」というアンケートに対する回答(5段階)	4
	算数理解	同上	4
	社会理解	同上	4
	理科理解	同上	4
授業がわかる調査データ (楽しさ)	国語楽しさ	「授業(じゅぎょう)が楽しいですか。」というアンケートに対する回答(5段階)	4
	算数楽しさ	同上	4
	社会楽しさ	同上	4
	理科楽しさ	同上	4
	英語楽しさ	同上	4
	音楽楽しさ	同上	4
	図工楽しさ	同上	4
	家庭楽しさ	同上	4
体育楽しさ	同上	4	
質問紙調査データ	興味・関心	「興味や関心をもったことについて、自分なりに調べて考えを深めていますか」というアンケートに対する回答(5段階)	4
	改善	「授業をきっかけに、学校や地域や社会をよくするために改善した方がよいことを考えたり、思いついたりすることはありますか」というアンケートに対する回答(5段階)	4
	意見	「授業で、友達と意見を伝えあったり見合ったりするなかで、自分の考えを深めたり広げたりすることができていると思いますか」というアンケートに対する回答(5段階)	4
	友人関係	「学校の友達は自分の良いところを認めてくれましたか」というアンケートに対する回答(5段階)	4
	先生	「先生はわかるところまで教えてくれたか」というアンケートに対する回答(5段階)	4

	質問できる	「わからないところを質問しやすい雰囲気か」というアンケートに対する回答（5段階）	4
	授業スピード	「授業のスピードはどうだったか」というアンケートに対する回答（5段階）	4
	家庭の会話	「家の人に学校でのことを話すか」というアンケートに対する回答（5段階）	4
	登下校時間	「通学班/登下校の時刻を守れるか」というアンケートに対する回答（5段階）	4
	授業時間	「授業の始まる時間を守れるか」というアンケートに対する回答（5段階）	4
	テレビゲーム使用時間	「1日にどの程度テレビゲームをするか」というアンケートに対する回答	6
	スマホ使用時間	「1日にどの程度スマートフォンを使うか」というアンケートに対する回答	6
その他	転校生フラグ	転校生か否か	1

## 5 要約と今後の指針

本資料では、令和3年度戸田市小学校のデータを用いた不登校予測分析の、手法及び使用特徴量に関する説明を行った。予測では、欠席数予測と長欠率予測の主に2つのパターンを用いてアプローチを行い、パラメータなどの詳細設定を変えることで、約1000種類のモデルを構築・検証した。また精度検証プロセスでは、分析に伴い起こりうる誤差を考慮し、サンプリングと検証プロセスの反復を通じて、予測の不確実性についての評価も併せて行った。

今後、モデルの精度をより高めていくためには、データ量や特徴量を充実させていくことが重要なステップとなっていく。しかし、突然大幅に真新しいデータを増やさずとも、まずは今あるデータの年度を2年度分、3年度分と拡張していくことで、より詳細な経年分析を行ったり、環境や季節性による傾向を見つけてモデルに組み込んだりすることも可能になる。加えて、中学校では小学校よりも不登校の頻度が高いため、小学校と同じ粒度の出欠データを整備することで不登校予測分析を中学校にも拡大することができれば、より不登校要因への理解が進むことが期待される。

データを用いた不登校予測は社会的にも新しい試みであり、実用のためには、学校現場とも連携した細かな試行と改善を長期的に行っていくことが不可欠である。その初期段階として、モデル構築のためのデータ準備方法や、網羅的な手法比較の考え方の検証を行った本分析は、今後より複雑で広範囲な分析を行う上での重要なステップになることが期待される。