

空き家推定システム LINKS SOMA

システム設計書

2026年3月 国土交通省 総合政策局 情報政策課

目次

1. 開発スコープ	5
1-1. 概要	5
1-2. システムを利用する業務全体像とシステム利用フロー	7
1-2-1. 業務フロー	7
1-2-2. システムシーケンス図	8
2. 開発するシステム：機能要件（FN/SL/AL/CO/HW/IF/SC）	9
2-1. システム機能(FN)	9
2-1-1. アーキテクチャ	9
2-1-2. システム機能一覧	10
2-1-3. システム機能の詳細	13
2-1-4. ソフトウェア・ライブラリ（SL）	26
2-1-5. 数理モデル・アルゴリズム（AL）の詳細	28
2-2. システムコンポーネント（CO）	39
2-2-1. システムコンポーネント図	39
2-2-2. システムコンポーネント一覧	40
2-3. ハードウェアアーキテクチャ(HW)	42
2-3-1. ハードウェアアーキテクチャ	42
2-3-2. ハードウェアの詳細	43
2-4. データインターフェース(IF)	45
2-4-1. データアーキテクチャ図	45
2-4-2. データインターフェース一覧	47
2-4-3. データインターフェースの詳細	48

- 2-5. ユーザーインターフェース(SC) 63
 - 2-5-1. ユーザーインターフェース一覧 63
 - 2-5-2. 画面遷移図 67
 - 2-5-3. ユーザーインターフェースの詳細 69
- 3. 開発するシステム：非機能要件（NF） 94
 - 3-1. 非機能要件一覧 94
 - 3-1-1. LINKS SOMA 94
 - 3-1-2. LINKS SOMA CityGML Converter 95
 - 3-1-3. ジオコーディングツール 95
 - 3-2. 非機能要件詳細 96
 - 3-2-1. 【NF001】 実行環境 96
 - 3-2-2. 【NF002】 OS 96
 - 3-2-3. 【NF003】 マシンスペック 96
 - 3-2-4. 【NF007】 処理速度 96
 - 3-2-5. 【NF201】 実行環境 96
- 4. 実証調査に利用するデータ（DT） 98
 - 4-1. 実証調査に利用するデータ一覧 98
 - 4-2. 実証調査に利用するデータの詳細 100
 - 4-2-1. 【DT001】 水道栓データ 100
 - 4-2-2. 【DT002】 水道使用量データ 100
 - 4-2-3. 【DT003】 住民基本台帳 101
 - 4-2-4. 【DT004】 登記情報 101
 - 4-2-5. 【DT005】 推定対象選定用データ 102

4-2-6. 【DT006】 国勢調査小地域データ	103
4-2-7. 【DT007】 地域集計用ポリゴン	103
4-2-8. 【DT008】 空き家調査結果.....	103
4-2-9. 【DT101】 建物ポリゴンデータ	104
4-2-10. 【DT102】 ジオコーディング用データ	106
4-2-11. 【DT103】 ジオコーディング済データ	106
4-2-12. 【DT201】 正規化水道栓データ	106
4-2-13. 【DT202】 正規化水道使用量データ	107
4-2-14. 【DT203】 正規化住民基本台帳.....	107
4-2-15. 【DT204】 正規化登記情報	107
4-2-16. 【DT205】 正規化推定対象選定用データ.....	107
4-2-17. 【DT206】 正規化ジオコーディング済データ	107
4-2-18. 【DT207】 正規化空き家調査結果.....	107
4-2-19. 【DT301】 住居単位水道データ	108
4-2-20. 【DT302】 住居単位住民基本台帳.....	110
4-2-21. 【DT303】 住居単位登記情報.....	112
4-2-22. 【DT401】 テキストマッチングデータ	113
4-2-23. 【DT402】 家屋単位 GIS データ	114
4-2-24. 【DT403】 家屋種別判定後データ	116
4-2-25. 【DT501】 建物単位空き家推定結果データ.....	116
4-2-26. 【DT502】 地域単位空き家推定結果データ.....	117
用語集.....	119

1. 開発スコープ

1-1. 概要

令和6年度に引き続き、国土交通省の分野横断的なデータ化およびオープンデータ化を推進し、官民の多様な分野における活用事例（ユースケース）の開発を進める新たな取組「Project LINKS」を展開する一環として、空き家問題をテーマとした EBPM ユースケースを開発する。

開発するシステムは以下の3点である。

1. LINKS SOMA

- 自治体が保有するデータを活用し、AI を用いて建物単位で空き家を推定できるシステム

2. LINKS SOMA CityGML Converter

- PLATEAU の建物モデルを LINKS SOMA で利用する形式に変換するためのツール

3. ジオコーディングツール

- インターネットに接続可能な環境で動作する住所の一覧に緯度・経度を付与することができるツール

LINKS SOMA は、自治体が保有するオープンデータと非公開データを組み合わせ、高精度な空き家推定をノーコードかつわかりやすい UI/UX で実現し、自治体職員が手軽に地域内の空き家を推定・把握し、現状分析を行えることを目的に開発したデスクトップアプリケーションである。本アプリケーションでは、個人情報を含むデータを扱うことを前提としており、自治体のネットワークモデル「三層分離」における「LGWAN 接続系」環境下での起動・実行可能である必要があった。そこで、完全オフライン環境でも必要な処理を実行できるよう、デスクトップアプリケーションの開発には OSS フレームワークである Electron を採用した。

また、本アプリケーションにおけるデータ処理および機械学習処理は Python で実装後、実行可能ファイルとしてビルドし、Electron から呼び出す構成とした。この構成により、UI/UX は React・TypeScript で構築しつつ、データ処理および機械学習ロジックは Python で実装するという役割分担を実現している。これにより、操作性に優れたインターフェースと高精度な推定処理を両立している。

また、空き家推定は、あらかじめ空き家の特徴を学習した機械学習モデルをシステムに内包することで実現している。さらに、ユーザー自身が保有するデータに基づき独自のモデルを構築・適用できるよう、モデル学習機能も備えている。

LINKS SOMA CityGML Converter は、LINKS SOMA において PLATEAU の建築物ポリゴンデータ、都市計画決定情報、土地利用データを活用できるようにするための変換ツールである。本アプリケーションの開発には、すでに公開されている「PLATEAU GIS Converter」を活用した。本アプリケーションでは、CityGML に触れたことがない自治体職員でも簡単に操作できるよう、G 空間情報センターにおいて公開されている PLATEAU

データを、元データの ZIP ファイルの状態から解凍せずにファイル読みを行い、GeoPackage 形式で出力する機能を実装した。また、PLATEAU データには標高データをはじめ大容量のファイルが含まれるため、解凍すると大幅にパソコンの容量を消費してしまう。本対応により、自治体職員が PLATEAU データをより使いやすくなるよう開発した。

ジオコーディングツールは、自治体が「住所と緯度・経度がセットになったデータ」を保有していない場合に利用できるアプリケーションとして開発した。本アプリケーションも LINKS SOMA と同様に Electron を用いたデスクトップアプリケーションであり、React および TypeScript をベースに構築されている。ただし、オンライン環境での動作を前提としており、外部のジオコーディング API 等と連携することで、自治体が住所データから緯度・経度情報を付与できる仕組みを実装している。

1-2. システムを利用する業務全体像とシステム利用フロー

1-2-1. 業務フロー

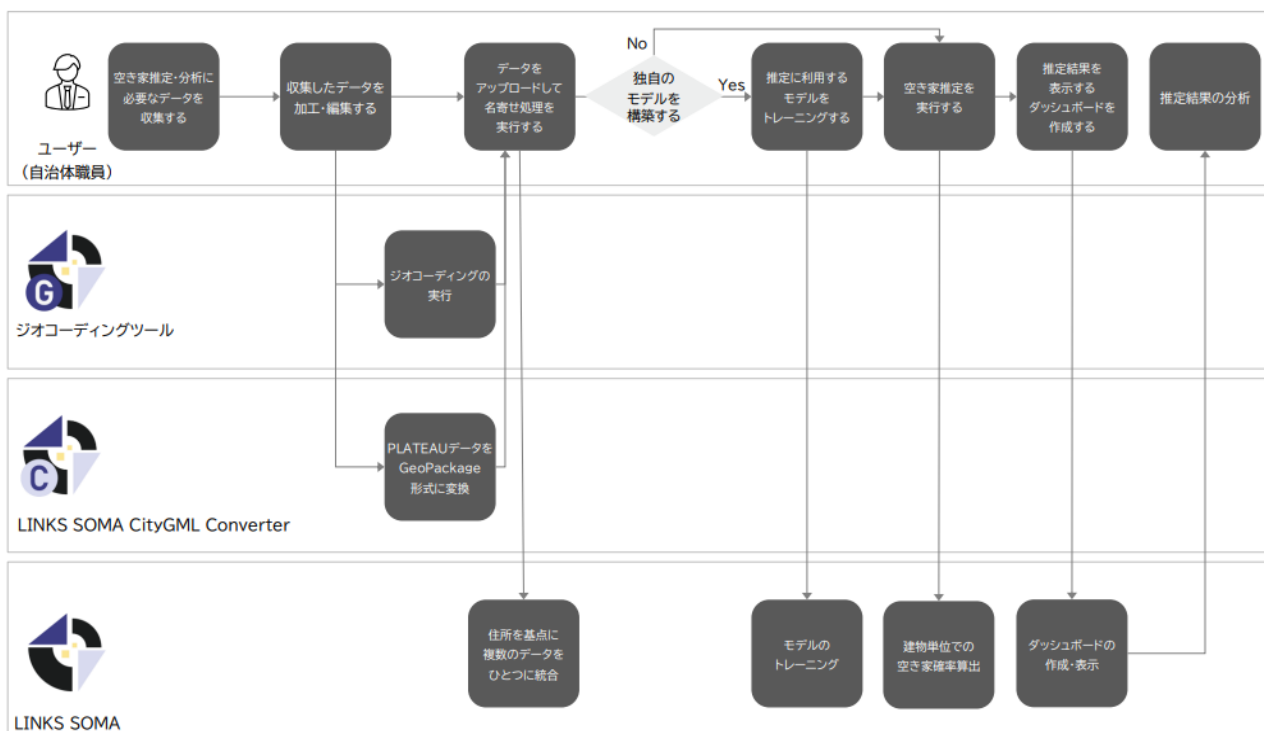


図 1-2-1 業務フロー

1-2-2. システムシーケンス図

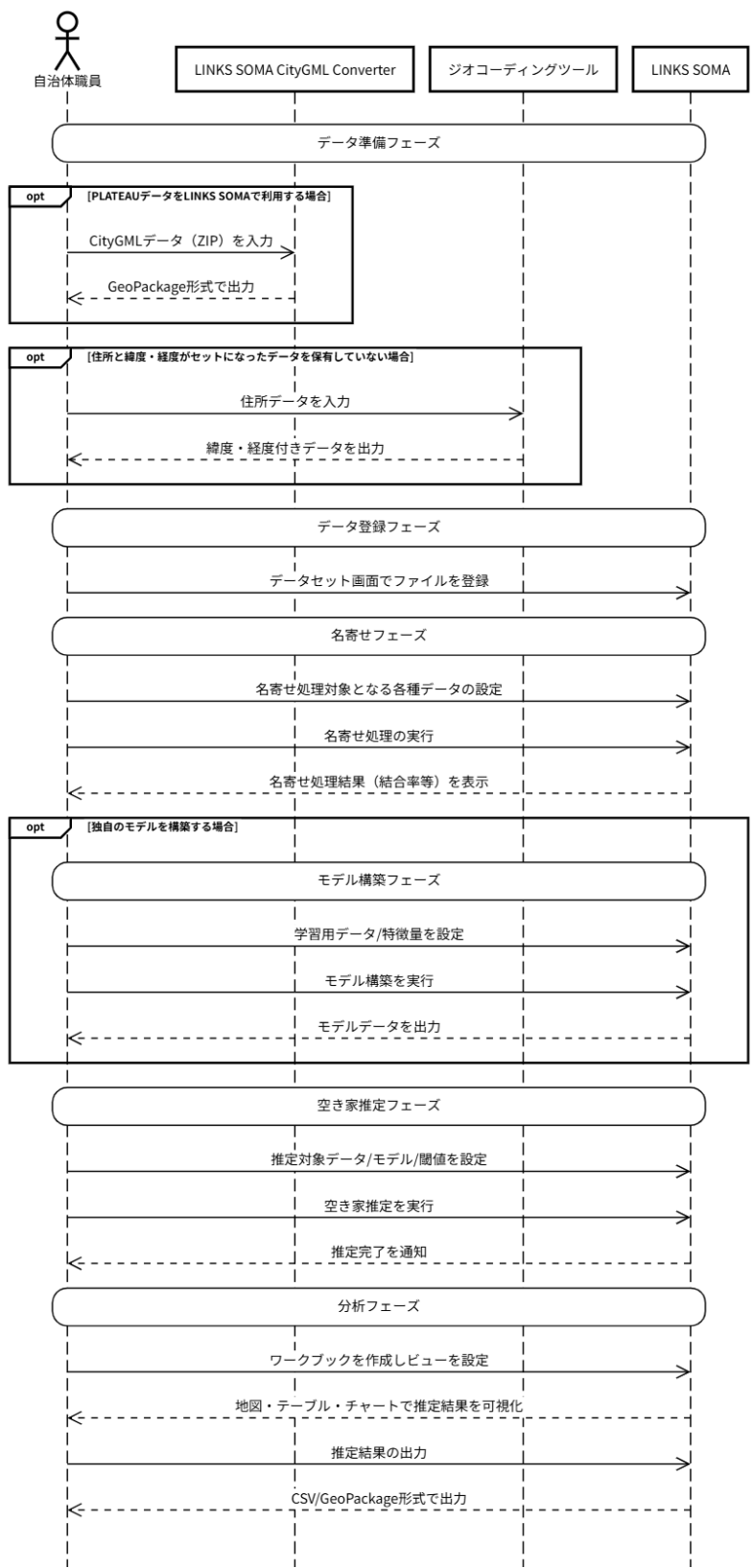


図 1-2-2 システムシーケンス図

2. 開発するシステム: 機能要件 (FN/SL/AL/CO/HW/IF/SC)

2-1. システム機能(FN)

2-1-1. アーキテクチャ

LINKS SOMA は、名寄せシステム、空き家推定 AI システム、空き家推定 BI ツールの 3 つのサブシステムと、システム共通機能から構成される。(各機能の詳細は 4-2 節を参照) LINKS SOMA CityGML Converter、ジオコーディングツールは、サブシステムを持たずそれぞれが独立した単一機能のツールとして提供される。

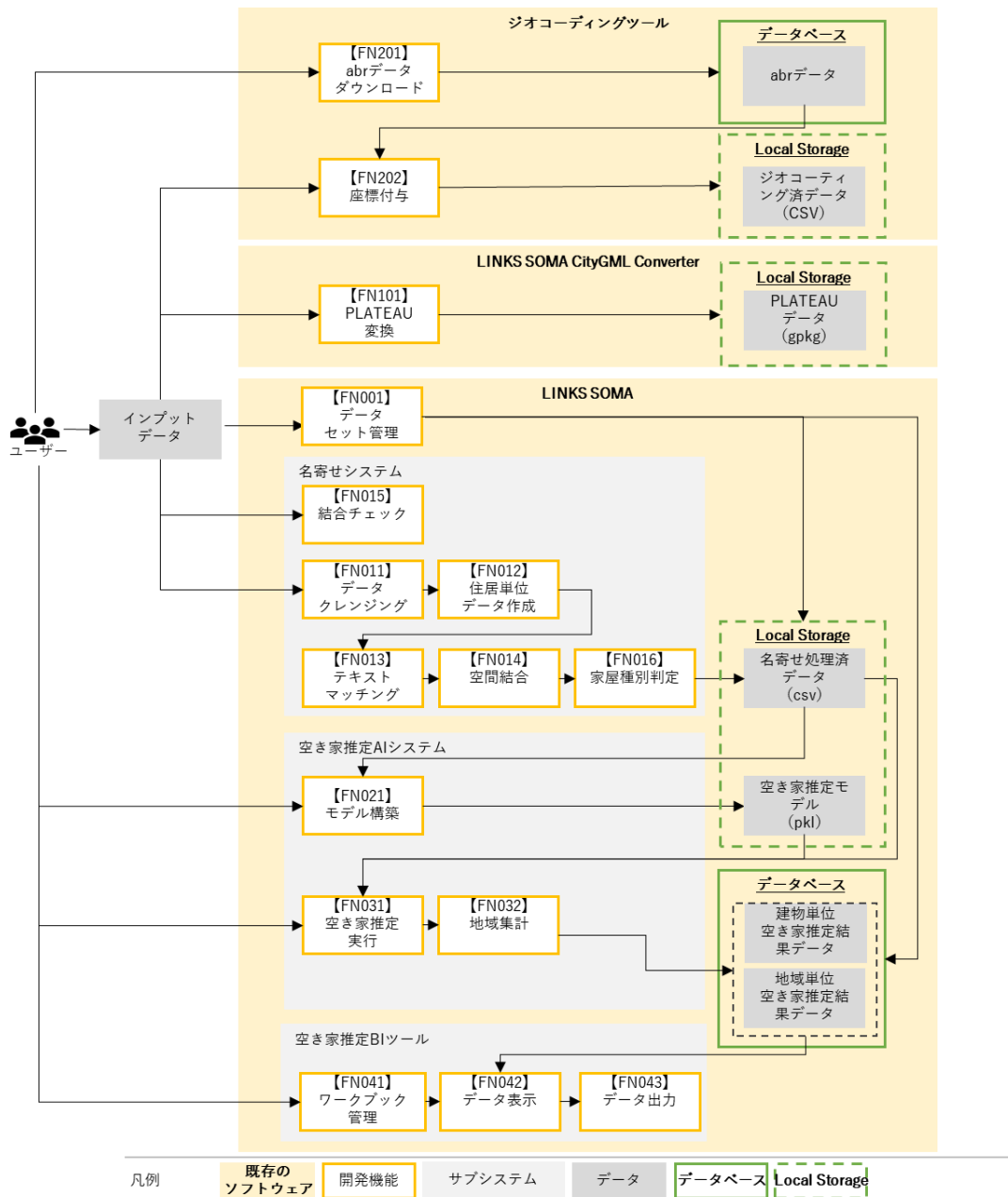


図 2-1-1-1 システムアーキテクチャ

2-1-2. システム機能一覧

開発機能の一覧を表 2-1-2-1 に示す。なお、本業務において開発（新規・改修）を行うシステム機能を朱文字で示す。

表 2-1-2-1 システム機能一覧

システム	ID	機能名	機能説明	区分
LINKS SOMA	FN001	データセット管理	● インputデータ・名寄せ済データ・ 推定結果データ を一覧管理する GUI を備える。	既存改修
			● データセットのアップロード・ダウンロード・編集・削除の操作を提供する。	
			● 管理しているデータセットの中身が確認できるよう、プレビュー機能を提供する。	
LINKS SOMA	FN011	データクレンジング	● 各データソースの住所カラムを正規化し、表記揺れを統一する。 (正規化のロジックを改修)	既存改修
			● 各カラムの値が指定型以外の場合、欠損値に変換する。	
			● 日付情報を西暦形式に統一する。	
LINKS SOMA	FN012	住居単位データ作成	● 水道栓・水道使用量・住民基本台帳・登記簿のデータを住居単位で再集計し、AI モデルの学習時、及び空き家推定時に利用する特徴量を生成する。 (特徴量生成のロジックを改修)	既存改修
LINKS SOMA	FN013	テキストマッチング	● 正規化済み住所に基づき、複数データソースを結合 (left join) する。	変更なし
			● 正規化済み住所が完全一致している場合に結合を実行する。	
LINKS SOMA	FN014	空間結合	● 建物ポリゴンに対して交差結合を実行し、家屋単位の GIS データを生成する。	変更なし
			● 座標系を WGS84 に統一する。	
LINKS SOMA	FN015	結合チェック	● 名寄せ処理の事前チェックとして、各データ間の住所の表記揺れを確認する。	新規開発
LINKS SOMA	FN016	家屋種別判定	● 空き家推定の対象を家屋の種別から選定する。	新規開発

LINKS SOMA	FN021	モデル構築	<ul style="list-style-type: none"> ● LightGBM を用いて AI モデルを構築する。 ● モデルの学習に Positive Unlabeled Bagging を利用する 	既存改修
LINKS SOMA	FN031	空き家推定実行	<ul style="list-style-type: none"> ● 学習済みモデルを用いて建物単位の空き家確率を推定する。 	変更なし
LINKS SOMA	FN032	地域集計	<ul style="list-style-type: none"> ● 建物単位の空き家推定結果を指定の地域区分（小地域単位等）で再集計した結果を出力する。 	変更なし
LINKS SOMA	FN041	ワークブック管理	<ul style="list-style-type: none"> ● 推定結果の分析用ワークブックを作成・編集・削除する。 	変更なし
LINKS SOMA	FN042	データ表示	<ul style="list-style-type: none"> ● 推定結果を地図・テーブル・チャート（円グラフ・棒グラフ・折れ線グラフ）で可視化する。 ● 緯度経度が重複しているデータの検知・表示を行う。 ● 地理座標を有するデータセットについては、テーブルの該当レコードの選択または地図上の該当箇所の選択により、地図表示機能と連動したテーブル表示を行う ● 地図表示時、建物ポリゴンと空間結合できている場合にはポリゴン単位で、結合できていない場合はポイントで表示する。 	既存改修
LINKS SOMA	FN043	データ出力	<ul style="list-style-type: none"> ● 推定結果をユーザー指定の項目・形式（CSV/GeoPackage/GeoJSON）・座標系・閾値で出力する 	既存改修
LINKS SOMA CityGML Converter	FN101	PLATEAU 変換	<ul style="list-style-type: none"> ● LOD0（平面的な外形表現）及び LOD1（建物図形に高さを与えて表現する形式）の PLATEAU データを、CityGML 形式データから GeoPackage 形式に変換する。 	既存改修
ジオコーディング ツール	FN201	abr データダウンロード	<ul style="list-style-type: none"> ● ジオコーディング処理で使用するためのデータとして、ユーザーが指定した市区町村のアドレス・ベース・レジストリ（abr）データをダウンロードサイトから取得する。 	新規開発

ジオコーディングツール	FN202	座標付与	<ul style="list-style-type: none">住所データに、外部ジオコーディングサービス（【IF301】）、または abr Geocoder を利用して緯度・経度を付与する。	既存改修
-------------	-------	------	---	------

2-1-3. システム機能の詳細

2-1-3-1. 【FN001】データセット管理<既存改修>

本システム機能の概要

本機能は、LINKS SOMA で扱う以下データセットのアップロード・閲覧・削除操作を提供する。

- インプットデータ：空き家推定に利用するファイル
- 名寄せ済処理済データ：名寄せ処理（FN011～FN014）の出力結果
- 空き家推定結果データ：空き家推定処理（FN031、FN032）の出力結果

本システム機能の処理の詳細

- 処理内容
 - ユーザーが選択したファイルを受信し、UUID によるファイル名を生成してストレージに保存する。
 - SQLite データベースにメタデータ（ファイル名・サイズ・種別・登録日時）を記録する。
- 利用するライブラリ
 - 【SL001】 Electron（IPC によるファイル受信）
 - 【SL008】 SQLite3 + Drizzle ORM（DB 操作）

本システム機能の入出力データの仕様

- 入力
 - データの内容：ユーザーがアップロードした各種ファイル（水道情報、住民基本台帳、登記簿等）
 - データの形式：CSV、Shapefile 等の静的ファイル
- 出力
 - データの内容：ファイルメタデータ（名称・種別・パス・登録日時等）
 - データの形式：SQLite テーブル

2-1-3-2. 【FN011】データクレンジング<既存改修>

本システム機能の概要

本機能は、名寄せ処理でデータを正しく結合できるようにするために、住所データの正規化を行う。

本システム機能の処理の詳細

「2-1-5-2. 開発したアルゴリズム」の「【AL001】住所正規化」を参照

本システム機能の入出力データの仕様

- 入力
 - データの内容：ユーザーが入力した住所カラムを持つテキストデータ（DT001、DT003～DT005、DT102）
 - データの形式：CSV（UTF-8 / Shift_JIS）
 - 利用するデータインターフェース：【IF001】名寄せ処理インターフェース
- 出力
 - データの内容：入力ファイルに「正規化住所」カラムを追加したテキストデータ（DT201～DT206）
 - データの形式：CSV（UTF-8 BOM 付き）

2-1-3-3. 【FN012】住居単位データ作成<既存改修>

本システム機能の概要

水道・住民基本台帳・登記の各データをユニークな住居（住所）単位に再集計し、後続のテキストマッチングで結合可能な形式に整備、および、【FN031】空き家推定実行に用いる特徴量を作成する。

本システム機能の処理の詳細

「2-1-5-2. 開発したアルゴリズム」の「【AL002】住居単位集約」を参照

本システム機能の入出力データの仕様

- 入力
 - データの内容：ユーザーがアップロードした水道使用量データ（DT102）、および住所正規化済みの水道・住民基本台帳・登記データ（DT201～DT203）
 - データの形式：CSV（UTF-8 BOM 付き）
 - 利用するデータインターフェース：【IF001】名寄せ処理インターフェース
- 出力
 - データの内容：住居単位で集約された各種データ（DT301～DT303）
 - データの形式：CSV（UTF-8 BOM 付き）

2-1-3-4. 【FN013】テキストマッチング

本システム機能の概要

正規化済み住所をキーとして複数のデータソースを結合し、住居単位の統合データを生成する。これにより、後続の空き家推定（【FN031】）に必要な特徴量を1つのレコードにまとめる。

本システム機能の処理の詳細

- 結合処理
 - 【DT301】 住居単位水道データの正規化住所をキーに各種データを左外部結合 (left join) する

本システム機能の入出力データの仕様

- 入力
 - データの内容：FN011,および FN012 の出力ファイル
 - データの形式：CSV (UTF-8 BOM 付き)
 - 利用するデータインターフェース：【IF001】 名寄せ処理インターフェース
- 出力
 - データの内容：FN011,および FN012 の全出力ファイルを統合したデータ (DT401)
 - データの形式：CSV (UTF-8 BOM 付き)

2-1-3-5. 【FN014】 空間結合

本システム機能の概要

本機能は、テキストマッチングデータ (DT401) に対し、建物ポリゴンデータおよび国勢調査メッシュとの空間結合を実施する。

本システム機能の処理の詳細

「2-1-5-2. 開発したアルゴリズム」の「【AL005】 空間結合 (交差結合)」 「【AL010】 平面直角座標系選定」を参照

本システム機能の入出力データの仕様

- 入力
 - テキストマッチング済データ
 - データの内容：FN013 の出力データ
 - データの形式：CSV (UTF-8 BOM 付き)
 - 建物ポリゴンデータ
 - データの内容：ユーザーが入力した建物ポリゴンデータ (DT101)
 - データの形式：GeoPackage / Shapefile / CSV (WKT 形式)
 - 国勢調査メッシュ
 - データの内容：ユーザーが入力した建物ポリゴンデータ (DT006)
 - データの形式：GeoPackage
 - 利用するデータインターフェース：【IF001】 名寄せ処理インターフェース
- 出力
 - データの内容：空間情報が付与された統合データ (DT402)

- データの形式：CSV（UTF-8 BOM 付き）

2-1-3-6. 【FN015】結合チェック<新規開発>

本システム機能の概要

本機能は、名寄せ処理（FN011～FN014）の実行前に、ユーザーがアップロードしたデータ品質の問題を把握できるように、各データソース間の住所の表記揺れチェック機能を提供する。

本システム機能の処理の詳細

- 住所正規化
 - 処理内容
 - 水道データ、および結合対象データの住所カラムに対し、【FN011】と同じ正規化処理を適用する
 - 利用するライブラリ
 - 【SL012】Pandas
 - 利用するアルゴリズム
 - 【AL001】住所正規化
- 結合
 - 処理内容
 - 正規化済み住所を町字名でグルーピングする
 - 結合対象の町字名のうち、水道データの町字名に存在しないものを抽出する
 - 利用するライブラリ
 - 【SL016】Polars
- 類似住所候補検索
 - 処理内容
 - 抽出した水道データに存在しない町字住所に対し、最も類似度の高い水道データの町字名をレーベンシュタイン類似度に基づき選定する
 - 類似度が0.8以上（10文字中8文字が一致する水準）1.0未満の住所を最大5件、類似住所候補として提示する
 - 利用するライブラリ
 - 【SL016】Polars
 - 利用するアルゴリズム
 - 【AL007】レーベンシュタイン距離

本システム機能の入出力データの仕様

- 入力
 - データの内容：ユーザーが入力した住所カラムを持つテキストデータ（DT001、DT003～DT005、DT102）

- データの形式：CSV (UTF-8)
- 利用するデータインターフェース：【IF005】結合チェックインターフェース
- 出力
 - データの内容：水道データに存在しない町字名、及び類似住所候補
 - データの形式：JSON

2-1-3-7. 【FN016】家屋種別判定<新規開発>

本システム機能の概要

本機能では、空き家推定の対象とするレコードを家屋種別から選定する。

本システム機能の処理の詳細

- 家屋種別が Null 値のレコードに対し、「種別不明」フラグを付与する
- 以下に該当しないレコードを削除する
 - ユーザー指定した家屋種別を持つレコード
 - 「種別不明」フラグが付与されたレコード

本システム機能の入出力データの仕様

- 入力
 - データの内容：DT401、または DT402
 - データの形式：CSV (UTF-8 BOM 付)
 - 利用するデータインターフェース：IF001
- 出力
 - データの内容：家屋種別でフィルタリング済のテキストマッチングデータ (DT403)
 - データの形式：CSV (UTF-8 BOM 付)

2-1-3-8. 【FN021】モデル構築<既存改修>

本システム機能の概要

本機能は、空き家推定に使用する学習済みモデルを構築する機能を提供する。

本システム機能の処理の詳細

- 処理内容
 - 入力データをモデルトレーニング用の学習用データと、精度検証のためのテスト用データに分割する。
 - 学習用データを用いて、モデルをトレーニングする。モデルのトレーニングには【AL101】Positive Unlabeled Bagging、【AL102】LightGBM を利用する。

- トレーニング後のモデル精度をテスト用のデータを用いて算出する。
- 利用するライブラリ
 - 【SL002】 Scikit-Learn
 - 【SL005】 GeoPandas
 - 【SL009】 Pandas
 - 【SL010】 Optun
- 利用するアルゴリズム
 - 【AL101】 Positive Unlabeled Bagging
 - 【AL102】 LightGBM

本システム機能の入出力データの仕様

- 入力
 - データの内容：FN014,または FN016 の出力結果
 - データの形式：csv
- 出力
 - データの内容：学習済みモデル
 - データの形式：pkl

2-1-3-9. 【FN031】 空き家推定実行

本システム機能の概要

本機能は、名寄せ済データセット（【FN011】～【FN014】の出力）と学習済みモデル（【FN008】で管理）を選択し、空き家推定処理を実行する。

本システム機能の処理の詳細

- 処理内容
 - 学習済み LightGBM モデルを ZIP から展開し、テキスト形式のモデルファイルを読み込む
 - 学習対象カラムの欠損フラグ (isnull_*) を生成する
 - モデルの予測 (predict) を実行し、空き家確率 (0～1 の実数値) を取得する
 - 基準閾値に従い空き家判定ラベルを付与する
- 利用するライブラリ
 - 【SL010】 LightGBM、【SL011】 scikit-learn、【SL012】 Pandas
- 利用するアルゴリズム
 - 【AL008】 LightGBM

本システム機能の入出力データの仕様

- 入力:名寄せ済データセット

- データの内容：【FN011】～【FN014】で生成された統合データ
- データの形式：CSV（UTF-8 BOM 付き）
- 入力:学習済みモデル
 - データの内容：LightGBM の学習済みモデルファイル
 - データの形式：ZIP
 - 利用するデータインターフェース：【IF003】空き家推定インターフェース
- 出力:建物単位推定結果
 - データの内容：建物単位の空き家推定結果データ（DT501）
 - データの形式：SQLite テーブル

2-1-3-10. 【FN032】地域集計

本システム機能の概要

本機能は、建物単位の空き家推定結果を指定の地域区分（小地域単位等）で再集計した結果を出力する。

本システム機能の処理の詳細

- 建物単位推定結果の座標を EPSG:4326 に変換する
- 建物の重心点が地域ポリゴンに含まれるかを空間結合（intersects）で判定する
- 地域区分ごとに以下を集計する
 - 建物総数、推定空き家数、空き家率
 - 世帯数合計、15歳未満比率、65歳以上比率
 - 各閾値（0.05～0.95）の空き家数・空き家率
- 利用するライブラリ
 - 【SL013】GeoPandas、【SL014】Shapely
- 利用するアルゴリズム
 - 【AL010】平面直角座標系選定

本システム機能の入出力データの仕様

- 入力
 - 建物単位推定結果
 - データの内容：FN031 の出力データ（DT501）
 - データの形式：SQLite テーブル
 - 地域ポリゴン
 - データの内容：地域区分のポリゴンデータ（DT007）
 - データの形式：Shapefile / GeoPackage / GeoJSON / CSV
 - 利用するデータインターフェース：【IF003】空き家推定インターフェース
- 出力
 - データの内容：地域ごとの空き家率・人口比率等の集計値（DT502）

- データの形式：SQLite テーブル

2-1-3-11. 【FN041】ワークブック管理

本システム機能の概要

本機能は、以下の空き家推定結果の分析用 UI への CRUD（作成・読取・更新・削除）操作を提供する。

- ワークブック：分析用のシートをまとめて管理するプロジェクト
- シート：複数ビューを配置して分析画面を構成するページ
- ビュー：地図やグラフ・表で推定結果を可視化する個別の領域

本システム機能の処理の詳細

- 処理内容
 - ユーザーからの入力内容に応じて、ワークブック、ビュー、シートの CRUD を行う。
- 利用するライブラリ
 - 【SL002】 React、【SL008】 SQLite3 + Drizzle ORM

本システム機能の入出力データの仕様

- 入力
 - データの内容：空き家推定処理（FN031、FN302）の出力結果
 - データの形式：SQLite テーブル
- 出力
 - データの内容：レイアウト構成・表示設定・フィルター条件
 - データの形式：SQLite テーブル

2-1-3-12. 【FN042】データ表示機能<既存改修>

本システム機能の概要

本機能は、推定結果データを BI ツールとして地図・テーブル・チャート等に表示する機能を提供する。

- 地図ビュー
 - 建物ポリゴン（またはポイント）を空き家推定確率に応じた 3 段階の色分けで表示する
 - ビューポート（画面に表示されている地図範囲）内のデータのみを段階的に読み込むことで、大量データの効率的な表示を実現する
 - 建物（またはポイント）クリックでポップアップ（詳細情報ウィンドウ）を表示する
 - 同一ジオメトリを持つデータは、地図上で識別可能な状態として表示する。
 - 地図ビューにテーブルを併設表示でき、一方の選択操作が他方に即時反映される。
 - テーブルのレコード選択により、地図上の該当箇所が表示される

- 地図上の建物（またはポイント）選択により、テーブル上の該当箇所がグレーアウトされる
- テーブルビュー
 - 推定結果を一覧表形式で表示する（50以上のカラムから表示項目を選択可能）
 - カラムヘッダクリックによるソート、ページネーション機能
- チャートビュー
 - 円グラフ（建物単位）：推定確率の分布を表示する
 - 折れ線グラフ（建物単位）：時系列推移を表示する
 - 棒グラフ（地域単位）：地域ごとの集計値を比較する
- フィルター・集計
 - テキスト型：部分一致・完全一致・前方一致等
 - 数値型：等しい・以上・以下・範囲指定
 - 日付型：指定日以降・指定日以前・範囲指定

本システム機能の処理の詳細

- 地図レンダリング
 - 処理内容
 - ビューポート内のデータを SQLite から取得し、GeoJSON 形式に変換する
 - 推定確率に応じた色分け表現を適用してポリゴン／ポイントレイヤーを描画する
 - 建物クリック時にポップアップを表示する
 - 利用するライブラリ
 - 【SL006】 MapLibre GL JS
- 重複データ検知・表示
 - 処理内容
 - 同一のジオメトリ（座標情報）を持つデータをグループ化し、重複フラグ・重複件数を付与する
 - 重複がある場合、地図上にアウトラインレイヤーとナビゲーション UI を表示する
 - 利用するライブラリ
 - 【SL006】 MapLibre GL JS
- チャート描画
 - 処理内容
 - ビュースタイルに応じて円グラフ・折れ線グラフ・棒グラフを描画する
 - 棒グラフは地域（area_group）×基準日（reference_date）単位で集計し、集計関数（平均・合計等）を適用する
 - 利用するライブラリ
 - 【SL007】 Recharts
 - フィルター処理
 - 処理内容

- フィルター条件を SQL WHERE 句に変換し、データベースクエリに適用する
- テキスト型・数値型・日付型の複合条件に対応する
- 利用するライブラリ
 - 【SL008】 SQLite3 + Drizzle ORM

本システム機能の入出力データの仕様

- 入力
 - データの内容：空き家推定処理（【FN031】）、地域集計処理（【FN032】）の出力結果
 - データの形式：SQLite テーブル
- 入力
 - データの内容：表示スタイル・カラム選択・閾値・フィルター条件
 - データの形式：SQLite テーブル
- 出力
 - なし

2-1-3-13. 【FN043】 データ出力

本システム機能の概要

本機能は、【FN031】 空き家推定及び【FN032】 地域集計による処理結果を、本システム外にて利用できるように BI ツール上で設定したフィルター条件・閾値に基づき、GIS データ形式で出力する。

本システム機能の処理の詳細

- 処理内容
 - SQLite に格納した空き家推定結果データに、ビューで設定されているフィルター条件を適用して取得する
 - ユーザーが選択した任意の座標系へ変換する
 - 英語カラム名を日本語名に変換する
 - 指定形式（CSV／GeoPackage／GeoJSON）でファイルを出力する
 - 利用するライブラリ
 - 【SL013】 GeoPandas、【SL014】 Shapely、【SL012】 Pandas
 - 利用するアルゴリズム
 - なし

本システム機能の入出力データの仕様

- 入力
 - データの内容：建物単位、または地域単位の推定結果データ（DT501,DT502）
 - データの形式：SQLite テーブル

- 利用するデータインターフェース：【IF004】データ出力インターフェース
- 出力
 - データの内容：フィルター適用済み推定結果
 - データの形式：CSV（UTF-8 BOM 付き） / GeoPackage / GeoJSON

2-1-3-14. 【FN101】 PLATEAU 変換

本システム機能の概要

本機能は、LINKS SOMA CityGML Converter が提供する機能であり、G 空間情報センターで公開されている PLATEAU の CityGML 形式データを LINKS SOMA で利用可能な GeoPackage 形式に変換する。

本システム機能の処理の詳細

- 処理内容
 - PLATEAU データの ZIP ファイルを直接読み込み、CityGML 形式を解析する
 - PLATEAU GIS Converter のコア機能を活用し、建築物 (bldg) を GeoPackage 形式に変換する。
- 利用するライブラリ
 - 【SL017】 PLATEAU-GIS-Converter
- 利用するアルゴリズム
 - なし

本システム機能の入出力データの仕様

- 入力
 - データの内容：PLATEAU データ
 - データの形式：ZIP（内部に CityGML 形式ファイルを含む）
- 出力
 - データの内容：PLATEAU データ
 - データの形式：GeoPackage (.gpkg)

2-1-3-15. 【FN201】 abr データダウンロード

本システム機能の概要

本機能は、ジオコーディングツール内で abr ジオコードを利用するために必要なアドレス・ベース・レジストリ (abr) データをダウンロードし、データベースに保存する。

本システム機能の処理の詳細

- 処理詳細
 - ユーザーが指定した市区町村のアドレス・ベース・レジストリ (abr) データをダウンロードサイトから取得する。
 - ダウンロードしたデータを SQLite DB に格納する。
- 利用するライブラリ
 - **【SL019】 abr Geocoder**

本システム機能の入出力データの仕様

- 入力:ユーザーが指定した市区町村のコード値 (システム入力)
- 出力: ユーザーが指定した市区町村の abr データ (SQLite DB)

2-1-3-16. **【FN202】 座標付与**

本システム機能の概要

本機能は、ジオコーディングツールが提供する機能であり、住所データに対して外部ジオコーディングサービスを利用し、緯度・経度情報を付与する。実行可能なジオコードは以下の3種類。

- AWS Location Service (Amazon Web Services 社が提供するジオコーディング API)
- **GEOSPACE API (NTT インフラネット社が提供するジオコーディング API)**
 - **abr ジオコード (デジタル庁が提供するジオコード)**

本システム機能の処理の詳細

- 座標付与
 - 処理内容
 - 住所データを読み込み、各住所に対してユーザーが選択したジオコードを実行し、取得した緯度・経度を元データに付与する。
 - ユーザーが「AWS Location Service」または「GEOSPACE API」を選択した場合、API 実行のためインターネット通信が発生する。詳細は「2-4-2 データインターフェース一覧」を参照すること。
 - ユーザーが「abr ジオコード」を選択した場合は、**【SL019】 abr Geocoder** を用いて処理を行う。この場合、インターネット通信は発生しない。
 - 利用するアルゴリズム
 - **【SL019】 abr Geocoder**

本システム機能の入出力データの仕様

- 入力:住所データ
 - データの内容: ジオコーディング対象の住所データ (DT102)

R7_LINKS07_行政情報を活用した空き家推定技術の実証調査_要件定義資料

- データの形式：CSV
 - データの内容：住所に緯度・経度が付与されたデータ（DT103）
 - データの形式：CSV
- 出力

2-1-4. ソフトウェア・ライブラリ (SL)

利用したソフトウェア、ライブラリを表 2-1-4-1 に示す。

表 2-1-4-1 ソフトウェア・ライブラリー一覧

ID	名称	内容
SL001	Electron	● デスクトップアプリケーション開発の OSS フレームワーク
SL002	React	● UI コンポーネントライブラリ
SL003	TypeScript	● 型安全なプログラミング言語
SL004	Vite	● フロントエンドビルドツール
SL005	Fluent UI	● Microsoft が開発したデザインシステム「Fluent Design System」を実現するための UI フレームワーク
SL006	MapLibre GL JS	● 豊富なファイル形式の空間情報の参照・加工・分析等が可能な GIS フリーソフトウェア
SL007	Recharts	● チャート描画ライブラリ
SL008	SQLite3 + Drizzle ORM	● ローカルデータベースおよび ORM
SL009	Python	● バックエンド処理の実行言語
SL010	LightGBM	● 勾配ブースティング (学習データから段階的に予測精度を高める手法。→9 章 用語集) による機械学習ライブラリ
SL011	scikit-learn	● 機械学習の補助ライブラリ (前処理・評価指標等)
SL012	Pandas	● データフレーム処理ライブラリ
SL013	GeoPandas	● 地理空間データ処理ライブラリ

SL014	Shapely	● ジオメトリ演算ライブラリ
SL015	NumPy	● 数値計算ライブラリ
SL016	Polars	● 高速データ処理ライブラリ
SL017	PLATEAU-GIS-Converter	● PLATEAU の CityGML 形式のファイルを GeoPackage 形式に変換するライブラリ
SL018	Optuna	● ベイズ最適化によるハイパーパラメータチューニングを行うことができるライブラリ
SL019	abr Geocoder	● アドレス・ベース・レジストリ (abr) で整備された住所・所在地を用いたジオコーディング機能を利用するためのライブラリ

2-1-5. 数理モデル・アルゴリズム（AL）の詳細

2-1-5-1. 利用したアルゴリズム

本業務において利用するアルゴリズムを以下に示す。

表 2-1-5-1-1 利用したアルゴリズム一覧

ID	名称	説明	選定理由	アルゴリズムを利用した機能
AL101	PU Bagging	空き家調査により空き家であることが確認された建物を「空き家（Positive）」、それ以外の建物を空き家かどうか未確認のデータとして「未知(Unlabeled)」として扱う学習手法	空き家として確認されていない建物をすべて「非空き家」として学習させることによる推定精度の低下を防ぐため	FN021
AL102	LightGBM	勾配ブースティング決定木(GBDT。複数の決定木を順番に組み合わせて高精度な推定を行う手法。→9章用語集)を利用した機械学習アルゴリズムの一手法	<ul style="list-style-type: none"> ● 単一の決定木によって構築したモデルに比べて高精度な推定が可能のため ● 学習時間が短く、高精度に推定する実績のあるアルゴリズムであるため 	FN021

2-1-5-1-1. 【AL101】 Positive Unlabeled Bagging

- 本アルゴリズムを利用する機能
 - 【FN021】 モデル構築
- アルゴリズム内で利用するライブラリ
 - なし
- アルゴリズムの詳細

Positive Unlabeled Bagging（以下、PU Bagging）とは、「正例は把握できるが、負例が明確でない」という状況において、正例以外のデータをランダムに暫定的な負例として扱う学習を複数回繰り返し、その結果を平均することで、ラベルの不完全性の影響を抑えながら安定した分類モデルを構築する手法である。

一般的な教師あり学習による分類モデルでは、「正例（Positive）」と「負例（Negative）」の両方の正解ラベルが付与されたデータを用いて学習を行う。空き家推定の文脈でいえば、「空き家」と「非空き家」のデータを学習させることで、未知のデータに対しても「空き家」が「非空き家」かの推定が可能になる。しかし、自治体を実施する空き家調査の結果には、空き家であることが確認された建物の情報は含まれる一方で、非空き家であることが確認された建物の情報は網羅的に整備されていない場合が多い。これは、調査の範囲や手法によっては、管内のすべての建物が網羅的に調査されているわけではなく、空き家かどうかを確認されていない建物が多数存在するためである。そのため、「空き家として確認されていない」ことは必ずしも「非空き家である」ことを意味しない。このような状況において、「空き家」として確認されていない建物をすべて「非空き家」として学習させると、実際には空き家である建物が誤って非空き家として扱われる可能性がある。その結果、誤ったラベルをもとにモデルが学習してしまい、推定精度の低下につながるおそれがある。

そこで LINKS SOMA では、「正例は把握できるが、負例が明確でない」というデータ構造に対応するためのアプローチである Positive-Unlabeled Learning(以下、PU Learning)を採用し、PU Learning の具体的な実装手法として PU Bagging を採用する。

PU Learning では、空き家であることが確認された建物を「空き家（Positive）」、それ以外の建物を空き家かどうか未確認のデータとして「未知（Unlabeled）」として扱う。また、Bagging とは、データからランダムに標本を繰り返し抽出し、それぞれの標本でモデルを学習させ、複数のモデルの予測結果を平均することで最終的な予測を行う手法である。PU Bagging は、この Bagging の考え方を PU Learning に適用した手法であり、空き家（Positive）データと Unlabeled データを用いて以下の手順での学習を行う。

<PU Bagging での学習手順>

1. Unlabeled データからランダムに建物データをサンプリングする
2. サンプリングしたデータを暫定的な負例として扱い、空き家（Positive）データと組み合わせ学習データを生成する

3. 上記で生成した学習データを用いて確認済み空き家とランダム抽出された未確認建物を区別するモデルを1本学習する
4. 1~3を複数回繰り返す

Unlabeled データの中には実際には空き家である建物が含まれている可能性があるが、毎回異なるランダムサンプリングにより学習データを生成することで、空き家が誤って負例として扱われる影響を統計的に分散させることができる。さらに、各モデルが出力する予測確率を平均することで、ラベルの不完全性による影響が緩和され、より安定した推定スコアを得ることが可能となる。

2-1-5-1-2. 【AL102】 LightGBM

- 本アルゴリズムを利用する機能
 - 【FN021】 モデル構築
- アルゴリズム内で利用するライブラリ
 - 【SL010】 LightGBM
- アルゴリズムの詳細

LightGBM は、勾配ブースティング決定木（Gradient Boosting Decision Tree; GBDT）を利用した機械学習アルゴリズムの一手法である。

決定木とは、「ある条件を満たすかどうか」という分岐を階層的に繰り返すことで、最終的な判定に到達するモデルである。しかし、単一の決定木は構造が単純であるため、データ全体の複雑なパターンを十分に表現できない場合がある。この課題に対しては、精度の低い決定木を多数組み合わせることで、全体として高い精度を実現する勾配ブースティング手法が有効であるとされている。具体的には、まず1本目の決定木で予測を行い、その予測が外れた事例に注目して2本目の決定木を学習し、さらにその誤差（残差）に対して3本目を学習するという過程を繰り返す。これにより、前のモデルで捉えきれなかったパターンを次のモデルが補完し、全体として精度の高い予測が可能となる。このような勾配ブースティング決定木の代表的な実装の一つが LightGBM である。

具体的には、以下の式で表される。

$$F_T(x) = f_0(x) + \sum_{t=1}^T \alpha_t f_t(x)$$

ただし

- $F_T(x)$ ：モデルの出力結果
- $f_0(x)$ ：初期学習器
- α_t ：学習率
- $f_t(x)$ ：探索方向

LINKS SOMA では、空き家推定におけるデータの特徴および運用上の要件を踏まえ、LightGBM を採用している。その理由は以下のとおりである。

第一に、大規模かつ多様なデータに対して効率的に学習できる点である。空き家推定においては、水道データや住民基本台帳データなど、件数が多く性質の異なる複数のデータを統合して扱う必要がある。そのため、大規模データに対しても高速に学習を行うことができるアルゴリズムが求められる。LightGBM はこのような条件下においても高い計算効率を持ち、実運用に適した特性を有している。

第二に、欠損値を含むデータに対して柔軟に対応できる点である。本システムでは複数のデータソース

を突合して特徴量を生成しているが、住所表記の揺れや不一致等により、一部のレコードでは十分にデータを結合できない場合がある。その結果、特徴量の一部が欠損したデータが発生する。このような状況において、欠損値を含むデータをどのように扱うかは重要な課題となる。LightGBM は、決定木の分岐過程において欠損値を自動的に処理する仕組みを備えており、欠損を含むデータを除外することなく予測に活用することが可能である。これにより、データの結合率が完全でない実運用環境においても、安定した推定を行うことができる。

以上の理由から、本システムでは教師あり機械学習のアルゴリズムとして LightGBM を採用している。

2-1-5-2. 開発したアルゴリズム

なお、本業務において開発（新規・改修）を行うアルゴリズムを朱文字で示す。

表 2-1-5-2-1 開発したアルゴリズム一覧

ID	名称	説明	アルゴリズムを利用した機能
AL001	住所正規化	<ul style="list-style-type: none"> アップロードされた住所カラムに該当するすべての列の名寄せ（住所の正規化）を行い、「住所_正規化」カラムに登録する機能 	FN011
AL002	住居単位集約	<ul style="list-style-type: none"> 住居単位では収録されていない【DT205】正規化水道栓データ、【DT204】正規化水道使用量データ、【DT206】正規化住民基本台帳から各住居の情報を説明するデータへ整形、再集計する機能 再集計後のデータのうち同一住所の住戸が複数収録されている場合において、データの種類別にバリデーション処理をする機能 	FN012
AL003	日付正規化	<ul style="list-style-type: none"> 時系列的に表現される日付情報（生年月日や異動年月日等）のうち、和暦表記（例：平成 25 年 03 月 20 日）や短縮表記（例：130320）となっているデータについて、西暦形式 8 桁（例：20030303）の表記に統一する機能 	FN011
AL004	空間結合（交差結合）	<ul style="list-style-type: none"> 【DT402】テキストマッチングデータ（ポイントデータ）に対して、建物ポリゴンデータ（ポリゴンデータ）を空間結合する。空間結合の条件には、ジオメトリ同士の交差または最近傍探索を指定することができる 	FN014
AL005	レーベンシュタイ	<ul style="list-style-type: none"> 2 つの文字列間の編集距離を計算し、住所の類似度を算出する。表記揺れの検知に利用する。 	FN015

	ン距離		
AL006	平面直角座標系選定	<ul style="list-style-type: none"> 対象自治体の位置に基づき、全国 19 ある平面直角座標系から最適な座標系を自動選定し、距離・面積計算の精度を確保する 	FN014、FN032

2-1-5-2-1. 【AL001】住所正規化

- 本アルゴリズムを利用する機能
 - 【FN011】データクレンジング
- アルゴリズム内で利用するライブラリ
 - なし
- アルゴリズムの詳細
 - 住所の正規化
 - アップロード時に、住所が含まれる列を対象に実行する。各データに含まれる住所表記は統一された基準がないため、一貫性が保たれておらず、表記ゆれが発生している場合が多い。例えば「△△町 1-3-2」と「〇〇市△△町 1 丁目 3-2」の場合、前者は大字から表記が始まることやハイフン (-) 表記で町丁目、番地、号を示しているが、後者は市区町村名から表記が始まり、町丁目は漢字表記となっている。他にも全角と半角の違いや「の」と「ノ」の違い等様々ある。データ間を住所で結合する場合にはこれらの表記を統一させることで、適切にデータを結合する必要がある。
 - 本アルゴリズムでは、以下の住所正規化を実行する。正規化後の住所は「正規化住所」というカラムで出力される。
 - 数値正規化
 - 漢数字の文字列を半角算用数字に変換
 - 全角数字を半角数字に変換
 - 例：〇〇町 1 - 3 - 2 → 〇〇町 1-3-2
 - 文字種統一（全角・半角・仮名統一）
 - 半角カタカナを全角カタカナへ変換
 - 単独カタカナをひらがなへ変換
 - 例
 - ツ / ツ → つ
 - エ → エ
 - ノ / 之 → の
 - リ → り
 - 表記ゆれ吸収（語彙統一）
 - センター → センタ
 - 通り → 通

- ふ頭 → 埠頭
- ケ / ケ / カ / ケ / ガ → が
 - 直後が地名語尾（丘 / 谷 / 崎 / 沢 / 澤 / 原 / 畑 / 尻 / 岡 / 森 / 田 / 野 / 島 等）の場合のみ変換
 - 例：旭ヶ丘 → 旭が丘
- 三つ塚 → 三塚
- 四つ屋 → 四屋
- 所在地表記の正規化
 - 丁目 →-
 - 番地 →-
 - 例：〇〇町1丁目3番地2号 → 〇〇町1-3-2
 - 大字 を削除
 - 字 を削除
 - 末尾の 番 を削除
 - 末尾の 番地 を削除
- 記号正規化
 - ハイフンを半角ハイフン（U+002D）に統一
 - 連続する半角ハイフンを一つに統合
 - 末尾のハイフンを削除
 - 不要記号を削除
 - すべてのピリオド（半角・全角）を削除
- 空白処理
 - 半角・全角スペースを削除
 - 先頭・末尾の空白を除去
 - 連続する空白を単一の空白に置換
- なお、次のような場合には対応していない。
 - 地番住所が住居表示になった場合
 - 市区町村合併等が発生した場合
 - 市区町村独自のルールで住所を管理している場合
 - 旧漢字体で住所が管理されている場合
 - 住所に一部欠落がある場合

2-1-5-2-2. 【AL002】住居単位集約

- 本アルゴリズムを利用する機能
 - 【FN012】住居単位データ作成
- アルゴリズム内で利用するライブラリ
 - 【SL012】Pandas
- アルゴリズムの詳細

入力データのうち、特に自治体保有データ（水道使用量情報、住民基本台帳）は、住居単位ではなく水道栓や個人単位で記録されている場合が多いため、各住居の情報を説明するためにデータを集約する必要がある。以下に具体的な集約、および変数の加工方法について詳述する。

- 【DT204】 正規化水道使用量データ
 - 【DT204】 正規化水道使用量データには水道栓・検針月ごとの使用水量が収録されている。
 - 各住居の水道使用量の時系列情報が必要であるため、水道栓番号を行として、月ごとの使用水量を列としたピボットテーブルを作成する。次に、使用水量の集計値（最大・平均・最小・合計）や連続で使用水量が 0 立米であった期間等を算出する。

- 【DT205】 正規化水道栓データ
 - 【DT205】 正規化水道栓データには、個々の水道栓の水道栓番号や開閉栓の日付、住所が収録されている
 - 使用開始日がユーザーの指定した空き家推定を実施したい時点（以下：推定日）以前のレコードのみを対象に、以下の処理を行う。
 - 水道栓番号に基づいて正規化水道使用水量データと結合する。
 - 結合したデータに新たに「閉栓フラグ」カラムを生成し、水道栓データのうち、推定日より前に使用中止日が記録されているレコードについては True（閉栓）、それ以外は False（開栓）を入力する。
 - なお、水道栓データでは、同一住所に対して複数のレコードが存在する可能性がある。これは、住民の転出入に伴う契約履歴の残存や、共同住宅・併用住宅等において複数の水道メーターが設置されていることなどにより発生する。この場合、当該住所における水道利用状況を一意に判断することができず、建物単位での空き家判定に利用する水道利用情報を適切に取得できなくなる。そのため、以下のロジックにより住所ごとに代表となる水道契約レコードを 1 件選定する処理を実施する。
 - 開栓中の契約が複数存在する場合
 - 同一住所において、基準日時点で開栓中の契約（閉栓日が NULL、または閉栓日が基準日より未来）が 2 件以上存在する場合、各契約の直近 1 年間の合計使用水量を算出し、その値が最も大きい契約を当該住所の代表水道契約として採用する。
 - 開栓中の契約が 1 件のみ存在する場合
 - 同一住所において、基準日時点で開栓中の契約が 1 件のみ存在する場合には、その契約を代表水道契約として採用する。
 - すべての契約が閉栓済みの場合
 - 同一住所に存在する契約がすべて閉栓済みである場合には、基準日に最も近い閉栓日を持つ契約を代表水道契約として採用する。

- これは、その住所において最も最近まで利用されていた水道契約を代表とするという考え方によるものである。

以上により、【DT305】住居単位水道データを得る。

● 【DT206】正規化住民基本台帳

- 【DT206】正規化住民基本台帳は、個人単位で世帯番号、住所、性別、生年月日、住定年月日が付与されている。
- アプリケーション側ではこれを同一世帯の構成員情報を集計して、世帯ごとの家族構成データとなるよう、世帯番号、住所をキー変数として、世帯ごとの世帯人数、年齢構成、年齢構成比、住定期間を計算する。
- ここで住居の利用実態を正確に把握することを目的に、データの対象を1住所1世帯のレコードであるものに限定する。つまり、同一時点、同一住所に複数の世帯番号が付与されている場合や、世帯番号が同一にも関わらず住所が異なる場合は、対象外となる。これは、2世帯住宅であることや住所表記上の誤り、同一年内に市内に引っ越した事例等様々な要因が考えられ、実態の把握が容易でないためである。

以上により、【DT306】住居単位住民基本台帳を得る。

● 【DT207】正規化登記情報

- 【DT207】登記情報は、同一住所に対する複数の登記履歴を有する場合がある。そのため、同一住所に紐づく登記履歴を集約し、住居単位で利用可能な情報に整理する。
- 具体的には、住所ごとに最も古い登記日付を取得し、当該住所における建物の登記日付として採用した。
- また、登記理由のうち、建物の所有状況や権利関係の変化を示す以下の登記理由については、履歴に存在するかどうかを判定し、それぞれフラグを付与した。
 - 相続
 - 贈与
 - 売買
 - 差押
- 登記構造は、「軽量鉄骨造」、「木造合金メッキ鋼板2階建て」等表記方法は多岐にわたり、またデータによってもその傾向は様々である。そのため大きく7種類の構造形式に分類して、住居の特徴を説明する変数の一つとする。具体的には以下の通りにする。
 - 木造：データの先頭文字列に「木」と記載があった場合
 - 鉄筋コンクリート造：データの先頭文字列に「RC」または「鉄筋コンクリート」と記載があった場合
 - 鉄骨造：データの先頭文字列に「S造」「鉄骨造」「軽量鉄骨」と記載があった場合
 - 煉瓦造：データの先頭文字列に「煉瓦」または「レンガ」と記載があった場合

- **コンクリートブロック造**：データの先頭文字列に「コンクリートブロック」と記載があった場合
- **土蔵造**：データの先頭文字列に「土」と記載があった場合
- **その他**：以上に該当しないデータ

以上により、【DT307】住居単位登記簿を得る。

2-1-5-2-3. 【AL003】日付正規化

- 本アルゴリズムを利用する機能
 - 【FN012】住居単位データ作成
- アルゴリズム内で利用するライブラリ
 - なし
- アルゴリズムの詳細
 - 日付情報（生年月日や異動年月日等）を西暦形式 8 桁（例：20030303）の表記に統一する。

2-1-5-2-4. 【AL004】空間結合（交差結合）

- 本アルゴリズムを利用する機能
 - 【FN014】空間結合
- アルゴリズム内で利用するライブラリ
 - 【SL013】GeoPandas
- アルゴリズムの詳細
 - まずユーザーは、空間結合を行うデータと、空間結合先であるインプットデータ（GIS）を選択する。選択後にアプリケーション側では、空間結合を行う。
 - 空間結合処理には以下の 2 パターンがある
 - 結合するデータがポイントデータで、結合先のベースデータがポリゴンデータである場合
 - 例えば、水道栓の情報をポイントデータとし、建物のポリゴンデータを結合先とする場合がこれに該当する。この場合、水道栓の情報と建物のポリゴンが空間上で交差するときに、当該水道栓の情報が対応する建物のポリゴンデータに付与される。
 - 結合するデータと結合先のベースデータの両方がポリゴンデータである場合
 - 例えば、用途地域ポリゴンを結合するデータとし、建物のポリゴンデータを結合先とする場合がこれに該当する。この場合、用途地域ポリゴンと建物ポリゴンが重なる場合、建物ポリゴンに対して重なる面積が最も大きい用途地域の情報を対応する建物ポリゴンに付与する。本処理はユーザーが特定の指

定をすることなく処理が行われる。

- 上記により結合が行われる際に、住居 ID を付与する。住居 ID は推定対象レコードに付与される。ID の生成方法は空間結合先であるインプットデータ (GIS) の建築物 ID を基準とし、建築物 ID に 4 桁からなるランダムな 16 進数の文字列を、ハイフンを入れて付与する。

2-1-5-2-5. 【AL005】レーベンシュタイン距離<新規開発>

- 本アルゴリズムを利用する機能
 - 【FN015】結合チェック
- アルゴリズム内で利用するライブラリ
 - なし
- アルゴリズムの詳細
 - レーベンシュタイン距離は、2 つの文字列間の「編集距離」を測定する手法である。ある文字列を別の文字列に変換するために必要な、挿入・削除・置換の最小操作回数を計算する。
 - 本システムでは、結合チェック (【FN015】) において、チェック対象データに含まれる町字名の類似候補を検索する際に利用する。具体的な処理は以下のとおりである。
 - チェック対象データの町字名に対し、水道データの町字名との編集距離を計算する
 - 編集距離を文字列長で正規化し、0~1 の類似度スコアに変換する (1.0 が完全一致)
 - 類似度 0.8 以上 (10 文字中 8 文字が一致する水準) の住所を最大 5 件、候補として提示する

これにより、住所の表記ゆれの原因を特定しやすくなり、データ品質の改善を支援する。

2-1-5-2-6. 【AL006】平面直角座標系選定

- 本アルゴリズムを利用する機能
 - 【FN014】空間結合、【FN032】地域集計
- アルゴリズム内で利用するライブラリ
 - 【SL013】GeoPandas
- アルゴリズムの詳細
 - 日本国内の距離・面積計算を正確に行うため、WGS84 (経緯度座標系) から平面直角座標系への変換が必要となる場合がある。日本には全国 19 の平面直角座標系 (EPSG:2443~EPSG:2461) が定義されており、対象地域に応じて適切な座標系を選定する必要がある。
 - 本システムでは、対象自治体の位置情報 (市区町村コード等) に基づき、全国 19 ある平面直角座標系から最適な座標系を自動選定する。これにより、バッファ計算 (【AL006】) や地域集計 (【AL009】) において正確な距離・面積計算を実現する。

2-2. システムコンポーネント (CO)

2-2-1. システムコンポーネント図

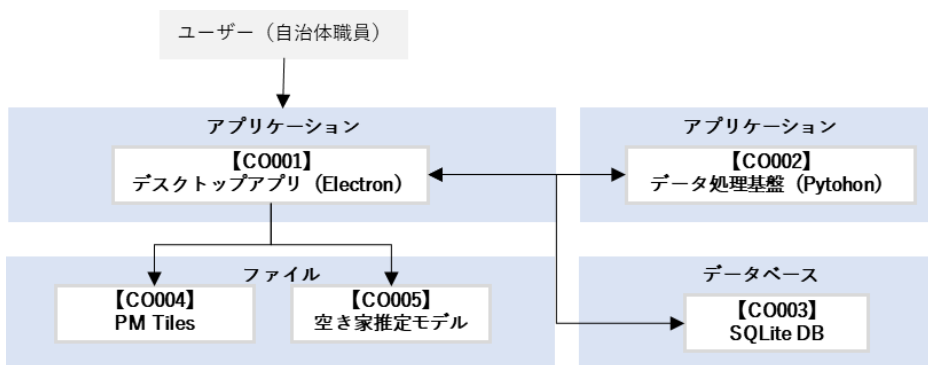


図 2-2-1-1 システムコンポーネント図 (LINKS SOMA)

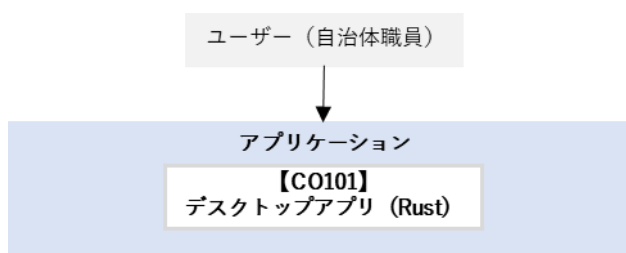


図 2-2-1-2 システムコンポーネント図 (LINKS SOMA CityGML Converter)

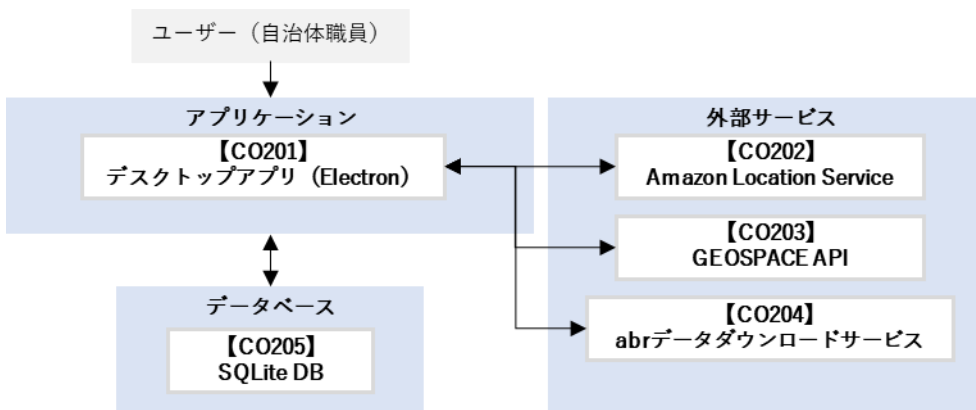


図 2-2-1-3 システムコンポーネント図 (ジオコーディングツール)

2-2-2.システムコンポーネント一覧

システム	ID	種別	コンポーネント名	用途
LINKS SOMA	CO001	アプリケーション	デスクトップアプリ (Electron)	デスクトップ上で動作するアプリケーション。データ処理基盤の呼び出しやデータベース・各種ファイルの読み込みを行い、空き家推定に必要な各処理の実行、及び実行結果の表示を行う。
	CO002	アプリケーション	データ処理基盤 (Pytohon)	PyInstaller により実行可能バイナリとして配布された Python のモジュール。空き家推定に必要な各種処理を行う。
	CO003	データベース	SQLite DB	空き家推定結果データ等、アプリケーションに必要な各種データを保存する。
	CO004	ファイル	PM Tiles	空き家推定結果をアプリケーション上に地図表示する際に利用する地図ファイル。
	CO005	ファイル	空き家推定モデル	LightGBM により構築された学習済みのモデルファイル。本モデルを利用して空き家推定を行う。
LINKS SOMA CityGML Converter	CO101	アプリケーション	デスクトップアプリ (Electron)	デスクトップ上で動作するアプリケーション。ユーザーが入力した PLATEAU データを既存ライブラリの PLATEAU GIS Converter を活用して、PLATEAU データを GeoPackage 形式に変換・出力する。
ジオコーディング ツール	CO201	アプリケーション	デスクトップアプリ (Electron)	デスクトップ上で動作するアプリケーション。ジオコーディング API の呼び出しやジオコーディング実行結果の表示・出力を行う。

	CO202	外部サービス	Amazon Location Service	Amazon Web Services (AWS) が提供するジオコーディング API。ジオコーディングに利用する。
	CO203	外部サービス	GEOSPACE API	NTT インフラネット社が提供するジオコーディング API。ジオコーディングに利用する。
	CO204	外部サービス	abr データダウンロードサービス	abr ジオコーダでのジオコーディングに必要なアドレス・ベース・レジストリのデータセットを提供する。
	CO205	データベース	SQLite DB	abr ジオコーダでのジオコーディングに必要なアドレス・ベース・レジストリのデータセットを保存する。

2-3. ハードウェアアーキテクチャ(HW)

2-3-1. ハードウェアアーキテクチャ

開発したアプリケーションはすべてデスクトップ上で動作するアプリケーションのため、ハードウェアの構成およびスペックについては設計の対象外となる。なお、アプリケーションの動作保証環境（PC スペック）については、「3. 開発するシステム：非機能要件（NF）」にて定義する。

また、LINKS SOMA および LINKS SOMA CityGML Converter はオフライン環境で処理が完結する一方で、ジオコーディングツールについては、オンライン環境に接続可能な PC での利用を前提とする。



図 2-3-1 ハードウェアアーキテクチャ

2-3-2：ハードウェアの詳細

本実証では LINKS SOMA のみを使用したため、本節ではオフライン環境で動作させた HW001 の詳細について記載する。

1) 【HW001】 PC

- 選定理由
実証対象地域の自治体が保有し、本実証を実施するために必要な通信環境を有しているため
- 仕様・スペック
 - 豊田市
 - 型番：PC-VKL41EZFB
 - OS：Windows 10 Professional
 - メモリ：8GB
 - CPU：Intel Core i3-10110U
 - GPU：Intel UHD Graphics
 - ストレージ：256GB
 - 豊橋市
 - 型番：PC-VKT42M3FB
 - OS：Windows 10 Professional
 - メモリ：8GB
 - CPU：12th Gen Intel(R) Core(TM) i5-1235U
 - GPU：Intel UHD Graphics
 - ストレージ：128GB
 - 岡崎市
 - 型番：PC-VKT42M3FB
 - OS：Windows 10 Professional
 - メモリ：8GB
 - CPU：12th Gen Intel(R) Core(TM) i5-1235U
 - GPU：Intel UHD Graphics
 - ストレージ：128GB
 - 下関市
 - 型番：PC-VKT42M3FB
 - OS：Windows 10 Professional
 - メモリ：8GB
 - CPU：12th Gen Intel(R) Core(TM) i5-1235U
 - GPU：Intel UHD Graphics
 - ストレージ：128GB
 - 熊谷市

- 型番：PC-VKT42M3FB
- OS：Windows 10 Professional
- メモリ：8GB
- CPU：12th Gen Intel(R) Core(TM) i5-1235U
- GPU：Intel UHD Graphics
- ストレージ：128GB

2-4. データインターフェース(IF)

2-4-1. データアーキテクチャ図

※各データ (DT) の詳細は 4 章「実証調査に利用するデータ (DT)」を参照すること

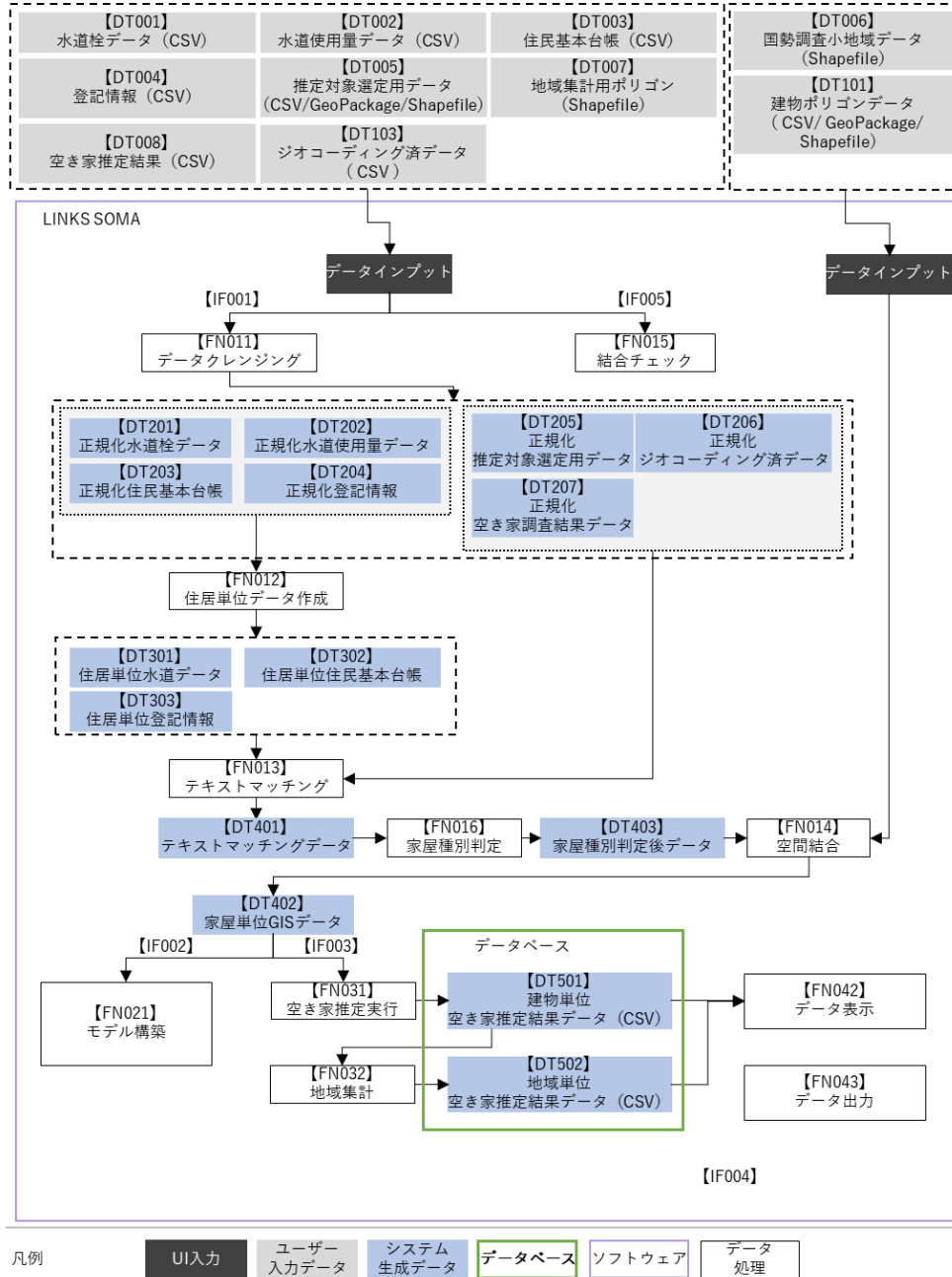
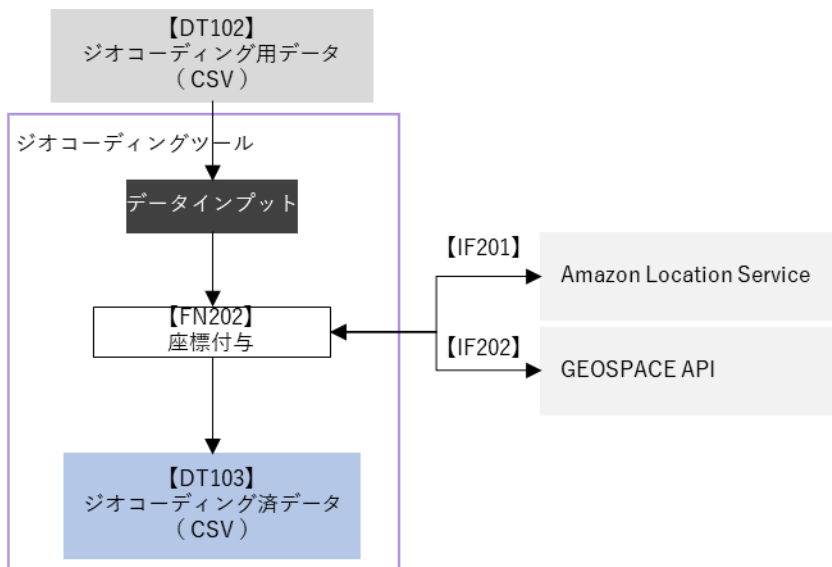


図 2-4-1-1 データアーキテクチャ図 (LINKS SOMA)



凡例

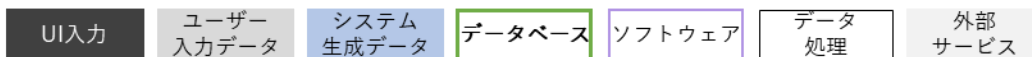


図 2-4-1-2 データアーキテクチャ図 (ジオコーディングツール)

2-4-2. データインターフェース一覧

データインターフェースの一覧を示す。なお、本業務において開発（新規・改修）を行うアルゴリズムを朱文字で示す。

表 2-4-2-1 データインターフェース一覧

システム	ID	IF 名	本インターフェースを利用する機能
LINKS SOMA	IF001	名寄せ処理	FN011～FN014、FN016
	IF002	モデル構築	FN202
	IF003	空き家推定	FN031～FN032
	IF004	データダウンロード	FN043
	IF005	結合チェック	FN015
ジオコーディングツール	IF201	ジオコーディング API (AWS)	FN202
	IF202	ジオコーディング API (NTT インフラネット)	FN202

2-4-3. データインターフェースの詳細

2-4-3-1. 【IF001】 名寄せ処理

- 本インターフェースの概要
 - 名寄せ処理の対象となる各種データや名寄せ処理の条件を取り込むためのインターフェース
- 本インターフェースを利用する機能
 - 【FN011】 データクレンジング
 - 【FN012】 住居単位データ作成
 - 【FN013】 テキストマッチング
 - 【FN014】 空間結合
 - 【FN015】 建物種別判定
- Executable
 - EXE
- メソッド
 - Local process execution (spawn)
- パス
 - /dist/IF001
- リクエストパラメータ

表 2-4-3-1 IF001 リクエストパラメータ

パラメータ	説明	値	必須
output_path	出力先フォルダのディレクトリパス	string	○
database_path	データベースファイルのディレクトリパス	string	○
job_id	ジョブ ID	string	○
settings.reference_date	推定基準日	string (YYYY-MM-DD)	○
data.geocoding	ジオコーディング済	object	

	データ		
data.geocoding.columns.address	住所カラム名	string	○
data.geocoding.columns.latitude	緯度カラム名	string	○
data.geocoding.columns.longitude	経度カラム名	string	○
data.building_polygon	建物ポリゴンデータ	object	
data.building_polygon.input_file_type	ファイル形式	"geopackage" / "shapefile"	—
data.building_polygon.data_type	データ種別	"plateau" / "others"	—
data.water_status	水道閉開栓状況	object	○
data.water_status.columns.water_supply_number	給水番号カラム名	string	○
data.water_status.columns.address	住所カラム名	string	○
data.water_status.columns.water_connection_date	給水開始日カラム名	string	○
data.water_status.columns.water_disconnection_date	給水停止日カラム名	string	○
data.water_usage	水道使用量	object	○
data.water_usage.columns.water_supply_number	給水番号カラム名	string	○
data.water_usage.columns.water_usage	使用量カラム名	string	○

data.water_usage.columns.water_recorded_date	計量日カラム名	string	○
data.resident_registry	住民基本台帳	object	
data.resident_registry.columns.household_code	世帯コードカラム名	string	○
data.resident_registry.columns.address	住所カラム名	string	○
data.resident_registry.columns.birth_date	生年月日カラム名	string	○
data.resident_registry.columns.resident_date	転入年月日カラム名	string	○
data.resident_registry.columns.reason_transfer	転入理由カラム名	string	○
data.resident_registry.columns.date_transfer	転出年月日カラム名	string	○
data.building_registry	登記情報	object	
data.building_registry.columns.address	住所カラム名	string	
data.building_registry.columns.structure_name	構造カラム名	string	
data.building_registry.columns.registration_reason	登記理由カラム名	string	
data.building_registry.columns.registration_date	登記日付カラム名	string	
data.census	国勢調査	object	○
data.building_type_determination	家屋種別判定用データ	object	

data.building_type_determination.input_file_type	ファイル形式	"csv" / "geopackage" / "shapefile"	
data.building_type_determination.columns.address	住所カラム名 (CSV 形式の場合)	string	
data.building_type_determination.columns.building_type	建物種別カラム名	string	
data.building_type_determination.residential_values	住宅地判定値リスト	string[]	
data.vacant_house	空き家調査結果	object	
data.vacant_house.columns.address	住所カラム名	string	
data.optional_data_source	追加データソース	object	
data.optional_data_source.columns.address	住所カラム名	string	

- レスポンスパラメータ

表 2-4-3-2 IF001 レスポンスパラメータ

パラメータ	説明	型	必須
status	処理の成否	string	○
type	インターフェースの種類	string	○
preprocess_type	名寄せ処理内で実行された処理の種類	string	○
error_code	エラー発生時のエラーコード	string	○ ※エラー発生時のみ
error_msg	エラー発生時のエラーメッセージ	string	○ ※エラー発生時のみ

2-4-3-2. 【IF002】モデル構築

- 本インターフェースの概要
 - モデル構築に必要となるデータや構築の条件を取り込むためのインターフェース
- 本インターフェースを利用する機能
 - 【FN021】モデル構築
- Executable
 - EXE
- メソッド
 - Local process execution (spawn)
- パス
 - /dist/IF002
- リクエストパラメータ

表 2-4-3-3 IF002 リクエストパラメータ

パラメータ	説明	値	必須
input_path	入力パス	string	○
settings	設定	object	○
settings.explanatory_variables	説明変数	string[]	
settings.advanced	高度な設定	object	○
settings.advanced.test_size	モデル構築・精度検証における学習とテストの比率。(デフォルト値は 7:3)	string[]	○
settings.advanced.n_splits	モデル構築において、データを分割(k-fold)して交差検証する回数を指定するオプション。回数が多いほど安定した結果を得られるようになるが、計算量が多くなるため注意が必要。(デフォルトでは3)	float	○

settings.advanced.undersample	学習用データの空き家件数と非空き家件数のサンプル数の比率を指定するオプション。【IF107】空き家調査結果が小規模で、空き家数が極端に少ない場合等により、精度が出ない場合に利用（デフォルトではオフ）。	boolean	○
settings.advanced.undersample_ratio	学習用データの空き家件数と非空き家件数のサンプル数の比も任意に設定（デフォルトでは 3.0）。	int	○
settings.advanced.threshold	空き家確率が一定以上の場合に、対象住居を空き家と推定するための閾値（デフォルトでは 0.3）。	0~1	○
settings.advanced.hyperparameter_flag	モデルの精度向上オプション。自動で LightGBM のハイパーパラメータ（学習率、木の深さ、正則化パラメータ等）の最適化を実施。ただし、計算時間が非常に長くなるため、精度が上がらない場合に実施（デフォルトではオフ）。	boolean	○
settings.advanced.n_trials	ハイパーパラメータ最適化の試行回数オプション。（デフォルトでは 100）	int	○
settings.advanced.lambda_l1	L1 正則化項の係数。説明変数が多く、過学習が起きる際に利用。値を大きくしすぎると、学習が進みにくくなる（精度が低下する）場合があるため、注意が必要。 （デフォルトでは 0(正則化なし)）	float	○

settings.advanced.lambda_l2	L2 正則化項の係数。サンプル数が少なく、過学習が起きる際に利用。値を大きくしすぎると、学習が進みにくくなる（精度が低下する）場合があるため、注意が必要。 (デフォルトでは 0(正則化なし))	float	○
settings.advanced.num_leaves	1 本の決定木の木 (分岐) の葉の最大数 (デフォルトでは 31)	float	○
settings.advanced.feature_fraction	各決定木においてランダムに抽出される列の割合 (デフォルトでは 1.0)	float	○
settings.advanced.bagging_fraction	各決定木においてランダムに抽出される標本の割合 (デフォルトでは 1.0)	float	○
settings.advanced.bagging_freq	ここで指定したイテレーションごとにバギング実施 (デフォルトでは 0)	int	○
settings.advanced.min_data_in_leaf	1 枚の葉に含まれる最少データ数 (デフォルトでは 20)	int	○
output_path	出力先フォルダのディレクトリパス	string	○
database_path	データベースファイルのディレク	string	○

	トリパス		
--	------	--	--

- レスポンスパラメータ

表 2-4-3-4 IF002 レスポンスパラメータ

パラメータ	説明	型	必須
status	処理の成否	string	○
type	呼び出したインターフェースの種類	string	○
error_code	エラー発生時のエラーコード	string	○ ※エラー発生時のみ
error_msg	エラー発生時のエラーメッセージ	string	○ ※エラー発生時のみ

2-4-3-3. 【IF003】 空き家推定

- 本インターフェースの概要
 - 空き家推定処理に必要となる各種データや推定の条件を取り込むためのインターフェース
- 本インターフェースを利用する機能
 - 【FN031】 空き家推定実行
 - 【FN032】 地域集計
- Executable
 - EXE
- メソッド
 - Local process execution (spawn)
- パス
 - /dist/IF003
- リクエストパラメータ

表 2-4-3-5 IF003 リクエストパラメータ

パラメータ	説明	型	必須
model_path	モデルファイルのパス	string	○
normalized_dataset_paths	推定に用いるデータセットのパス	string	○
settings.threshold	空き家判定の閾値	int	○
area_grouping.path	地域集計用ファイルのパス	string	○
area_grouping.columns.area_group_id	地域 ID カラム	string	○
area_grouping.columns.area_group_name	地域名称カラム	string	○
output_path	出力先フォルダのディレクトリパス	string	○
database_path	データベースファイルのディレクトリパス	string	○

- レスポンスパラメータ

表 2-4-3-6 IF003 レスポンスパラメータ

パラメータ	説明	型	必須
status	処理の成否	string	○
type	呼び出したインターフェースの種類	string	○
error_code	エラー発生時のエラーコード	string	○ ※エラー発生時のみ
error_msg	エラー発生時のエラーメッセージ	string	○ ※エラー発生時のみ

2-4-3-4. 【IF004】 データダウンロード

- 本インターフェースを利用する機能
 - 【FN043】 データ出力
- Executable
 - EXE
- メソッド
 - Local process execution (spawn)
- パス
 - /dist/IF004
- リクエストパラメータ

表 2-4-3-7 IF004 リクエストパラメータ

パラメータ	説明	型	必須
status	処理の成否	string	○
type	呼び出したインターフェースの種類	string	○

error_code	エラー発生時のエラーコード	string	○ ※エラー発生時のみ
error_msg	エラー発生時のエラーメッセージ	string	○ ※エラー発生時のみ

- レスポンスパラメータ

表 2-4-3-8 IF004 レスポンスパラメータ

パラメータ	説明	型	必須
status	処理の成否	string	○
type	呼び出したインターフェースの種類	string	○
error_code	エラー発生時のエラーコード	string	○ ※エラー発生時のみ
error_msg	エラー発生時のエラーメッセージ	string	○ ※エラー発生時のみ

2-4-3-5. 【IF005】結合チェック<新規開発>

- 本インターフェースの概要
 - 名寄せ処理の事前チェックとして、水道栓データと各種データの結合有無を確認するためのインターフェース
- 本インターフェースを利用する機能
 - 【FN015】結合チェック

- Executable
 - EXE
- メソッド
 - Local process execution (spawn)
- パス
 - /dist/IF005
- リクエストパラメータ

表 2-4-3-9 IF005 リクエストパラメータ

パラメータ	説明	型	必須
data.water_status	水道栓データ	object	○
data.resident_registry	住民基本台帳データ	object	
data.building_registry	建物情報データ	object	
data.geocoding	ジオコーディング済データ	object	
data.building_type_determination	家屋種別判定データ	object	
settings.threshold	レーベンシュタイン距離の類似度閾値 (デフォルト 0.8)	int	○
settings.max_number	類似住所候補の最大提示数 (デフォルト 5)	int	○
output_path	出力先フォルダのディレクトリパス	string	○
database_path	データベースファイルのディレクトリパス	string	○

- レスポンスパラメータ
なし (IF005 の実行結果は DB に JSON 形式で保存され、UI からポーリングで取得する)

2-4-3-6. 【IF201】 ジオコーディング API (AWS)

※本 API の詳細仕様は公式ドキュメント

(<https://docs.aws.amazon.com/location/latest/developerguide/geocode.html>) を参照すること

- 本インターフェースの概要
 - AWS が提供するサービスでジオコーディングをおこなうための WEB API
- 本インターフェースを利用する機能
 - 【FN202】 座標付与
- メソッド
 - POST
- リクエストパラメータ

表 2-4-3-10 IF201 リクエストパラメータ

パラメータ	説明	値	必須
Text	座標に変換したい住所を指定する	string	○

- レスポンスパラメータ

表 2-4-3-11 IF201 レスポンスパラメータ

パラメータ	説明	値	必須
lat	緯度	number	
lon	経度	number	
label	正規化住所	string	
success	処理の成否	boolean	○
errorMessage	エラーメッセージ	string	

2-4-3-7. 【IF202】 ジオコーディング API (GEOSPACE WEB API)

※本 API の詳細仕様は公式ドキュメント (<https://api-docs-dev.geospace.services/api/WebManual/geocoding/directory>) を参照すること

- 本インターフェースの概要
 - NTT インフラネットが提供するサービスでジオコーディングをおこなうための WEB API
- 本インターフェースを利用する機能
 - 【FN202】 座標付与
- メソッド
 - POST
- リクエストパラメータ

表 2-4-3-12 IF202 リクエストパラメータ

パラメータ	説明	値	必須
string	座標に変換したい住所を指定する	string	○
appid	アプリケーション ID(API 利用申請時に発行された ID)	string	○

- レスポンスパラメータ

表 2-4-3-13 IF202 レスポンスパラメータ

パラメータ	説明	値	必須
lat	緯度	number	
lon	経度	number	
label	正規化住所	string	
success	処理の成否	boolean	○
errorMessage	エラーメッセージ	string	

2-5. ユーザーインターフェース(SC)

本節では、画面の一覧、遷移図、および各画面の仕様を示す。朱文字は今年度の新規・変更箇所を示す。

2-5-1. ユーザーインターフェース一覧

- LINKS SOMA 空き家推定システム

凡例：

- SC001～：分析画面
- SC101～：名寄せ処理実行画面
- SC201～：モデル管理画面
- SC301～：空き家推定実行画面
- SC401～：データセット管理画面
- SC501～：処理一覧画面

表 2-5-1-1 LINKS SOMA 画面一覧

ID	画面名	画面説明	本画面から利用する機能
SC001	ワークブック一覧画面	● ワークブックを一覧で表示する	FN041
SC002	ワークブック編集画面	● ワークブックを編集する	FN041
SC003	ワークブック閲覧画面	● 編集したワークブックを確認する	FN041、FN042
SC101	名寄せ処理一覧画面	● 名寄せ処理の開始・ステータス管理を行う	FN001
SC102	名寄せ処理実行画面	● インプットデータの選択や設定を行い、名寄せ処理を実行する	FN011～FN015
SC103	表記揺れチェック画面	● インプットデータ間の住所の表記揺れをチェックする	FN015
SC201	モデル管理画面	● 学習済モデルのアップロード・一覧・管理を行う	FN021
SC202	モデル構築画面	● モデルのトレーニングを実行する	FN021

SC203	名寄せ処理済みデータ選択画面	<ul style="list-style-type: none"> ● 名寄せ処理済みデータの選択を行う 	FN021
SC204	説明変数の選択画面	<ul style="list-style-type: none"> ● 推定に利用する説明変数の選択を行う 	FN021
SC205	高度な設定の変更画面	<ul style="list-style-type: none"> ● モデルトレーニングの設定を変更する 	FN021
SC301	空き家推定一覧画面	<ul style="list-style-type: none"> ● 空き家推定の実行・ステータス管理を行う 	FN001
SC302	空き家推定実行画面	<ul style="list-style-type: none"> ● モデル・データ・設定の選択を行い推定を実行する 	FN031、FN032
SC303	モデル選択画面	<ul style="list-style-type: none"> ● モデルの選択を行う 	FN021
SC304	名寄せ済み処理データ選択画面	<ul style="list-style-type: none"> ● 名寄せ済み処理データの選択を行う 	FN001
SC305	地域集計用データ選択画面	<ul style="list-style-type: none"> ● 地域集計用データの選択を行う 	FN001
SC306	高度な設定の変更画面	<ul style="list-style-type: none"> ● 高度な設定の選択を行う 	-
SC401	インプットデータ管理画面	<ul style="list-style-type: none"> ● インプットデータを入力・管理・出力する 	FN001
SC402	名寄せ処理済みデータ管理画面	<ul style="list-style-type: none"> ● 名寄せ処理済みデータを管理・出力する 	FN001
SC403	空き家推定結果データ管理画面	<ul style="list-style-type: none"> ● 推定結果データを管理・出力する 	FN001
SC404	データセットプレビュー画面	<ul style="list-style-type: none"> ● データセットのプレビューを確認する 	FN001

SC501	処理一覧画面	<ul style="list-style-type: none"> ● 名寄せ・推定・出力のステータスを確認する 	FN001
SC502	名寄せ処理結果確認画面	<ul style="list-style-type: none"> ● 名寄せ処理の結果確認・再実行 	-
SC503	モデル確認画面	<ul style="list-style-type: none"> ● モデルの詳細確認 	FN021
SC504	ダウンロード結果確認画面	<ul style="list-style-type: none"> ● 出力処理の結果確認・ダウンロード 	FN043
SC505	推定結果確認画面	<ul style="list-style-type: none"> ● 空き家推定の完了確認 	-

- LINKS SOMA CityGML Converter

表 2-5-1-2 LINKS SOMA CityGML Converter 画面一覧

ID	画面名	画面説明	関連する機能 (ID)
SC601	メイン画面	<ul style="list-style-type: none"> ● CityGML 形式データの選択と出力先を設定する ● GeoPackage 形式への変換を実行する 	【FN101】
SC602	処理ステータス画面	<ul style="list-style-type: none"> ● 変換のステータスを確認する ● 実行を途中でキャンセルする 	【FN101】

- ジオコーディングツール

表 2-5-1-3 ジオコーディングツール画面一覧

ID	画面名	画面説明	関連する機能 (ID)
SC701	メイン画面	<ul style="list-style-type: none">● 住所データの入力、外部サービスの API キーの入力をする● ジオコーディングのテスト実行および本実行をする● 本実行結果をダウンロードする	【FN202】

2-5-2. 画面遷移図

- LINKS SOMA 空き家推定システム

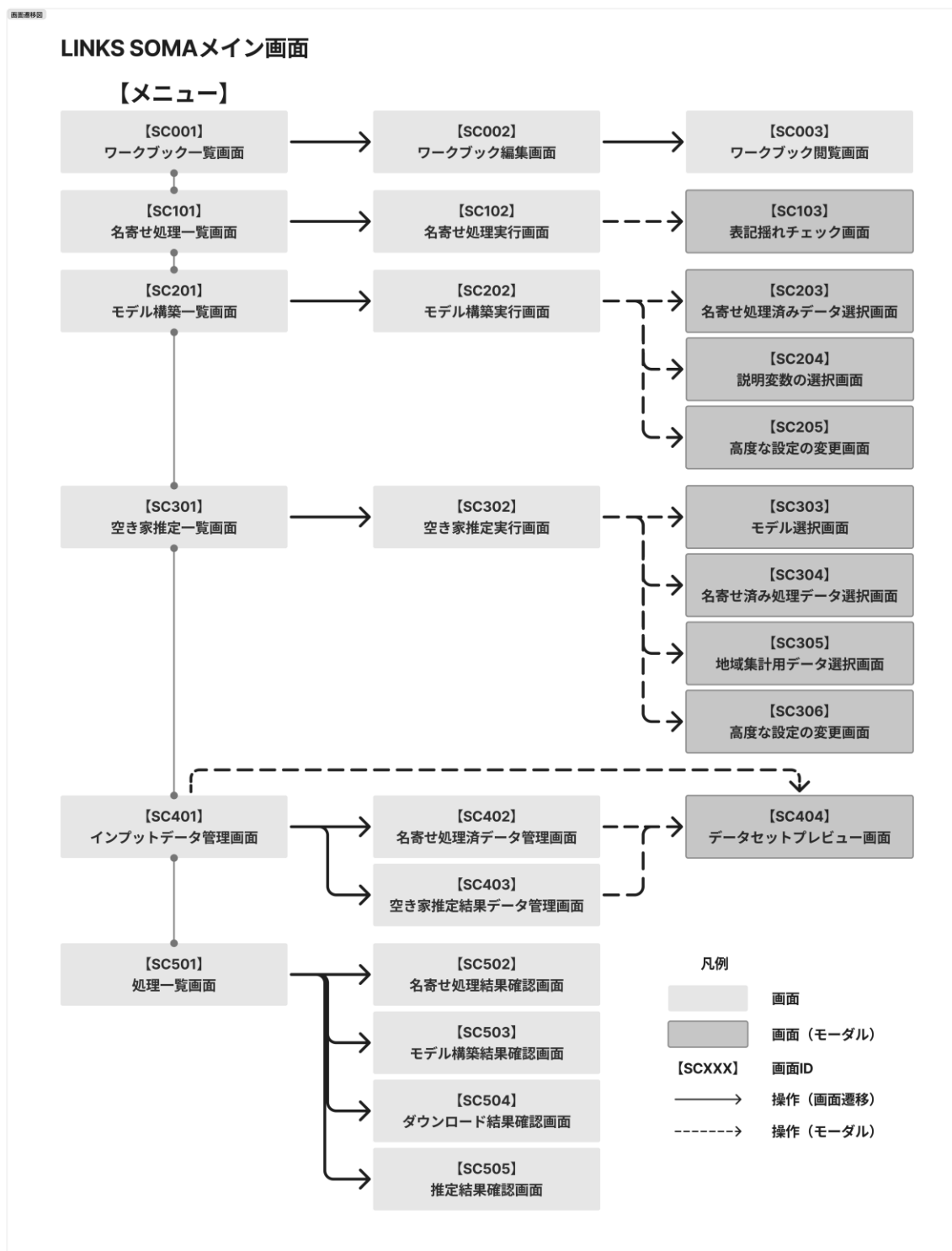


図 2-5-2-1 LINKS SOMA 画面遷移図

- LINKS SOMA CityGML Converter



図 2-5-2-2 LINKS SOMA CityGML Converter 画面遷移図

- ジオコーディングツール



図 2-5-2-3 ジオコーディングツール画面遷移図

2-5-3. ユーザーインターフェースの詳細

2-5-3-1. LINKS SOMA 空き家推定システム

2-5-3-1-1. 【SC001】ワークブック一覧画面

- 画面の目的・概要
 - ユーザーが新規作成/編集したワークブック（分析結果を整理するためのファイル単位。→9章用語集）を一覧で表示する
- 画面イメージ

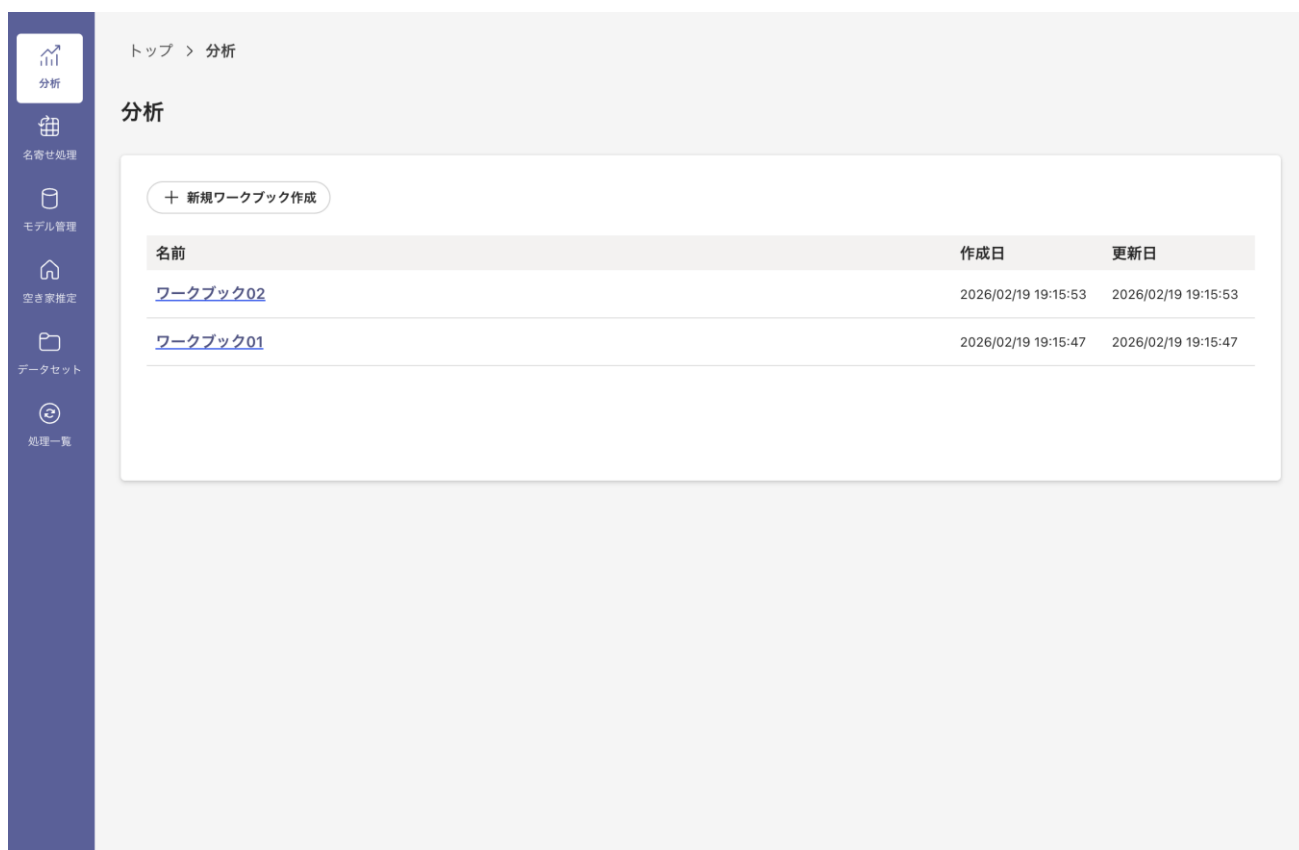


図 2-5-3-1-1 ワークブック一覧画面のイメージ

2-5-3-1-2. 【SC002】ワークブック編集画面

- 画面の目的・概要
 - ワークブックを編集する。地図・グラフ・テーブルの各ビュー（表示形式。→9章 用語集）を設定し、データを可視化する
- 画面イメージ



図 2-5-3-1-2 ワークブック編集画面のイメージ

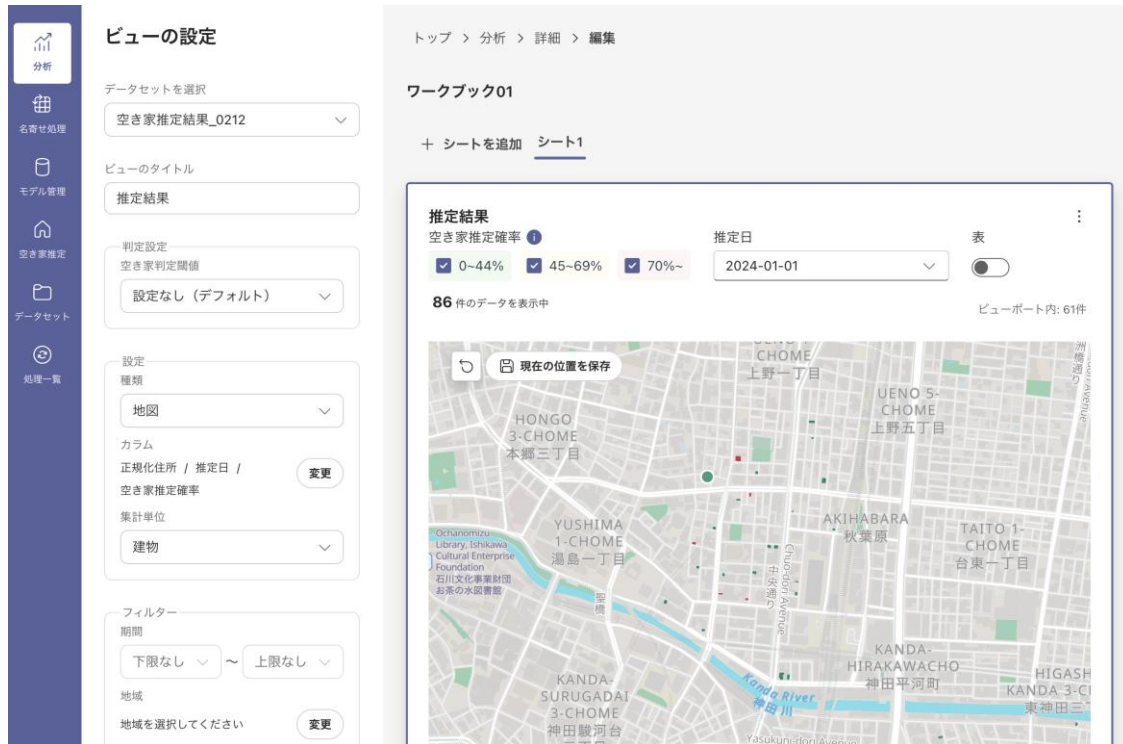


図 2-5-3-1-3 ワークブック編集画面（建物単位の地図ビュー）のイメージ

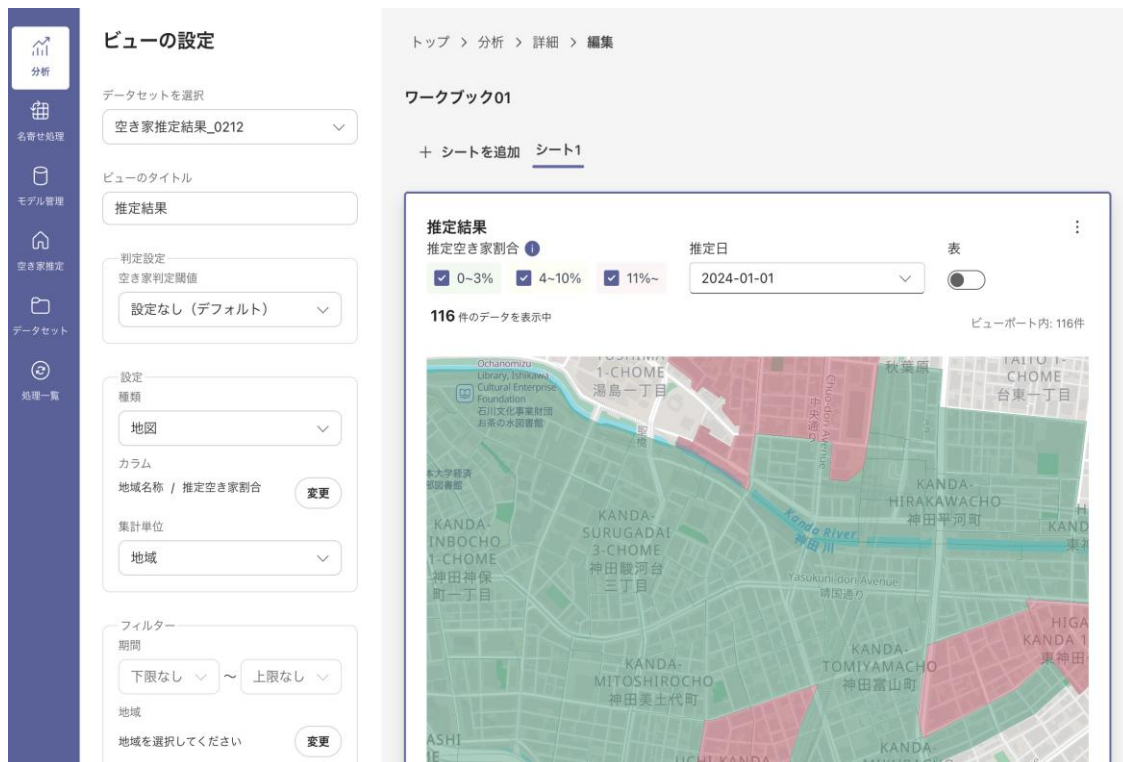


図 2-5-3-1-4 ワークブック編集画面（地域単位の地図ビュー）のイメージ

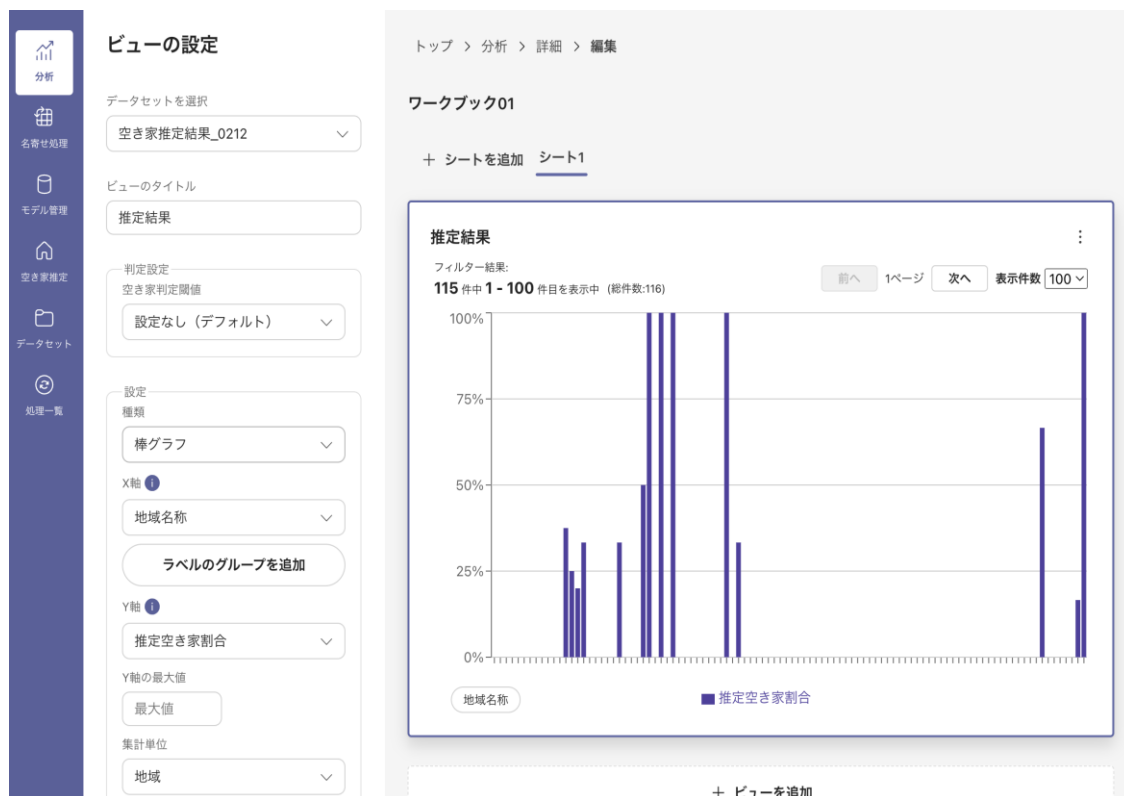


図 2-5-3-1-5 ワークブック編集画面（棒グラフビュー）のイメージ

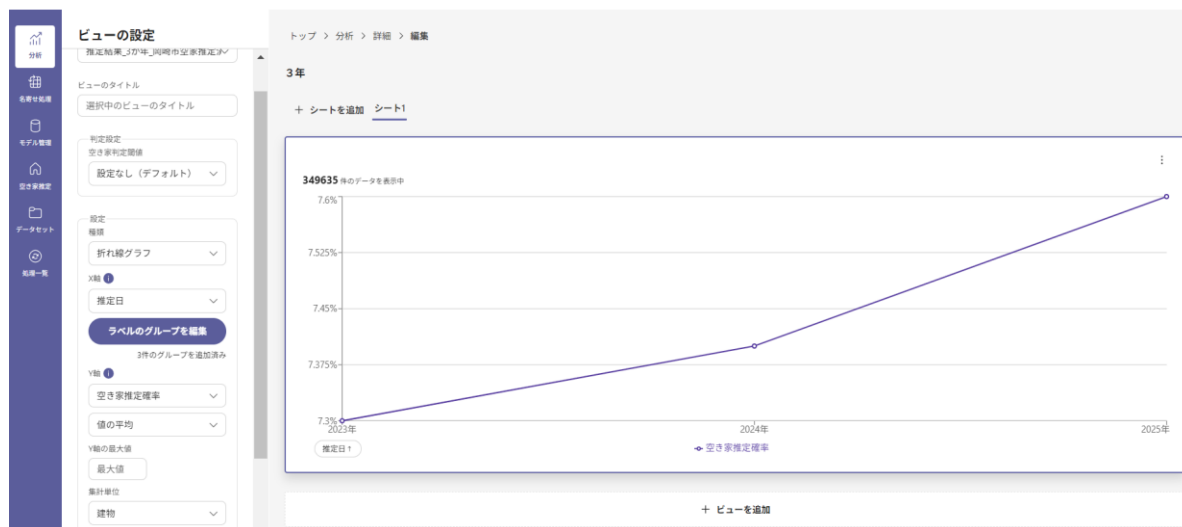


図 2-5-3-1-6 ワークブック編集画面（折れ線グラフビュー）のイメージ

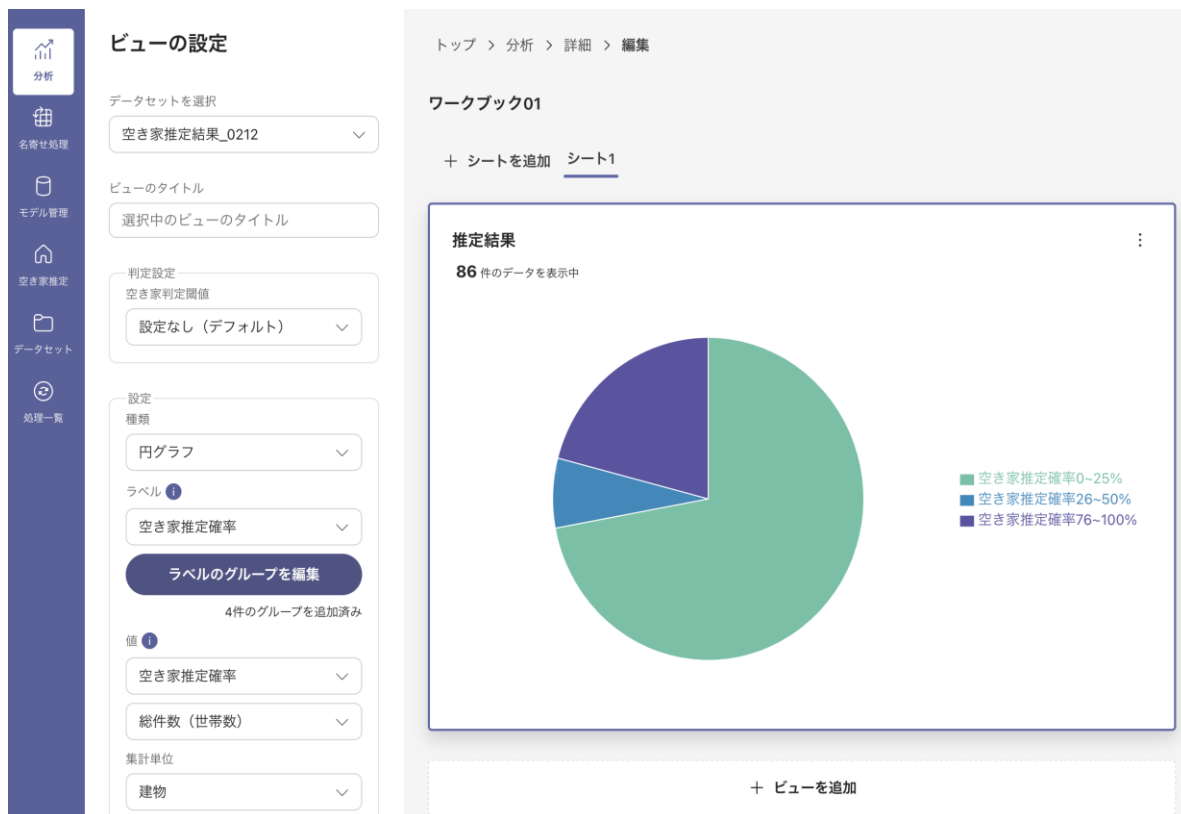


図 2-5-3-1-7 ワークブック編集画面（円グラフビュー）のイメージ

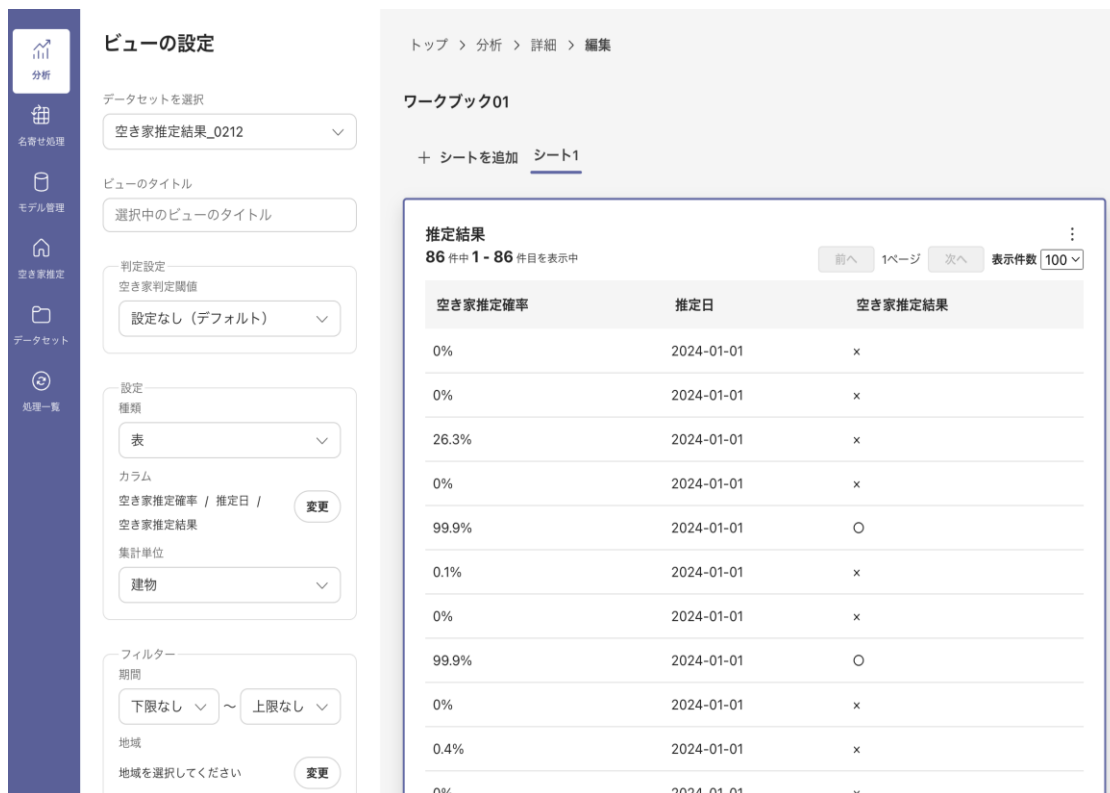


図 2-5-3-1-8 ワークブック編集画面（表ビュー）のイメージ



図 2-5-3-1-9 ワークブック編集画面（グループの設定）のイメージ



図 2-5-3-1-10 ワークブック編集画面（フィルターの設定）のイメージ

2-5-3-1-3. 【SC003】ワークブック閲覧画面

- 画面の目的・概要
 - 編集したワークブックを確認/プレビューする
- 画面イメージ

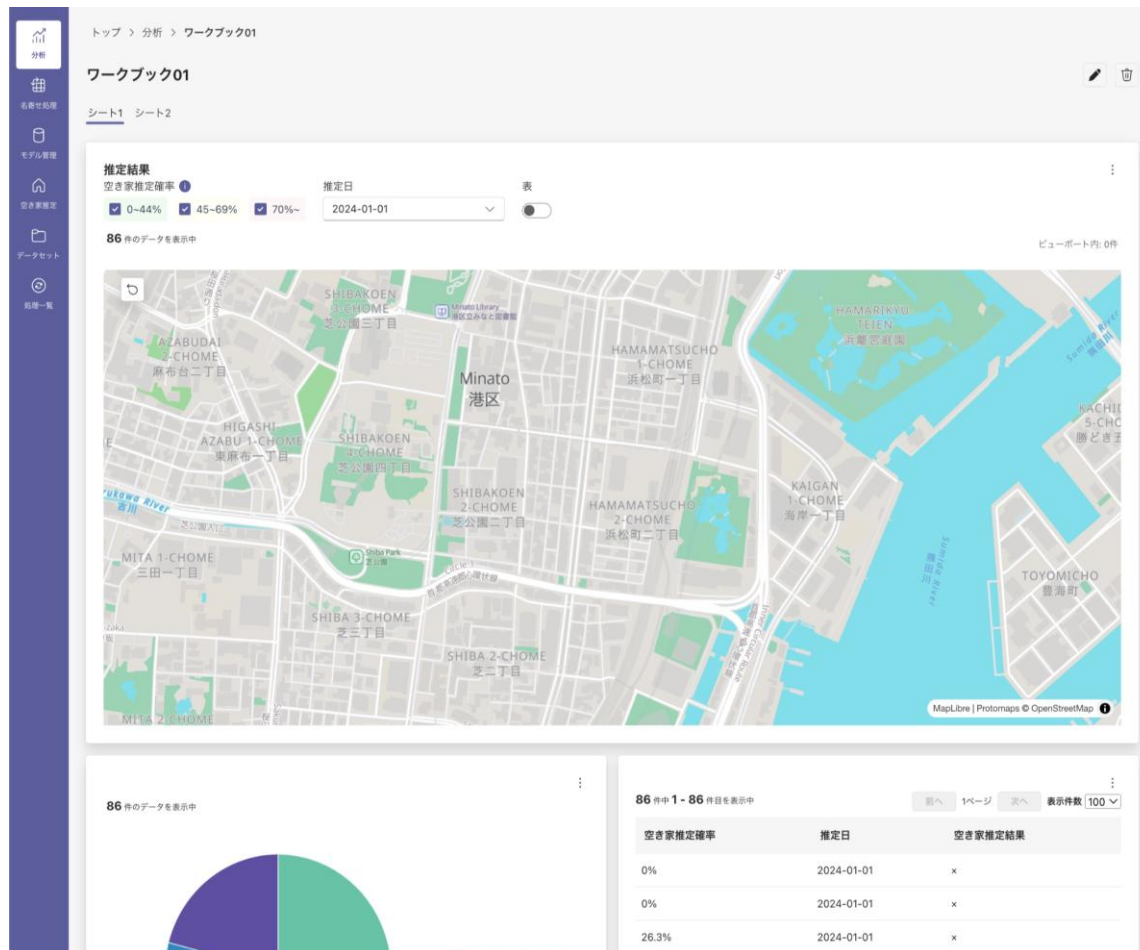


図 2-5-3-1-11 ワークブック閲覧画面のイメージ

2-5-3-1-4. 【SC101】名寄せ処理一覧画面

- 画面の目的・概要
 - 名寄せ処理を開始する
 - ユーザーが実行した名寄せ処理のステータスや結果を管理する
- 画面イメージ

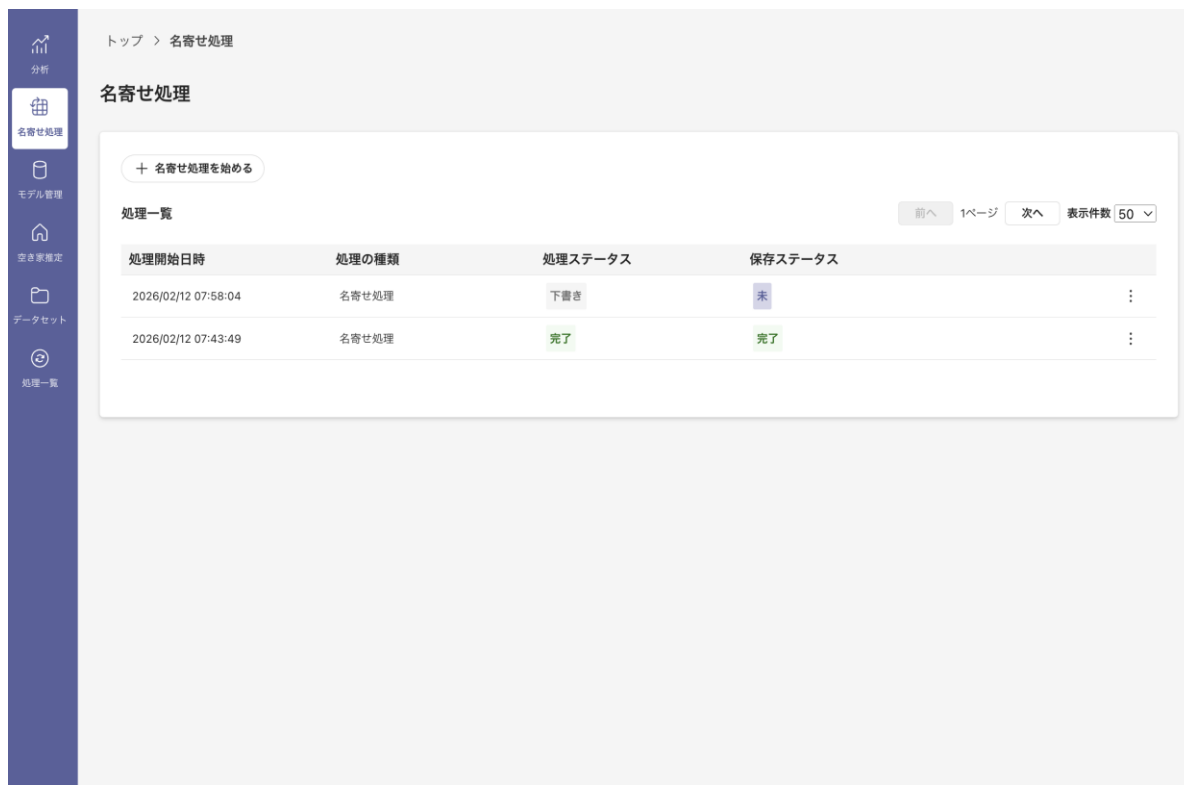


図 2-5-3-1-12 名寄せ処理一覧画面のイメージ

2-5-3-1-5. 【SC102】名寄せ処理実行画面

- 画面の目的・概要
 - 13 ステップのウィザード形式で、インプットデータの選択・基準日の設定・各種データセットの入力を行い、名寄せ処理を実行する
- 画面イメージ



図 2-5-3-1-13 名寄せ処理実行画面（ウィザード）のイメージ

2-5-3-1-6. 【SC103】表記揺れチェック画面

- 画面の目的・概要
 - インputデータ間の住所（町字名）の表記揺れを確認する
- 画面イメージ

住所の表記ゆれチェック

各データに含まれる住所（町字まで）のうち、水道データに存在しない住所を抽出します。該当する住所については、元データの表記を修正することで、名寄せ処理での結合率が改善できる可能性があります。

▼ 住民基本台帳

表記ゆれの可能性がある住所（1件） ⓘ

↓ CSVダウンロード

件数	不一致の住所	水道データの類似住所	水道データ件数	類似度
1	千代田区鍛冶町	千代田区鍛冶町 ⓘ	9	86%

※住所の表記ゆれを確認するための参考情報です。結合できない原因の特定や、名寄せ結果の予測はできません

※名寄せ処理では水道データの住所をマスタとして各データとの結合を行います

※表記の修正は本システム外で行い、修正後に再度アップロードしてください

再実行

閉じる

図 2-5-3-1-14 表記揺れチェック画面のイメージ

2-5-3-1-7. 【SC201】モデル構築一覧画面

- 画面の目的・概要
 - 学習済モデルのアップロード・一覧・管理を行う
 - モデルの学習を開始する
- 画面イメージ



図 2-5-3-1-15 モデル構築画面のイメージ

2-5-3-1-8. 【SC202】モデル構築実行画面

- 画面の目的・概要
 - 名寄せ処理済みデータの選択や必要な設定を行い、モデル構築を実行する
- 画面イメージ

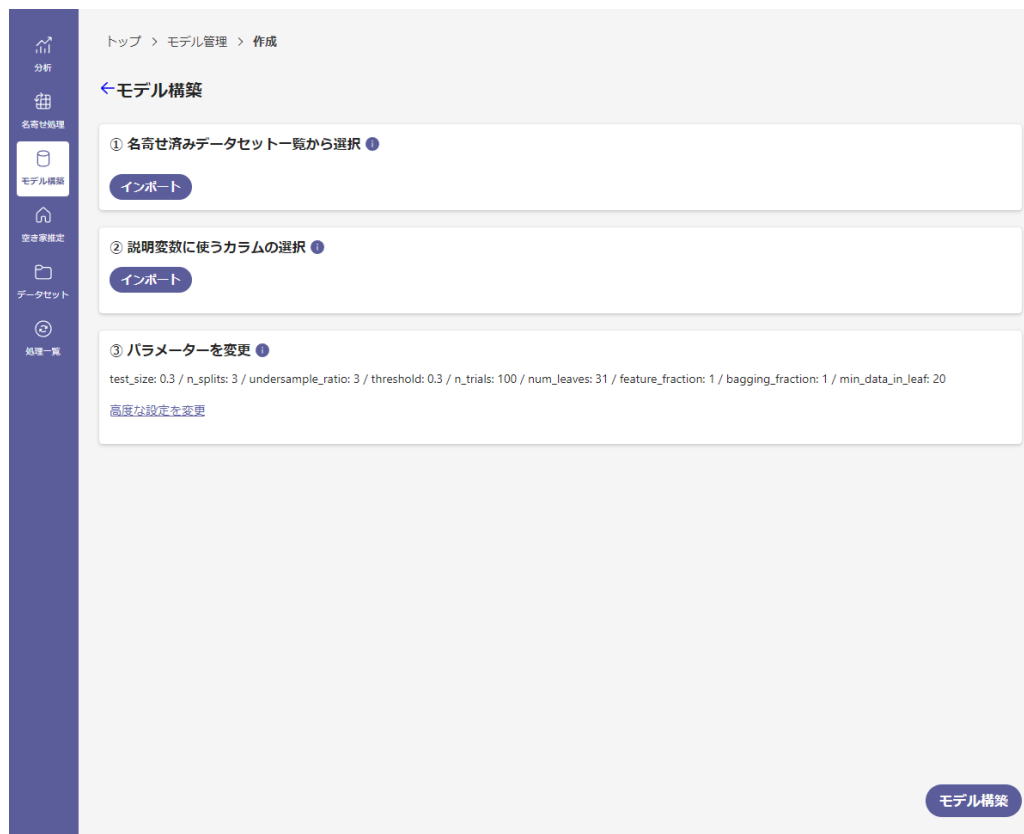


図 2-5-3-1-16 モデル構築実行画面のイメージ

2-5-3-1-9. 【SC203】名寄せ処理済データ選択画面

- 画面の目的・概要
 - モデル構築に利用する名寄せ処理済みデータを選択をする
- 画面イメージ

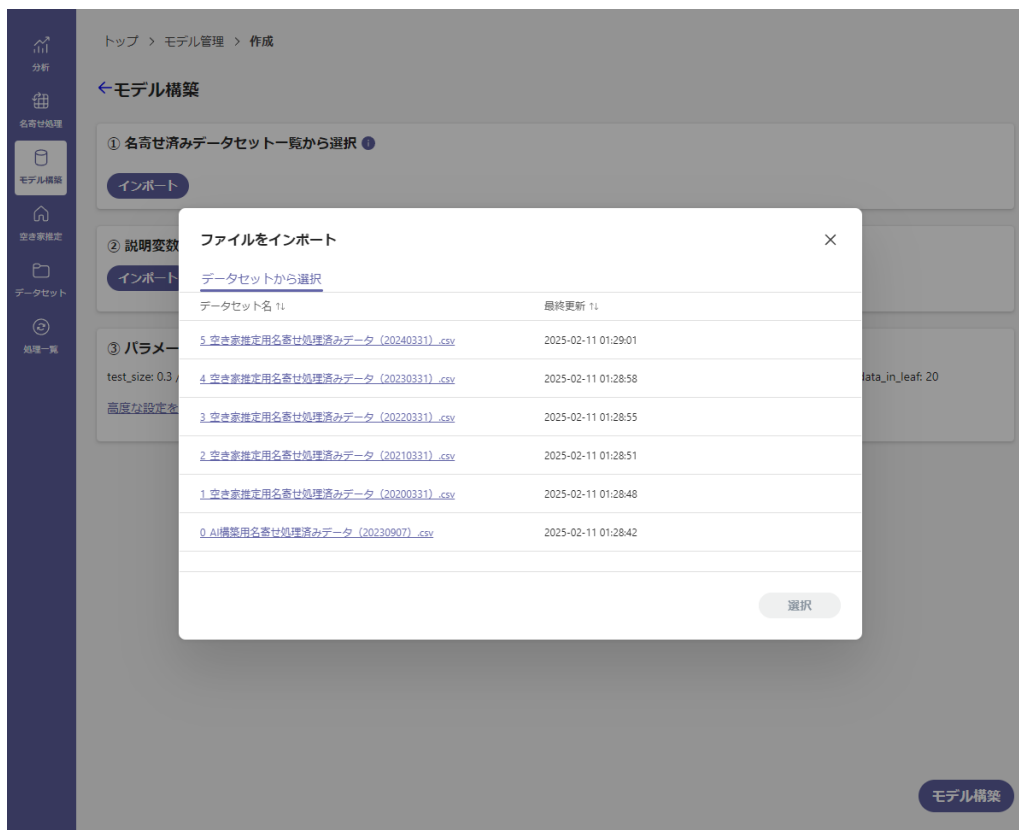


図 2-5-3-1-17 名寄せ処理済データ選択画面のイメージ

2-5-3-1-10. 【SC204】説明変数選択画面

- 画面の目的・概要
 - モデル構築に利用する説明変数を選択する
- 画面イメージ

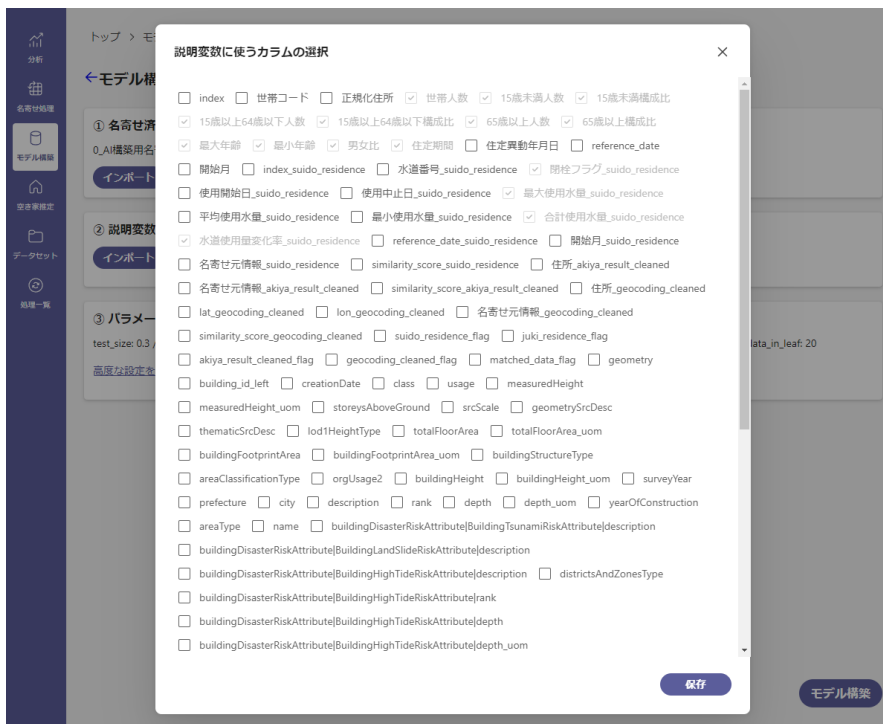


図 2-5-3-1-18 説明変数選択画面のイメージ

2-5-3-1-11. 【SC205】 高度な設定の変更画面

- 画面の目的・概要
 - モデル構築のための各種設定を変更する
- 画面イメージ

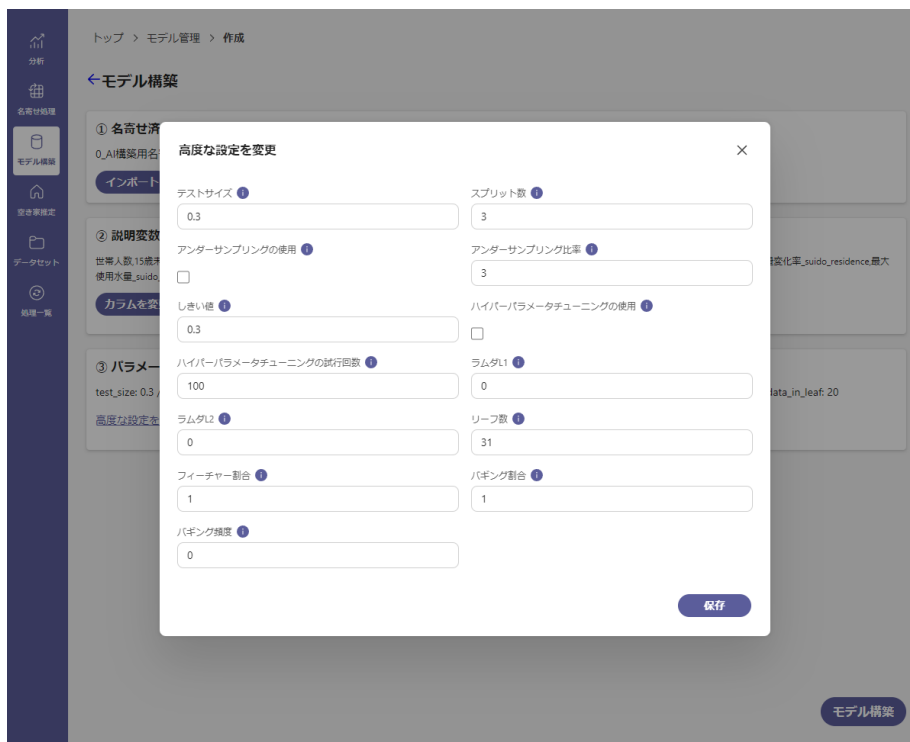


図 2-5-3-1-19 高度な設定の変更画面のイメージ

2-5-3-1-12. 【SC301】 空き家推定一覧画面

- 画面の目的・概要
 - 空き家推定を実行する
 - ユーザーが実行した空き家推定のステータスや結果を管理する
- 画面イメージ



図 2-5-3-1-20 空き家推定一覧画面のイメージ

2-5-3-1-13. 【SC301】 空き家推定一覧画面

- 画面の目的・概要
 - 空き家推定を実行する
 - ユーザーが実行した空き家推定のステータスや結果を管理する
- 画面イメージ



図 2-5-3-1-21 空き家推定一覧画面のイメージ

2-5-3-1-14. 【SC302】 空き家推定実行画面

- 画面の目的・概要
 - モデル・名寄せ処理済データ・地域集計用データの選択をカード UI 上で行い、空き家推定を実行する
- 画面イメージ

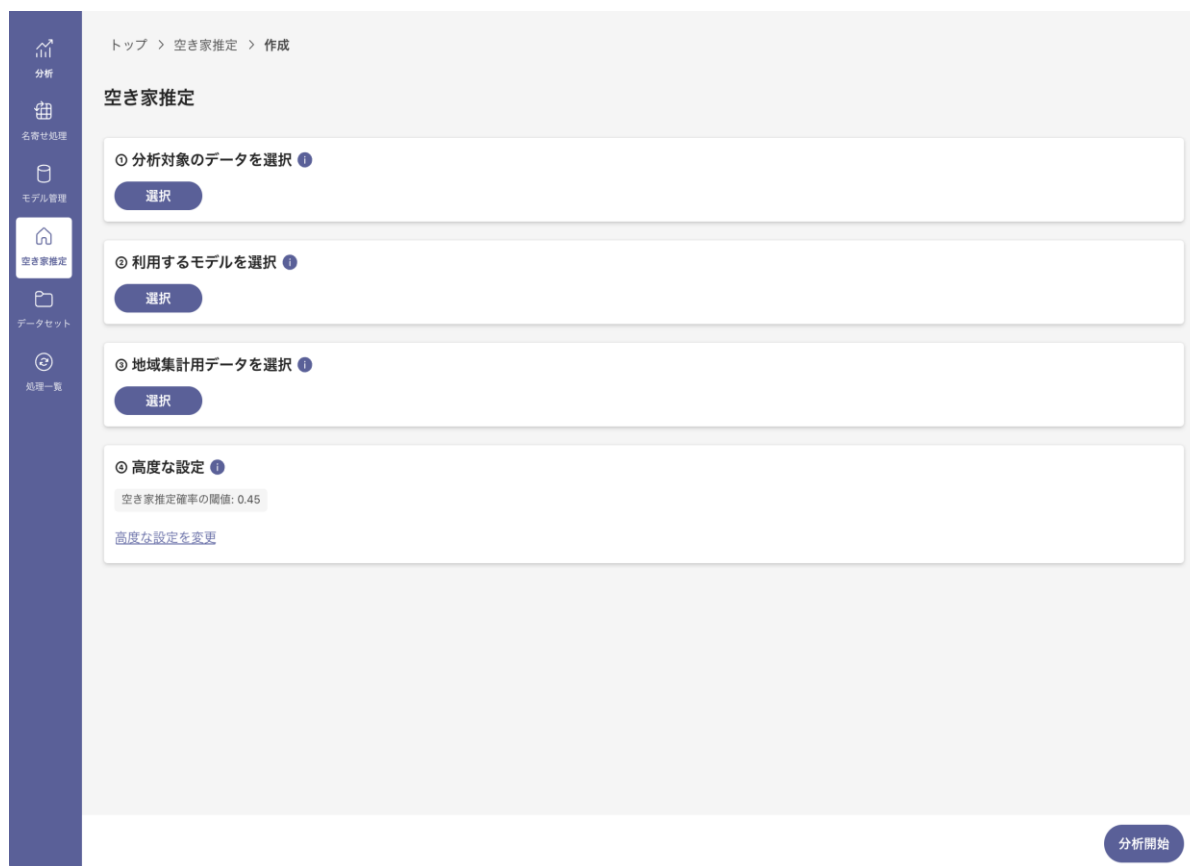


図 2-5-3-1-22 空き家推定実行画面のイメージ

2-5-3-1-15. 【SC401】 インプットデータ管理画面

- 画面の目的・概要
 - インプットデータを入力・管理・出力する
- 画面イメージ

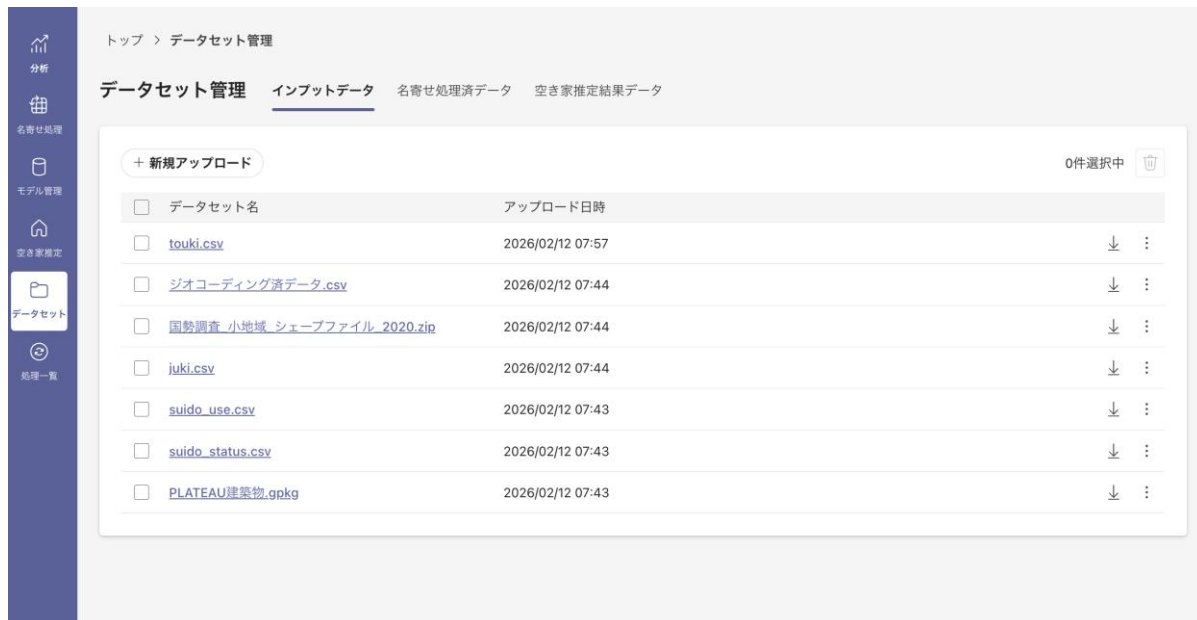


図 2-5-3-1-23 インプットデータ管理画面のイメージ

2-5-3-1-16. 【SC402】 名寄せ処理済データ管理画面

- 画面の目的・概要
 - 名寄せ処理済データを入力する
 - ユーザーが入力・保存した名寄せ処理済データを管理・出力する
- 画面イメージ



図 2-5-3-1-24 名寄せ処理済データ管理画面のイメージ

2-5-3-1-17. 【SC403】 空き家推定結果データ管理画面

- 画面の目的・概要
 - ユーザーが実行した空き家推定の結果データを管理・出力する
- 画面イメージ



図 2-5-3-1-25 空き家推定結果データ管理画面のイメージ

2-5-3-1-18. 【SC501】 処理一覧画面

- 画面の目的・概要
 - ユーザーが実行した名寄せ処理、モデル構築、空き家推定、出力のステータスを確認する
- 画面イメージ



図 2-5-3-1-26 処理一覧画面のイメージ

2-5-3-1-19. 【SC502】 名寄せ処理結果確認画面

- 画面の目的・概要
 - ユーザーが実行した名寄せ処理の結果を確認する
 - 必要に応じて再実行に進む
- 画面イメージ

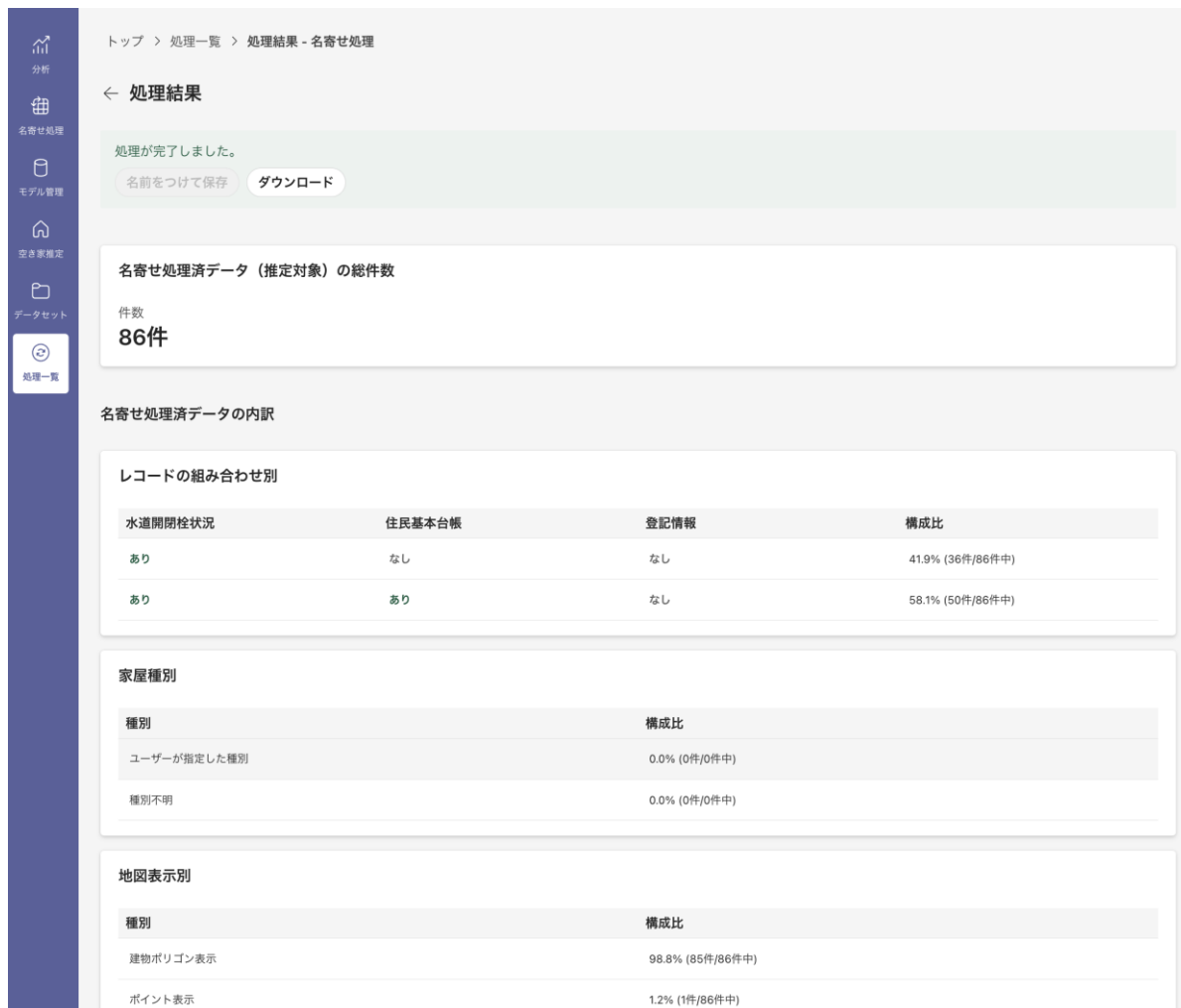


図 2-5-3-1-27 名寄せ処理結果確認画面のイメージ

2-5-3-1-20. 【SC503】モデル確認画面

- 画面の目的・概要
 - ユーザーがアップロードしたモデルの詳細を確認する
- 画面イメージ

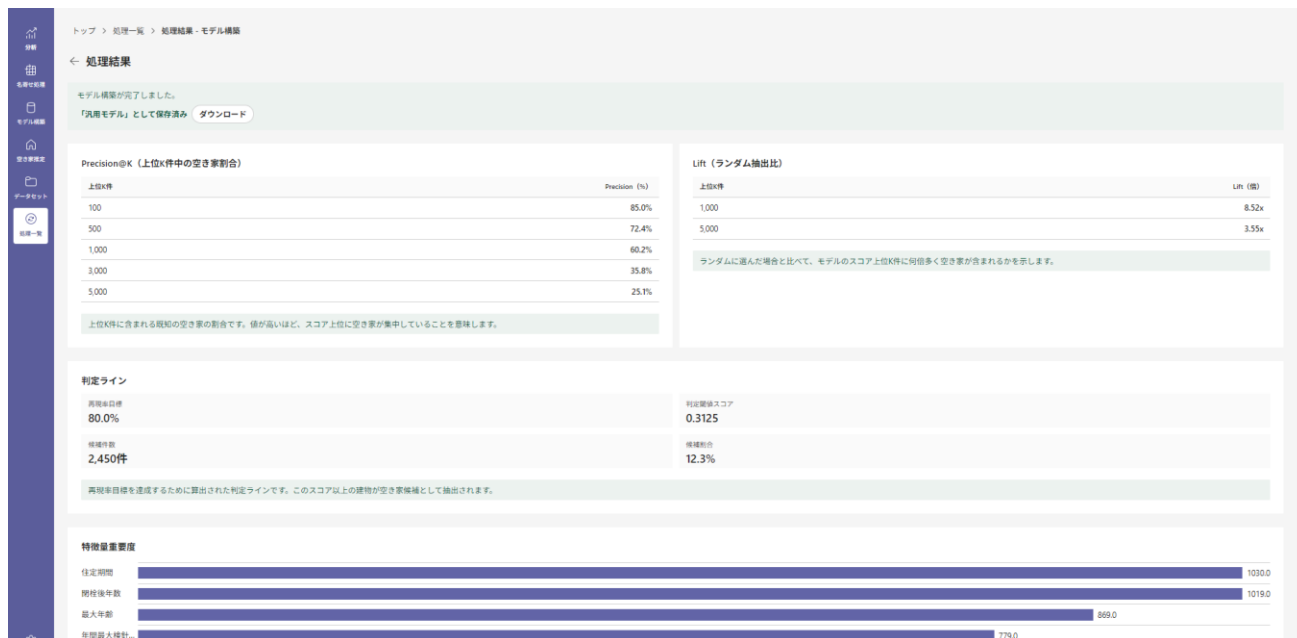


図 2-5-3-1-28 モデル構築結果確認画面のイメージ

2-5-3-1-21. 【SC504】ダウンロード結果確認画面

- 画面の目的・概要
 - ユーザーが実行した出力処理の結果を確認する
 - データをダウンロードする
- 画面イメージ

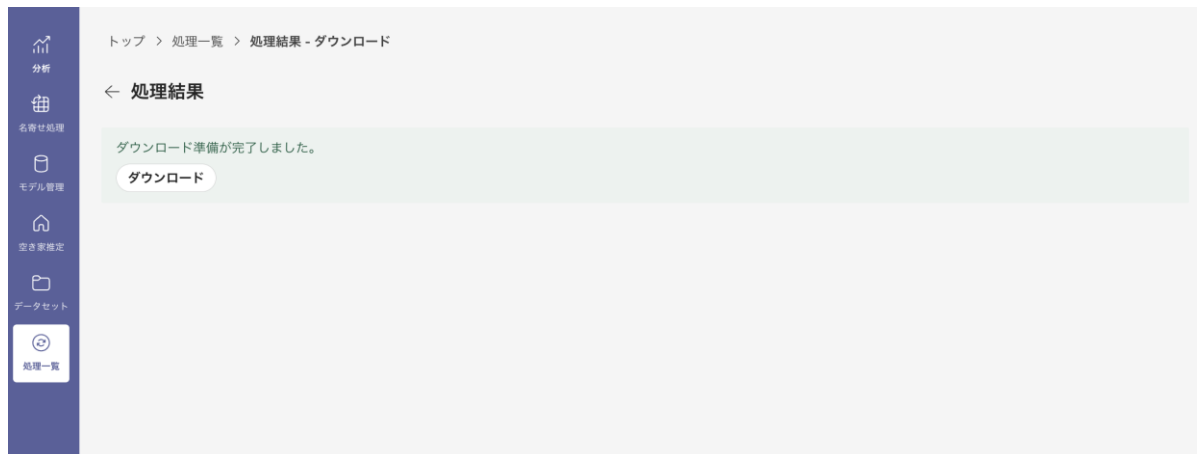


図 2-5-3-1-29 ダウンロード結果確認画面のイメージ

2-5-3-1-22. 【SC505】推定結果確認画面

- 画面の目的・概要
 - 空き家推定の完了を確認し、推定結果の概要を表示する
- 画面イメージ



図 2-5-3-1-30 推定結果確認画面のイメージ

2-5-3-2. LINKS SOMA CityGML Converter

2-5-3-2-1. 【SC601】メイン画面

- 画面の目的・概要
 - 変換を行う CityGML 形式データの選択と出力先を設定する
 - GeoPackage 形式への変換を実行する
- 画面イメージ

LINKS SOMA CityGML Converter

☒ 入力

☑ フォルダ選択 📄 ファイル選択

選択 フォルダが選択されていません

☒ 出力

選択 出力先が選択されていません

▶ 変換

図 2-5-3-2-1 メイン画面のイメージ

2-5-3-2-2. 【SC602】 処理ステータス画面

- 画面の目的・概要
 - ユーザーが実行した変換のステータスを確認する
 - 実行を途中でキャンセルする
- 画面イメージ



図 2-5-3-2-2 処理ステータス画面のイメージ

2-5-3-3. ジオコーディングツール

2-5-3-3-1. 【SC701】メイン画面

- 画面の目的・概要
 - ジオコーディングのテスト実行および本実行をする
 - 本実行結果をダウンロードする
- 画面イメージ

ジオコーディング

ジオコーディング用データ

データを選択

ファイルが選択されていません

住所に対応するカラムを選択

選択してください

利用するAPI

AWSジオコーディングAPI

NTTジオコーディングAPI

ABRジオコーディング

APIトークン

APIキーを入力

実行

※ テスト実行：CSVの1行目のデータのみを使用して、ジオコーディングが正常に動作するかを確認します。

テスト実行 本番実行

テスト実行 モック実行 前回途中から実行

テスト実行結果

まだ実行されていません

図 2-5-3-2-3 メイン画面のイメージ

3. 開発するシステム: 非機能要件 (NF)

3-1. 非機能要件一覧

3-1-1. LINKS SOMA

表 3-1-1-1 LINKS SOMA の非機能要件

カテゴリ	ID	項目	詳細
動作保証環境	NF001	実行環境	自治体職員が利用するネットワークのうち非インターネット接続かつ個人情報を扱うことが可能な環境で、実行形式ファイル (.exe) を起動して動作すること
	NF002	OS	Windows 10 Pro / Home (64 ビット版) 以上で動作すること
	NF003	マシンスペック	以下のマシンスペックで動作すること CPU : Intel Core i3 第 10 世代以上、または同等性能の AMD プロセッサ メモリ : 8GB 以上 GPU : Intel UHD Graphics 以上、または同等性能の GPU
性能	NF004	処理速度	デスクトップアプリケーションとして実行する常識的な時間として、1.5 分 / 1 万件をめどに各処理を完了する

3-1-2. LINKS SOMA CityGML Converter

LINKS SOMA と同様の非機能要件とする。

3-1-3. ジオコーディングツール

表 3-1-3-1 ジオコーディングツールの非機能要件

カテゴリ	ID	項目	詳細
動作保証環境	NF201	実行環境	自治体職員が利用するネットワークのうちインターネット接続が可能な環境で、実行形式ファイル (.exe) を起動して動作する。
	NF202	OS	NF002 と同様の非機能要件とする
	NF203	マシンスペック	NF003 と同様の非機能要件とする

3-2. 非機能要件詳細

3-2-1. 【NF001】 実行環境

- 本非機能要件を適用するシステム
 - LINKS SOMA 空き家推定システム
 - LINKS SOMA CityGML Converter
- 設定理由
 - 総務省が定義する自治体のネットワークモデル「三層分離」のうち個人情報を扱う業務が実施できる「LGWAN 接続系」を想定し、非インターネット接続環境での実行が可能な設定

3-2-2. 【NF002】 OS

- 本非機能要件を適用するシステム
 - LINKS SOMA 空き家推定システム
 - LINKS SOMA CityGML Converter
 - ジオコーディングツール
- 設定理由
 - 自治体職員の業務用端末として想定され、開発環境として最低限保証可能な OS を設定した

3-2-3. 【NF003】 マシンスペック

- 本非機能要件を適用するシステム
 - LINKS SOMA 空き家推定システム
 - LINKS SOMA CityGML Converter
 - ジオコーディングツール
- 設定理由
 - 実証対象地域の職員の業務用端末での実行が可能で、最低限の性能を保障できるスペックを設定した

3-2-4. 【NF007】 処理速度

- 本非機能要件を適用するシステム
 - LINKS SOMA 空き家推定システム
 - LINKS SOMA CityGML Converter
- 設定理由
 - デスクトップアプリケーションであるため処理速度はユーザーの端末の性能によることになるが、最低限の処理速度を設定した

3-2-5. 【NF201】 実行環境

- 本非機能要件を適用するシステム
 - ジオコーディングツール

- 設定理由
 - ジオコーディングを行うためにはインターネット接続環境下で実行する必要があるため

4. 実証調査に利用するデータ (DT)

4-1. 実証調査に利用するデータ一覧

表 4-1-1 実証調査に利用するデータ一覧

※朱文字：本実証で変換・作成するデータ

ID	データ名称	データ形式	出所
DT001	水道栓データ	CSV TXT Excel	実証調査協力自治体利用資料
DT002	水道使用量データ	CSV TXT Excel	実証調査協力自治体利用資料
DT003	住民基本台帳	CSV TXT Excel	実証調査協力自治体利用資料
DT004	登記情報	CSV Excel	実証調査協力自治体利用資料
DT005	推定対象選定用データ	CSV GeoPackage Shapefile	実証調査協力自治体利用資料 国土交通省
DT006	国勢調査小地域データ	Shapefile	e-Stat
DT007	地域集計用ポリゴン	Shapefile	e-Stat
DT008	空き家調査結果	CSV TXT Excel	実証調査協力自治体利用資料
DT101	建物ポリゴンデータ	GeoPackage Shapefile	国土交通省 国土地理院
DT102	ジオコーディング用データ	CSV	DT001
DT103	ジオコーディング済データ	CSV	DT101

DT201	正規化水道栓データ	CSV	DT001
DT202	正規化水道使用量データ	CSV	DT002
DT203	正規化住民基本台帳	CSV	DT003
DT204	正規化登記情報	CSV	DT004
DT205	正規化推定対象選定用データ	CSV	DT005
DT206	正規化ジオコーディング済データ	CSV	DT102
DT207	正規化空き家調査結果データ	CSV	DT008
DT301	住居単位水道データ	CSV	DT201～DT202
DT302	住居単位住民基本台帳	CSV	DT203
DT303	住居単位登記情報	CSV	DT204
DT401	テキストマッチングデータ	CSV	DT205、DT206、DT301～DT303
DT402	家屋単位 GIS データ	CSV	DT101、DT401
DT403	家屋種別判定後データ	CSV	DT401、または DT402
DT501	建物単位空き家推定結果データ	CSV	DT401、または DT403
DT502	地域単位空き家推定結果データ	CSV	DT501

4-2. 実証調査に利用するデータの詳細

4-2-1. 【DT001】水道栓データ

水道栓ごとの使用期間および所在地情報を管理するデータ。

表 4-2-1 水道栓データの必須カラム

カラム名称	説明	データ型
水道番号	水道栓を一意に識別するための番号	string
住所	水道栓が設置されている所在地	string
使用開始日	水道の使用を開始した日	date
使用中止日	水道の使用を中止した日	date

4-2-2. 【DT002】水道使用量データ

各水道栓の検針日ごとの使用水量を管理するデータ。

表 4-2-2 水道使用量データの必須カラム

カラム名称	説明	データ型
水道番号	水道栓データと連携するための識別 ID	int
検針年月日	水道メーターを検針した日	date
使用水量	当該検針期間における使用水量	int

4-2-3. 【DT003】 住民基本台帳

住民単位で世帯情報および居住履歴情報を管理するデータ。

表 4-2-3 住民基本台帳の必須カラム

カラム名称	説明	データ型
世帯番号	世帯を識別する番号	int
住所	住民の居住地住所	string
生年月日	生年月日	date
住定日	当該住所に住民登録した日	date
異動日	転出・死亡等により住民異動が発生した日	date
異動事由	異動の理由（例：転出、死亡、職権消除等）	string

4-2-4. 【DT004】 登記情報

家屋の登記上の所在・用途・構造・登記履歴等を管理するデータ。

表 4-2-4 登記情報の必須カラム

カラム名称	説明	データ型
住所	登記上の家屋所在地	string
登記構造	登記された建物の構造	string
登記日付	登記が行われた日	date

登記理由	登記が発生した理由	string
------	-----------	--------

4-2-5. 【DT005】 推定対象選定用データ

推定対象とする家屋の種別を推定したい場合に用意するデータ。

表 4-2-5 推定対象選定用データの必須カラム

カラム名称	説明	データ型
住所	家屋の住所	string
種別	家屋の種別	string

本実証においては、愛知県岡崎市・豊田市では【DT004】登記情報、愛知県豊橋市、埼玉県熊谷市では【DT101】建物ポリゴンデータに家屋の種別情報が含まれていたため、推定対象選定用データとして利用した。山口県下関市では、ゼンリンの建物ポイントデータを利用した。（建物ポイントデータの仕様詳細は公式サイト（<https://www.zenrin.co.jp/product/category/gis/contents/building-point/index.html?tab=6830>）を参照）

4-2-6. 【DT006】 国勢調査小地域データ

町丁（ちょうちょう）や字（あざ）などの小地域の範囲を示した境界データ

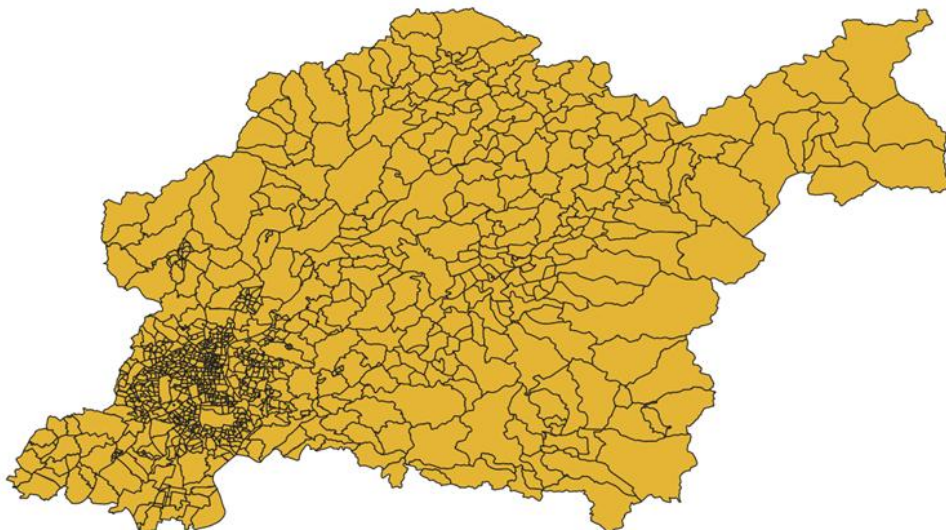


図 4-2-1 国勢調査小地域データ（町丁・字等）を地図描画した様子

4-2-7. 【DT007】 地域集計用ポリゴン

空き家推定結果をユーザーが指定する任意の地域単位に集計し、地域別集計データを生成するために用いるデータ。本実証では、【DT006】 国勢調査小地域データを地域集計用ポリゴンとして利用した。

4-2-8. 【DT008】 空き家調査結果

自治体職員または委託業者等による現有家屋の目視調査等の結果に関するデータ。本データに記載されている家屋を空き家と扱った。

表 4-2-6 空き家調査結果の必須カラム

カラム名称	説明	データ型
住所	空き家とみなした家屋の住所	string

4-2-9. 【DT101】建物ポリゴンデータ

家屋の形状を地図上に表示するために必要なデータ。本実証においては、以下2種類のデータを利用。

4-2-9-1. PLATEAU データ（利用自治体：愛知県岡崎市・豊田市・豊橋市、埼玉県熊谷市）

建物の形状をCGモデルで再現し、それに名称や用途・建設年都市などの活動情報を付与した地理空間データ。本実証での利用にあたっては、CityGML形式のPLATEAUデータをLINKS SOMA CityGML Converterにて表に変換した。

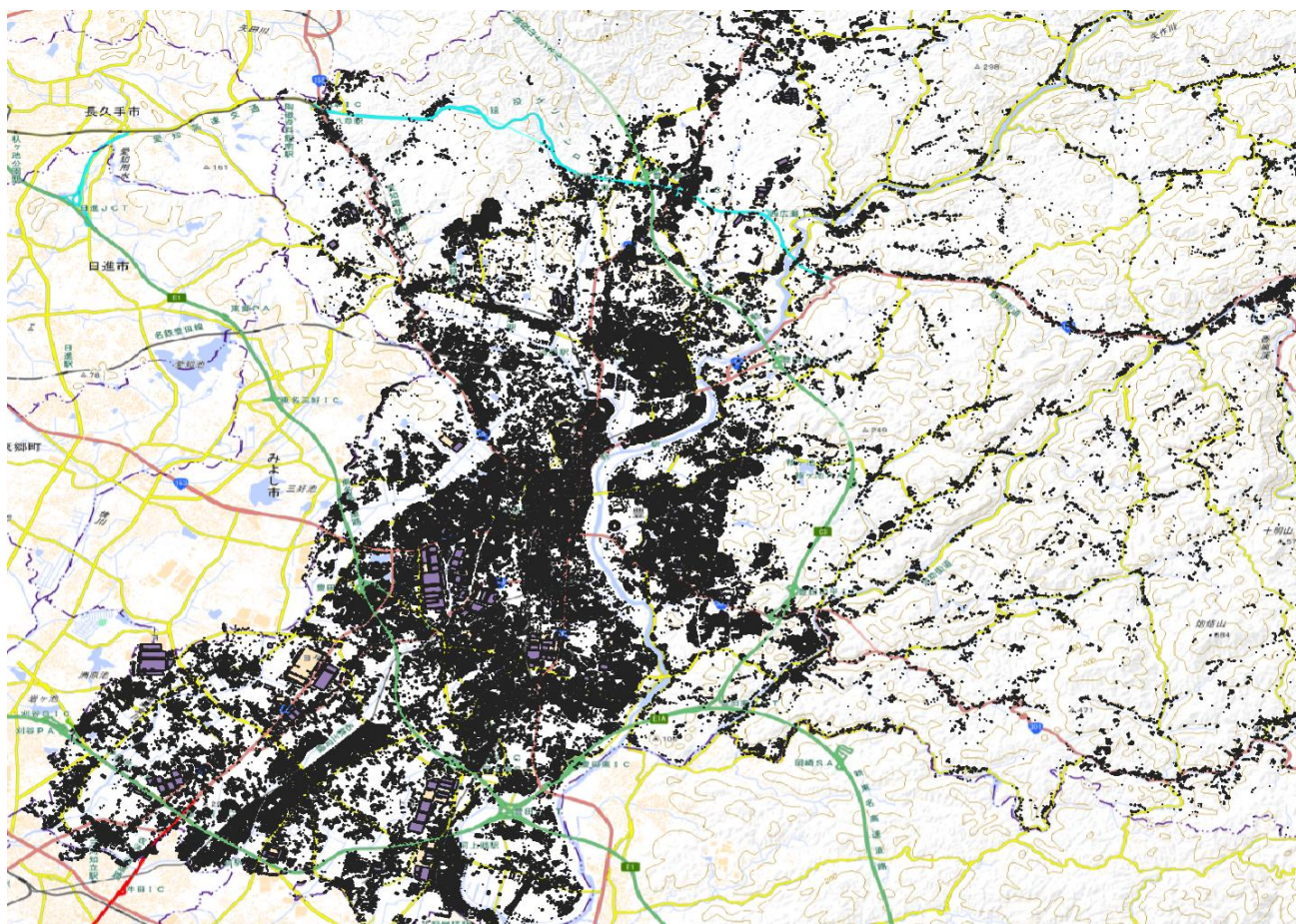


図 4-2-2 PLATEAU 建物ポリゴンデータを地図描画した様子（背景地図：地理院地図）

4-2-9-2. 基盤地図情報 建物データ (利用自治体：山口県下関市)

国土地理院が整備する日本全域をカバーする高精度なデジタル地図(ベクトル形式)において、建物の外形(輪郭)と位置を表したデータ。都市部では縮尺 1/2500 相当、それ以外の地域では 1/25000 相当の精度で公開。本実証での利用にあたっては、xml 形式の基盤地図情報を GIS ソフトにて GeoPackage 形式に変換した。

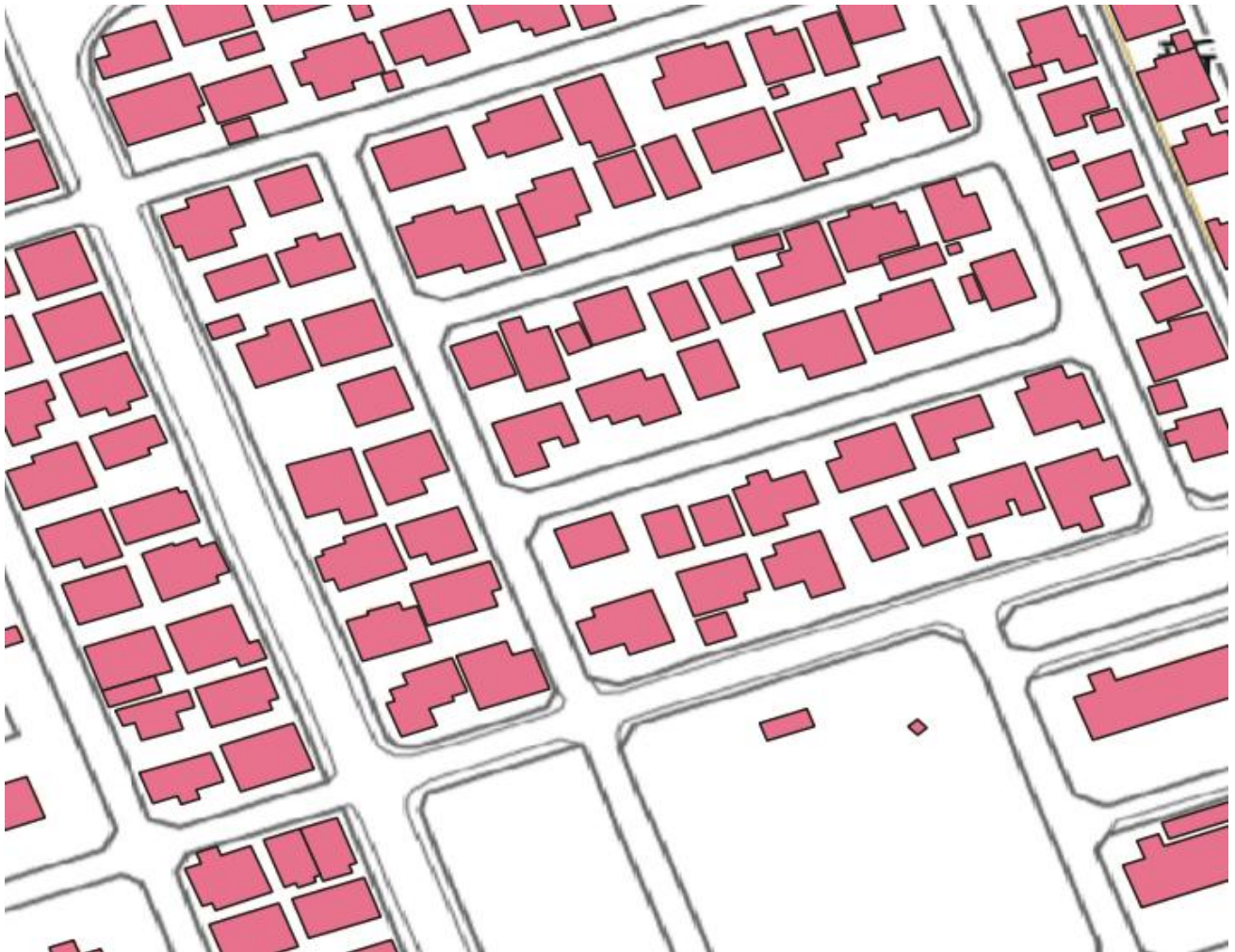


図 4-2-3 基盤地図情報 建物データを地図描画した様子 (背景地図：地理院地図)

4-2-10. 【DT102】 ジオコーディング用データ

ジオコーディングを実行する必要がある場合に用意するデータ。

表 4-2-7 ジオコーディング用データの必須カラム

カラム名称	説明	データ型
住所	ジオコーディングの対象となる住所	string

4-2-11. 【DT103】 ジオコーディング済データ

【DT102】ジオコーディング用データに対して、住所列を対象にジオコーディングを行い、緯度経度情報を付与したデータ。

表 4-2-8 ジオコーディング済データの必須カラム

カラム名称	説明	データ型
住所	ジオコーディングの対象となった住所	string
緯度	住所に対応する緯度	string
経度	住所に対応する経度	string

4-2-12. 【DT201】 正規化水道栓データ

【DT001】水道栓データにおける住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。

4-2-13. 【DT202】 正規化水道使用量データ

【DT002】 水道使用量データにおける日付型データの表記揺れを修正したデータ。

4-2-14. 【DT203】 正規化住民基本台帳

【DT003】 住民基本台帳における住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。

4-2-15. 【DT204】 正規化登記情報

【DT004】 登記情報における住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。

4-2-16. 【DT205】 正規化推定対象選定用データ

【DT005】 推定対象選定用データにおける住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。（推定対象選定用データが CSV 形式の時のみ生成される）

4-2-17. 【DT206】 正規化ジオコーディング済データ

【DT103】 ジオコーディング済みデータにおける住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。

4-2-18. 【DT207】 正規化空き家調査結果

【DT008】 空き家調査結果における住所の表記揺れを修正したデータ。正規化後の住所は「正規化住所」として、元データのカラムに追加される。

4-2-19. 【DT301】住居単位水道データ

【DT201】正規化水道栓データと【DT202】正規化水道使用量データを結合し、住居単位に成形したデータ。

表 4-2-9 住居単位水道データのカタログ

カラム名称	説明	データ型
住所	水道栓の住所	string
正規化住所	正規化後の住所	string
水道番号	水道栓ごとに付与された識別番号	int
水道使用開始日	水道の使用開始日	date
水道使用中止日	水道の使用中止日	date
閉栓フラグ	閉栓している場合に付与されるフラグ	boolean
平均検針水量	推定日から1年以内の平均検針水量	float
年間最大検針水量	推定日から1年以内で最大の検針水量	float
年間合計検針水量	推定日から1年以内の合計検針水量	float
直近4ヶ月の使用量増減率	直近期間の水量増減率（直近2ヶ月÷直近4ヶ月）	float
検針水量（11・12ヶ月前）	推定月の11～12ヶ月前の検針水量	float
検針水量（9・10ヶ月前）	推定月の9～10ヶ月前の検針水量	float
検針水量（7・8ヶ月前）	推定月の7～8ヶ月前の検針水量	float

検針水量（5・6ヶ月前）	推定月の5～6ヶ月前の検針水量	float
検針水量（3・4ヶ月前）	推定月の3～4ヶ月前の検針水量	float
検針水量（1・2ヶ月前）	推定月の1～2ヶ月前の検針水量	float
使用量データあり	水道使用量データが存在する場合に付与されるフラグ	boolean
ゼロ使用期数	使用量がゼロの期間数	int
最小水道使用量	最小の水道使用量	float
閉栓後年数	閉栓から推定日までの経過年数	float
使用量データなし	水道使用量データが存在しない場合に付与されるフラグ	boolean
前半平均使用水量	推定月の6～12ヶ月前における検針水量の平均	float
後半平均使用水量	推定月の1～6ヶ月前における検針水量の平均	float
半期変化率	後半平均使用水量 ÷ 前半平均使用水量	float
直近使用水量	直近2期の検針での平均使用量	float

4-2-20. 【DT302】住居単位住民基本台帳

【DT202】正規化住民基本台帳を世帯ごとに集計後、住居単位に成形したデータ

表 4-2-10 住居単位住民基本台帳のカラム

カラム名称	説明	データ型
住所	住所	string
正規化住所	正規化後の住所	string
世帯番号	世帯を識別する番号	int
最大年齢	世帯内の最大年齢	int
最大年齢欠損	最大年齢が欠損かどうか	boolean
65歳以上人数	推定日時点で65歳以上の人数	int
15歳未満人数	推定日時点で15歳未満の人数	int
住定期間	住定日から基準日までの経過日数。世帯内で最も長い人の値を採用	int
世帯人数	推定日時点で居住が確認された世帯の人数	int
死亡人数	世帯内で記録された死亡件数	int
転入数	世帯内で記録された転入件数	int
転出・転居数	世帯内で記録された転出・転居件数	int
削除イベントあり	世帯内で削除イベントがあったかどうか	boolean

転出イベント数	世帯内で記録された転出の件数	int
最終異動後経過年数	世帯内で最後に異動があったから経過した年数	float
最終異動経過年数欠損	最終異動年数が欠損かどうか	boolean
独居高齢者	高齢者の単身世帯かどうか。最大年齢 ≥ 65 、かつ世帯人数=1の場合に付与される。	boolean
死亡後入居者なし	死亡が記録された時点以降に新たな入居者が記録されていない場合に付与されるフラグ	boolean
世帯縮小率	世帯人数の減少率	float

4-2-21. 【DT303】住居単位登記情報

【DT204】正規化登記情報から住居単位のデータに成形したデータ。

表 4-2-11 住居単位登記情報のカラム

カラム名称	説明	データ型
住所	元住所	string
正規化住所	正規化後の住所	string
登記構造	登記情報に記載された建物構造	string
登記事由発生日	登記事由が発生した日	date
登記事由発生からの経過日	登記事由発生日から推定日までの経過日数	int
コンクリートブロック造	当該構造の場合に付与されるフラグ	int
煉瓦造	当該構造の場合に付与されるフラグ	int
鉄筋コンクリート造	当該構造の場合に付与されるフラグ	int
鉄骨造	当該構造の場合に付与されるフラグ	int
木造	当該構造の場合に付与されるフラグ	int
土造	当該構造の場合に付与されるフラグ	int
その他構造	上記以外の構造の場合に付与されるフラグ	int
相続	登記事由が相続の場合に付与されるフラグ	int

贈与	登記事由が贈与の場合に付与されるフラグ	int
売買	登記事由が売買の場合に付与されるフラグ	int
差押	登記事由が差押の場合に付与されるフラグ	int

4-2-22. 【DT401】 テキストマッチングデータ

【DT205】～【DT206】、【DT301】～【DT303】のデータを正規化住所をキーに結合したデータ。本データは結合元のデータが保有していたカラムを全て保持、かつ以下のカラムが付与される。

表 4-2-12 テキストマッチングデータのカラム（追加のみ記載）

カラム名称	説明	データ型
住民データ有無フラグ	住民データが存在するかどうかのフラグ	boolean
一人当たり検針水量	世帯員 1 人あたりの水道使用量	xfloat
複合ルールスコア	複数条件を組み合わせたスコア	float

4-2-23. 【DT402】家屋単位 GIS データ

【DT401】テキストマッチングデータに、【DT101】建物ポリゴンデータを空間結合したデータ。【DT401】テキストマッチングデータに以下のカラムが付与される。

表 4-2-13 家屋単位 GIS データのカラム（テキストマッチングデータとの差分のみ記載）

カラム名称	説明	データ型
土地区分	建築物が立地する土地の区域区分	string
建築物に紐づく不動産 ID の情報	建築物に紐づく不動産識別 ID	string
構造種別	建築物の構造種別（GIS 側属性）	string
標高	建物位置の標高値	int
建物付属物	小屋根・外階段・バルコニー等の付属設備情報	string
地上階の階数	地上階の階数	int
建築された年	建築年	int
解体された年	解体年	int
建物ポリゴンジオメトリ情報	建物ポリゴンのジオメトリ情報	string
建物 ID	建物を一意に識別する ID	int
経度	建物の経度	int
緯度	建物の緯度	int
空間レベル	ジオコーディング結果の空間レベル	string

空間精度	ジオコーディング結果の精度	string
地域名称コード	地域集計用の地域コード	string
地域名称	地域名称	string
住居 ID	システム内部で付与された番号	int
建物種別	推定対象選定用データから結合された建物種別	string
家屋種別判定不可フラグ	家屋種別が不明な場合に 1	int
高潮浸水時の浸水の深さ	高潮浸水想定時の浸水深	int
内水浸水リスク深さ	内水浸水リスクの深さ	int
洪水浸水想定時の深さ	洪水浸水想定時の浸水深	int
土砂災害警戒区域の有無	土砂災害警戒区域に該当する場合に 1	int
指定河川の名称	関連する指定河川の名称	String
住民データ有無フラグ	住居単位住民基本台帳が結合できない場合に付与されるフラグ	int
水道データ有無フラグ	住居単位水道データが結合できない場合に付与されるフラグ	int
登記情報データ有無フラグ	住居単位登記情報が結合できない場合に付与されるフラグ	int
空き家フラグ	空き家調査結果と結合できた場合に付与される	int

	フラグ	
--	-----	--

4-2-24. 【DT403】家屋種別判定後データ

【DT401】テキストマッチングデータ、または【DT402】家屋単位 GIS データに含まれる家屋種別のうち、ユーザーが指定した種別、及び種別不明のデータを抽出したデータ。

4-2-25. 【DT501】建物単位空き家推定結果データ

建物単位の空き家推定結果データ。家屋単位 GIS データに以下のカラムが付与される。

表 4-2-14 建物単位空き家推定結果データのカラム（家屋単位 GIS データとの差分のみ記載）

カラム名称	説明	データ型
空き家確率	空き家である確率（0～1）	int
空き家推定結果	空き家判定結果（0：非空き家、1：空き家）	int
空き家推定データ作成日	推定結果データ作成日	date

4-2-26. 【DT502】 地域単位空き家推定結果データ

【DT501】 建物単位空き家推定結果データを、【DT007】 地域集計用ポリゴンで集計したデータ。

表 4-2-15 地域単位空き家推定結果データのコラム

コラム名称	説明	データ型
地域データ番号	地域データを一意に識別する番号	int
地域データ出力番号	出力時に付与された管理用番号	int
推定日	空き家推定の基準日	date
若年層率（15歳以下人口）	地域内人口に占める15歳以下人口の割合	int
高齢者率（65歳以上人口）	地域内人口に占める65歳以上人口の割合	int
家屋数	地域内の建物数	int
地域面積	地域ポリゴンの面積	int
地域ポリゴンジオメトリ	地域のポリゴン形状情報（WKT形式）	string
地域コード	地域を識別するコード	string
作成日	当該レコードの作成日時	date
更新日	当該レコードの最終更新日時	date
推定空き家数	推定により算出された空き家数	int
推定空き家割合	地域内家屋数に対する推定空き家の割合	int

R7_LINKS07_行政情報を活用した空き家推定技術の実証調査_要件定義資料

地域名称	地域の名称	string
------	-------	--------

用語集

A) アルファベット順

表 9-1 用語集（アルファベット順）

No.	用語	説明
1	AI	コンピューターが人間のように考えたり、学んだり、問題を解決したりする能力を指す。機械学習等の技術を使って、パターンを認識したり、自動でタスクを実行したりすることができる。
2	AI モデル	データを学習して新しいことを判断・予測できるようになった AI の頭脳のこと。
3	API	「Application Programming Interface」の略。異なるソフトウェア同士が情報をやり取りするための「決まりごと」や「仕組み」のこと。
4	API キー	API を安全に使うための「認証コード」のこと。
5	AWS	「Amazon Web Services」の略。Amazon が提供するクラウドコンピューティングサービスの総称。
6	CityGML	「City Geography Markup Language」の略。都市や地域の 3D モデルを表現するための国際標準規格。建物や道路、公園等の都市要素を、地理空間データとともに 3 次元で表現することができる。
7	CSV	「Comma-Separated Values (カンマ区切り値)」の略。表のデータをカンマで区切って保存するシンプルなファイル形式。ファイルの拡張子は「.csv」
8	Electron	Chromium (Web ブラウザエンジン) と Node.js (サーバーサイド JavaScript 実行環境) を組み合わせ、Web 技術でデスクトップアプリケーションを構築するフレームワーク。
9	F 値	AI モデルがどれだけ「正確かつバランスよく」予測できているかを評価する指

		標のこと。適合率と再現率の調和平均で算出される。
10	GeoJSON	「Geographic JavaScript Object Notation」の略称。地理空間データを表現するための JSON ベースのオープンスタンダード形式で、位置情報や地形特性等を効率的に記述できる。
11	GeoPackage	GIS で地理データを 1 つのファイルにまとめて効率的に保存・管理できるファイルベースのデータベース形式。ファイルの拡張子は「.gpkg」
12	GIS	「Geographic Information System」の略で、地理情報システムとも呼ばれる。地理的位置を手がかりに、位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術のこと。
13	LGWAN 接続系	地方自治体が利用する情報ネットワークシステム「LGWAN」を通じて、外部のシステムやネットワークと安全にデータのやり取りを行うための仕組みのこと。
14	LightGBM	「Light Gradient Boosting Machine」の略。高速かつ高精度な勾配ブースティング（弱い予測器を順番に組み合わせて精度を高める手法）フレームワーク。テーブルデータの分類・回帰タスクで広く使われる機械学習アルゴリズム。
15	N-gram	テキストを N 文字ずつ区切って部分文字列を生成する手法。住所の類似度比較等に利用される。例: 「東京都」を 2-gram で分割すると「東京」「京都」の 2 つになる。
16	ORM	「Object-Relational Mapping」の略。プログラムのデータ構造とデータベースのテーブルを自動的に対応づけ、SQL を直接記述せずにデータベース操作を可能にする仕組みのこと。
17	OSS	「Open Source Software」の略称。プログラムのソースコードが一般に公開されていて、無償で利用できるソフトウェアのこと。
18	pkl	Python Pickle File 形式のファイルのこと。Python のオブジェクトをシリアル

		化して保存するためのファイル形式。
19	PLATEAU	国土交通省が推進する日本全国の 3D 都市モデルを整備し、オープンデータとして提供する、まちづくりの DX の実現を目指したプロジェクトのこと。
20	PU Bagging	機会学習において、「正例 (Positive)」と「ラベルなしデータ」だけで学習する手法。「ラベルなしデータ」をランダムにサンプリングして複数のモデルを構築し、それぞれに予測させることで安定した分類性能を得ることができる。
21	QGIS	オープンソースで提供されている GIS ソフトウェアのこと。誰でも無料でダウンロードして利用できる。
22	Shapefile	GIS で地理空間データを保存・共有するための標準的なファイル形式の一つ。複数のファイル (.shp, .shx, .dbf 等) で構成される。
23	UI	「User Interface」の略称。ユーザーがコンピュータシステムとやり取りするための視覚的な要素や操作方法の総称。
24	UTF-8	文字をコンピュータで扱う際の文字コードの一つ。世界中のさまざまな言語の文字を统一的に表現できる仕組みのこと。
25	UX	「User Experience」の略称。ユーザーが製品やサービスを利用する際の全体的な体験を指す。使いやすさ、効率性、満足度等の要素を包括する。
26	WGS84	「World Geodetic System 1984」の略。緯度・経度で位置を表す経緯度座標系のこと (EPSG:4326)。GPS をはじめ広く利用される世界測地系の一つ。
27	xml	「Extensible Markup Language」の略称。マークアップ言語の一種で、データの構造を階層的に表現し、異なるシステム間でのデータ交換に広く使用されている。
28	ZIP	ファイルを圧縮してサイズを小さくしたり、複数のファイルを 1 つにまとめたりする形式や仕組みのこと。拡張子は「.zip」

B) 五十音順

表 9-2 用語集（五十音順）

No.	用語	説明
1	空き家、空家	居住その他の使用がなされていないことが常態である建築物のことを指す（空家等対策の推進に関する特別措置法2条より抜粋）。具体的には、1年間を通して人の出入りの有無や、水道・電気・ガスの使用状況等から総合的に見て「空家」かどうか判断する、とされる。
2	空き家調査	空き家の実態を調査することにより、国及び地方公共団体における空き家に関する基本的施策を推進する上での基礎資料を得ることを目的として実施される調査。調査票とデジタルカメラを持参して現地調査による把握が一般的。
3	アドレス・ベース・レジストリ	デジタル庁が整備する、住所や位置情報等の地理的アドレスデータを一元管理する誰でも無料で利活用可能な住所・所在地関係データベース。
4	カラム	表の「縦の列」で、同じ種類の情報をまとめて管理するもの。たとえば、「名前」や「住所」等、同じ項目を入れる「場所」のこと。
5	機械学習	コンピューターに大量のデータを読み込ませ、データ内に潜むパターンを学習させることで、未知のデータを判断するためのルールを獲得することを可能にするデータ解析技術。
6	教師データ	AIが機械学習に利用するためのデータのこと。正解ラベル（空き家か否か等）が付与された学習用データ。
7	空間結合	位置情報を基に、2つの空間データ（位置情報を含むデータ）を結びつける手法。
8	結合率	結合が成功した割合のこと。例えば、100件のデータのうち80

		件が結合できた場合、結合率は 80%。
9	交差結合	データベース操作のひとつで、テーブルのすべての組合せを作成する結合方法のこと。
10	勾配ブースティング	学習データから段階的に予測精度を高める手法。
11	勾配ブースティング決定木	GBDT。複数の決定木を順番に組み合わせて高精度な推定を行う手法。
12	最近傍結合	あるデータ点に最も近い他のデータ点を見つけるための方法のこと。
13	再現率	実際に正解であるものを AI がどれだけ見逃さずに予測できたかを示す指標のこと。
14	閾値	しきい値。判定基準となる値のこと。本システムでは、空き家確率に対して複数の閾値を設け、空き家判定ラベルを付与する。
15	ジオコーディング	住所テキストから対応する緯度・経度の座標情報を取得する処理のこと。
16	ジオメトリ	地図上で「場所」や「形」を表すためのデータのこと。点・線・面（ポリゴン）で構成される。
17	正解率	AI が出した予測結果が、実際の正解とどれだけ一致しているかを示す指標のこと。
18	説明変数	AI のモデルで「結果に影響を与える原因や要因として考えるデータ」のこと。予測に使う入力項目。
19	ダッシュボード	データや情報を一目でわかりやすく確認できるようにした「情報のまとめ画面」のこと。

20	適合率	AI が予測した結果が、どれだけ正しかったかを示す指標のこと。
21	テキストマッチング	複数のデータ内にあるテキストのうち、一致する部分を見つける処理のこと。住所文字列の類似度を計算し、同一住所を特定する。
22	統合型 GIS	GIS の機能を一つのプラットフォームにまとめて提供し、さまざまな地理空間データの作成、管理、分析、表示を効率的に行えるようにしたシステムのこと。
23	特異度	実際に正解でないものを、AI がどれだけ正しく正解ではないと判断できたかを示す指標のこと。
24	特徴量	機械学習モデルに入力するデータ項目のこと。説明変数とも呼ばれる。本システムでは、世帯情報・水道使用量・建物属性等が特徴量にあたる。
25	名寄せ	データベースに存在しているデータを名前、住所、電話番号等の情報から読み取り、同一人物や同一企業等の重複しているデータをひとつにまとめる作業、データ処理。
26	ノーコード	プログラミングをしないでシステム等を開発する手法。
27	バッファ	地理データにおいて、ある地物（建物、道路等）から一定距離以内の領域を示す範囲データ。空間結合時に「最も近い建物」を判定する際に利用される。
28	ハイパーパラメーター	機械学習モデルの学習過程を制御するために、学習開始前から設定されるパラメータのこと。
29	BI ツール	「Business Intelligence」ツールの略。蓄積されたデータを分析・可視化し、意思決定を支援するソフトウェアのこと。

30	ビュー	データを視覚的に表示する形式のこと。本システムでは地図・棒グラフ・折れ線グラフ・円グラフ・テーブルの5種類を提供する。
31	ピボットテーブル	データの行と列を入れ替えて集計する表形式。水道使用量データを住居単位に変換する際などに利用される。
32	平面直角座標系	日本全国を19の区域に分割し、各区域にXY座標を割り当てた座標系。測量や地理情報処理で使用され、EPSGコード（EPSG:2443～EPSG:2461）で参照される。
33	編集距離	2つの文字列間の違いを数値化した指標。レーベンシュタイン距離とも呼ばれる。
34	ポイントデータ	GISにおいて、緯度・経度の一点で表現される地理データのこと。水道栓の位置情報等が該当する。
35	ポリゴン	線で囲まれた多角形の面データのこと。GISでは、ビルや家の形をポリゴンで表す。
36	目的変数	AIのモデルが予測する対象のデータ。本システムでは「空き家か否か」のフラグ（0または1）が目的変数にあたる。
37	モデル構築	インプットデータを使ってルールやパターンを見つけ、そのルールを使って新しいデータに基づいて何かを判断したり、予測したりする仕組みを作ること。
38	レーベンシュタイン距離	2つの文字列がどれだけ異なるかを測る指標。一方の文字列をもう一方に変換するために必要な「挿入・削除・置換」の最小回数で表す。住所の類似度判定に利用される。
39	ワークブック	本システムにおいて、分析用の地図・グラフ・テーブル等の表示設定をまとめて管理するファイル単位のこと。



空き家推定システム LINKS SOMA

システム設計書

発行日: 2026年3月

委託者: 国土交通省 総合政策局 情報政策課

受託者: 株式会社ユーカリヤ

マイクロベース株式会社