

公的データの機械学習による空き家の推定手法開発と発生予測分析

主査 稲田 浩也^{*1}

委員 三浦 研^{*2}

空き家問題に対する解決策として機械学習技術の活用を提案し、神戸市全域を対象としての空き家の推定手法開発と発生予測分析を行った。複数の機械学習手法の中から CatBoost が最も有効であることを示し、物件情報や電力データから空き家分布の高精度の予測結果が得られた。また、この予測モデルを活用し 2050 年までの空き家分布の分布傾向を推定し、神戸市の中で空き家が増加する可能性の高いエリアを詳細に特定した。さらに、機械学習の解釈手法である SHAP を用いてモデルの予測過程を解釈し、空き家発生に寄与する主要な要因を明らかにした。これまでは自治体全域で単一の対策を行われていたが、本研究の成果により、従来よりも狭い範囲で空き家問題に対処することが可能になることが期待される。

キーワード：1) 空き家, 2) 機械学習, 3) 電力データ, 4) XGBoost, 5) LightGBM,
6) CatBoost, 7) 機械学習の解釈手法, 8) SHAP

DEVELOPMENT OF MACHINE LEARNING-BASED ESTIMATION METHOD FOR VACANT HOUSES USING PUBLIC DATA AND PREDICTIVE ANALYSIS OF OCCURRENCE

Ch. Kouya Inada

Mem. Ken Miura

This study proposes the use of machine learning technology as a solution to the vacant house problem, conducting estimation and predictive analysis of vacant houses across Kobe City. It demonstrates that CatBoost is the most effective among several machine learning methods, achieving high-accuracy predictions of vacant house distributions based on property information and electricity data. The study forecasts the trends of vacant house distributions up to 2050, identifying areas within Kobe City with a high potential for increased vacancies. Furthermore, by employing the SHAP interpretative method, it clarifies the main factors contributing to the occurrence of vacant houses.

1. 序論

1.1 研究の背景

近年、我が国では空き家の数が増加の一途をたどっており、国土交通省の報告によると、1998 年の 576 万戸から 2018 年には 846 万戸へと、20 年間で約 1.5 倍に増加している^{文1)}。特に、「その他空き家」に分類される空き家の増加が顕著であり、同期間において約 1.9 倍の 347 万戸へと拡大している^{文1)}。「その他空き家」には、長期間居住されていない家や、建て替え予定の家など、賃貸や売却を目的としない住宅が含まれる^{文1)}。これらの空き家の増加は、老朽化や放火による火災リスクを高めると考えられる。

この問題に対処するため、国は「空家等対策の推進に関する特別措置法（平成 26 年法律第 127 号）^{文2)}」や「新たな住生活基本計画（令和 3 年 3 月 19 日閣議決定）^{文3)}」を通じて、空き家の情報収集と活用促進を図っている。

地方自治体も、これらの法律に基づき、空き家の有効活用に向けた情報収集や措置を講じることが可能となったが、これらの対策は主に個別の建物や所有者を対象としており、地域全体を視野に入れた包括的な対策の展開には至っていないのが現状である。

空き家問題は、単に建物の老朽化だけでなく、周辺地域の人口動態や道路整備の状況といった地域特性と深く関連しているため^{文4)}、個別の対象だけでなく、地域全体を考慮した取り組みが求められる。

また、市町村には空き家等のデータベース整備に関する努力義務が課せられている^{文2)}。多くの自治体では、空き家の分布を把握するための悉皆調査やサンプリング調査から全体の空き家の分布を推計する手法がとられている^{文5)} ^{文6)} が、特に大規模な自治体では、全域にわたる悉皆調査は現実的ではなく、サンプリングによる調査では偏りが生じる可能性がある。これにより、空き家の実

^{*1} 京都大学大学院博士後期課程 ^{*2} 京都大学大学院 教授 博士（工学）

際の分布傾向を正確に把握することが難しくなり、結果として効果的な対策の策定が妨げられる可能性がある。

現在の空き家の分布を正確に把握することは、将来の空き家の発生を予測する上での基礎となる。さらに、空き家の発生を抑制し、未利用地がランダムに増加することによる「都市のスポンジ化」を防ぐためには、様々な分析手法やデータベースを活用し、より高精度の空き家予測モデルを開発することが求められる。こうしたアプローチによって、未来の空き家問題に対しても、予防的かつ効果的な対策を講じることが可能になると考えられる。

1.2 既往研究

これまでの空き家に関する研究は数多くあり、その研究手法は現地調査やアンケート調査、聞き取り調査が多数を占めている^{文7)}。秋山らは鹿児島市全域の空き家分布を計算機により推定するための教師データの確保のために現地調査を行っている^{文8)}。実際に、多くの自治体で戸別訪問による外観目視マニュアル^{文5)} ^{文6)}を作成しているが、時間や人員に関して多大な消費を伴う実地調査においては対象地域が限定されること、目視確認ではその精度が調査員の主観に依存すること、外観の老朽化度合と居住の有無が必ずしも一致しないことなどの問題点がある。

数理的な手法を活用した研究では、重回帰分析を用いて、都道府県単位の空家数の推移を予測した研究^{文9)}やベイズ統計を用いた空き家数の推計^{文10)}などがあるが、どれも予測の単位が大きくまた推定の精度にも課題が残った。

そこで、近年は機械学習を用いて空き家の分布を推定する手法が提案されている。^{文11)}ここで秋山らは欠損値への対応と最適な閾値設定が可能な決定木モデルの機械学習モデルであるXGBoostを採用し、建物単位および250mメッシュでの精度の高い空き家・非空き家判定を行っている。XGBoostのような決定木モデルの勾配ブースティングによる機械学習は、一般化線形モデルよりも推定制度の向上を図るうえで有用であり、ニューラルネットワーク(NN)では加味できない欠損値を扱える点で優れているとしている。この手法で課題として挙げられているのは大規模な自治体に対してXGBoostでは計算コストが高くなる可能性がある点である。大規模な自治体では悉皆調査による全戸数の空き家・非空き家の教師データを得ることは現実的ではなく、別の教師データを検討する余地がある。

また、機械学習はその高い精度の一方で、予測結果の解釈が一般に困難である。このデメリットに対して、機械学習のモデル解釈法を用いる研究も見られるが、一面的な活用に留まっている^{文12)}。

データでは、公的な統計情報ではものの、民間企業等による空き家の情報もしばしば活用される。代表的なものとしては株式会社ゼンリンの住宅地図^{文13)}があり、近年は東京電力パワーグリッドとNTTデータが出資し、2018年11月に設立されたグリッドデータバンク・ラボ 有限責任事業組合(以下、GDBL)が電力データを活用した空き家調査の可能性を探っている^{文14)}。本研究でもこうした民間企業等による空き家調査データを活用しているが、同データは有償であり自治体によっては財政的な面からその導入や継続的な維持・管理が困難になることも考えられる。

以上の既往研究の概観から、空き家問題には様々なアプローチが試みられているものの、実地調査の実施の困難さ、数理的手法の推定精度の課題、そしてデータの活用に伴うさまざまな制約などの課題があることが明らかである。

1.3 本研究の目的

以上を踏まえ、本研究では、個々の建物や敷地に留まらず、都市全体に焦点を当てた空き家対策を目指す。増加傾向にある空き家問題に対する新たな統合的なアプローチを提案し、空き家対策策定における実用的な手法を開発することを目的とする。具体的には、以下の三つの主要目的を設定する。

①高精度な空き家分布の推定：大規模な自治体で実施が困難な悉皆調査に代わり、公的データを用いて空き家の分布を推計する機械学習手法を開発する。この手法により、大都市における空き家の正確な分布の把握を目指す。

②将来的な空き家の発生予測：人口動態データなどを活用し、従来の手法では困難であった将来の空き家発生を高精度に予測する。この予測により、予防的な対策の検討や計画が可能となり、空き家問題の長期的な解決に向けた戦略を支援する。

③機械学習の解釈手法を用いた実証的分析：近年進展している機械学習の解釈手法を活用し、空き家が発生する条件や要因を明らかにする実証的な分析を行う。

この分析を通じて、空き家の発生を防ぐための具体的な対策や政策提言を行政の現場での議論に提供し、より先進的な空き家政策の策定に繋げる。

2. 本研究の対象

2.1 研究対象地域

本研究の対象地域は、都市部から農村まで多様な地域特性を持ちつつ空き家問題が顕在化している神戸市に設定した。神戸市は、人口約150万人で、近年はその数が減少傾向にあり、空き家率は平成30年時点で約13.3%と、全国平均の約13.6%に近い水準である^{文15)}。この人口

減少は今後も継続し、それに伴い空き家の数も増加すると考えられるため、効果的な空き家対策が急務とされている。

さらに、本研究の実施にあたり先述した GDBL からの協力を得て神戸市内の各住戸における電力使用データの提供を受けた。また、神戸市から登記情報に基づく住宅に関する公的データの提供を受けている。

なお、本研究では研究対象の住戸については、その管理形態や住戸の利用状況が一戸建て住宅とは大きく異なることから、集合住宅は除外し一戸建て住宅に焦点を当てている

2.2 分析に用いるデータと指標の整理

既往研究^{文12)}で挙げられた空き家発生と関連のある指標をもとに、データを目的変数と説明変数に分類して整理した(表 2-1)。これらのデータは、GIS 等による空間解析を可能にするためメッシュデータ形式で加工する。

目的変数には、GDBL から提供された各住戸の電力使用データを使用した。この研究では、後述のような特定の条件を満たす一戸建てを空き家と定義し、悉皆調査を実施できない自治体でも利用可能なラベルデータとしての活用を模索する。本研究の提案手法で、将来的にシミュレーションを行う際に GDBL からのデータ購入の回数を減らすことができれば、自治体の財政負担を軽減する可能性がある。

説明変数には 2020 年度、それ以前のデータしかない場合は取得できる最新年度の公的データを活用した。この公的データには住民の個人情報の保護及び情報の利用のしやすさから公開情報のみを用いる。具体的には、世帯数など住民に関する情報や用途地域などの地域に関する情報は国土交通省数値ダウンロードサイト^{文16)}等のオープンデータから、また、物件の建設年度など物件に関する情報は登記簿の表題部記載の情報を用いる^{注1)}。

3. 適切な機械学習モデルの選定

3.1 モデリングの前提

本研究の目的を達成するため、決定木モデルの中でも特に勾配ブースティング技術 (Gradient Boosting Decision Tree; GBDT) を用いたモデリングを行う。勾配ブースティングの代表的な手法である XGBoost, LightGBM (Light Gradient Boosting Machine), および CatBoost の三つを試行し、最も適合性の高いモデルを選定する

LightGBM は推定精度が既往研究で用いられた XGBoost とほぼ同等でありながら、計算コストが低いという長所がある。このことから、大規模な自治体に対して有効な勾配ブースティング手法であると考えられる^{文17)}。

CatBoost も GBDT の一種で、特にカテゴリ変数を含むデータセットに対して優れた性能を発揮する。また、木構造の計算方法の特徴から過学習を軽減する能力が高い^{文18)}。

表 2-1 使用データ一覧

種類	カテゴリ	名称	出典
—	—	メッシュコード (500mメッシュ)	国土交通省：国土数値情報ダウンロードサイト
目的変数	—	推定空家数 (件)	GDBLより受領
説明変数	人口要因	2020年における推計人口 (人) 2020年における男女計65歳以上推計人口割合 (%)	国土交通省：国土数値情報ダウンロードサイト 国土交通省：国土数値情報ダウンロードサイト
	建物要因	平均一階床面積 (㎡) 平均延床面積 (㎡) 平均築年数 (年) 階数 (地上) 階数 (地下) 階数 (棟) 最多構造種別 (カテゴリ変数)	神戸市：建物登記情報 神戸市：建物登記情報 神戸市：建物登記情報 神戸市：建物登記情報 神戸市：建物登記情報 神戸市：建物登記情報 神戸市：建物登記情報
	周辺要因	公示地価 (円/㎡)	国土交通省：国土数値情報ダウンロードサイト
	地形要因	平均標高 (m) 平均傾斜度 (度)	国土交通省：国土数値情報ダウンロードサイト 国土交通省：国土数値情報ダウンロードサイト
	道路要因	幅員3.0m未満の道路密度 (㎡)	国土交通省：国土数値情報ダウンロードサイト
	土地利用要因	用途地域別面積割合 (%) 最多の土地利用項目 (カテゴリ変数) (土地利用細分メッシュデータ500mメッシュに再集計して活用) (田、その他の農用地、森林、荒地、建物用地、幹線交通用地、湖沼、河川等)	国土交通省：国土数値情報ダウンロードサイト
		土地利用項目の種類数 (件) (土地利用細分メッシュデータ500mメッシュに再集計して活用) (田、その他の農用地、森林、荒地、建物用地、幹線交通用地、湖沼、河川等)	国土交通省：国土数値情報ダウンロードサイト
		最多の土地利用項目の種類数 (件) (土地利用細分メッシュデータ500mメッシュに再集計して活用) (田、その他の農用地、森林、荒地、建物用地、幹線交通用地、湖沼、河川等)	国土交通省：国土数値情報ダウンロードサイト

分析環境としては、Python 3.8.16 をベースにした Anaconda の JupyterLab を使用する。これにより、コードの管理と分析の再現性が保証される。

モデルの比較には、Root Mean Squared Error (RMSE) と R-squared (決定係数) の二つの指標を用いた。RMSE は予測値と実際の値の差の平方根を取ったもので、予測の精度を測る指標として広く用いられている。値が小さいほど、モデルの予測精度が高いことを示す。決定係数はモデルがデータにどれだけ適合しているかを示す指標で、0 から 1 の範囲の値を取り 1 に近づくほどデータに良くフィットしていることを示す。しかし、実際の複雑な問題設定では、1 に近い値は珍しく 0.7~0.9 の範囲が良いモデルとされることが多い。

3.2 コードの実装

一般的な機械学習の分析手順に則り、データセットを説明変数（予測に使用される変数）と目的変数（予測される変数）に分割する。その後、データは学習セットとテストセットに分割し、モデルの訓練と評価のために用いる。このプロセスでは、一般的な train_test_split 関数を使用し、データセットの 20%をテストセットに割り当てる。次に、XGBoost, LightGBM, CatBoost の各モデルを、学習データセットを用いて訓練し、テストデータセット上での性能を評価した。性能評価の指標には、先述の通り RMSE および決定係数を使用した。

3.3 推定結果

表 3-1 に実行結果を示す。ここから、CatBoost モデルが最も優れた予測性能を示したことが分かる。その決定係数は約 0.749 であり、これは他のモデルよりも高いことから、データに最も良く適合していると評価できる。この結果は、CatBoost がカテゴリ変数を含むデータセットに対して特に有効であるというその特性が、本研究のデータセットに適していることを示唆している。

表 3-1 各種機械学習モデルによる分析結果比較

比較項目	XGBoost	LightGBM	CatBoost
RMSE	13.1	12.9	11.5
決定係数	0.672	0.684	0.749

4. 機械学習による空き家分布の将来推定

4.1 Catboost による予測モデル構築

本節では 3 章で最も良い結果となった Catboost を用いた予測モデルのより精密な構築手順について述べる。3 章のモデリングと同様に説明変数と目的変数を読み込んだ後、これを分割しさらにこれらを訓練データ、検証データ、試験データに分割する。検証データは必ずしも必要ではないため 3 章では省略したが、4 章では過学習

（モデルが学習データに過剰適合し、新しいデータに対して一般化できない状況）が発生してもこれを検出し、より正確な予測モデルを作成するために採用した。続いて、学習データへの過学習を抑制するためハイパーパラメータの調整を行う。本研究では GridSearch を用いた。GridSearch は全てのパラメータの組合せを試して、最適なハイパーパラメータの組み合わせを見つけるための手法である。計算コストは高いが、最適なパラメータを網羅的に探索できる。本研究では表 3-2 に示すパラメータから最適なパラメータの調整を行った。

表 3-2 ハイパーパラメータの調整範囲と最適なパラメータ

パラメータ名称	探索範囲	最適なパラメータ
iterations	50~1000	250
learning rate	0.1~1	0.1
depth	3~8	8
l2 leaf reg	2~9	6

この最適なパラメータを用いて再度 CatBoost モデルを学習した。図 3-1 に学習ログを示す。横軸に Iteration(モデルが学習データセットを通じて学習を繰り返す回数)、縦軸に RMSE をとりモデルの学習の進行状況を視覚的に確認する。ここで、訓練データと検証データの RMSE が共に低下しており、モデルが訓練データと検証データに対して正しく学習し、適切に一般化していることを示している。

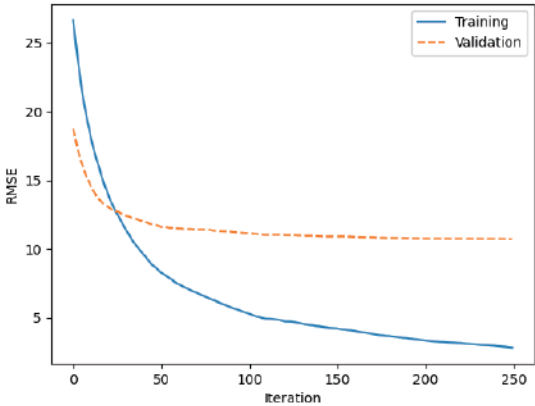


図 3-1 学習ログ

この時、RMSE 及び決定係数は表 3-3 の通りとなり、ハイパーパラメータの導入より、特に決定係数で示される予測モデルの精度が向上していることが分かる。

表 3-3 各種機械学習モデルによる分析結果比較

比較項目	ハイパーパラメータ 探索前	ハイパーパラメータ 探索後
RMSE	11.5	11.5
決定係数	0.749	0.792

4.2 Catboost による予測結果

図 4-1 に目的変数の分布、つまり 2020 年の空き家の分布を示し、図 4-2 に予測モデルより導かれた同年の空き家分布を示す。

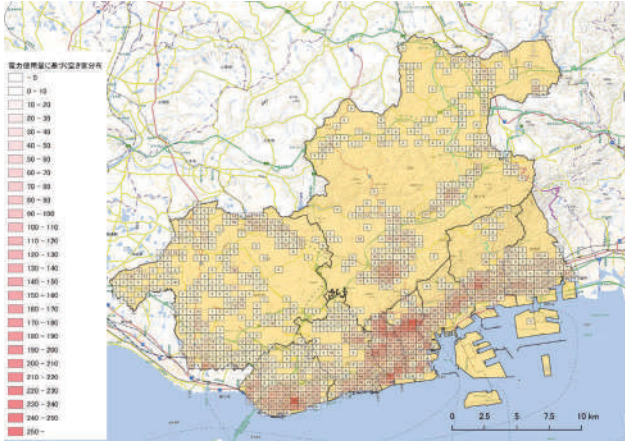


図 4-1 2020 年の空き家の分布（目的変数）

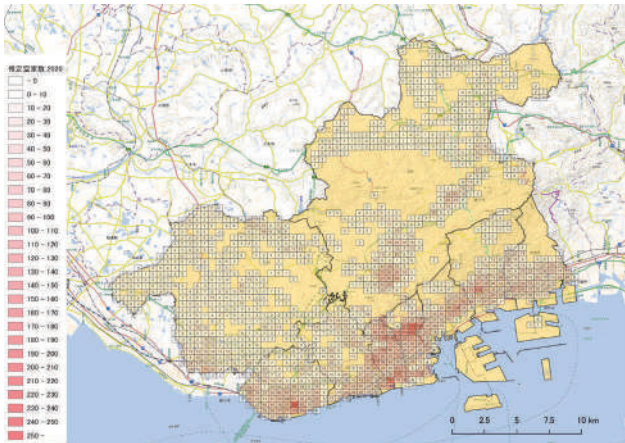


図 4-2 予測モデルによる空き家分布

ここから、視覚的にも目的変数と遜色ない予測値が得られたことが分かる。

また、図 4-2 で、空き家分布が表示されているメッシュが増えているが、これは目的変数が無いメッシュについても推定モデルから分布を推定できるためである。

さらに、この予測モデルを用いて、人口に関する変数を変更し、再度予測を行うことで、将来のシミュレーションを行う。変更する部分は「20XX 年の推定人口」および「20XX 年の男女計 65 歳以上推計人口割合」年である。

この人口のデータは年代別に表 4-1 に示す。これらを 2025 年から 5 年毎に 2050 年まで代入しシミュレーションを行う。データは国土交通省数値ダウンロードサイト文¹⁶⁾から入手した。シミュレーション結果を図 4-3 (a) ~ (f) に示す。

表 4-1 シミュレーション用に変更した変数

変更項目	モデル作成時	モデルを用いた推定時
推定人口	2020年	2025年、2030年、2035年 2040年、2045年、2050年
男女計65歳以上 推計人口割合	2020年	2025年、2030年、2035年 2040年、2045年、2050年

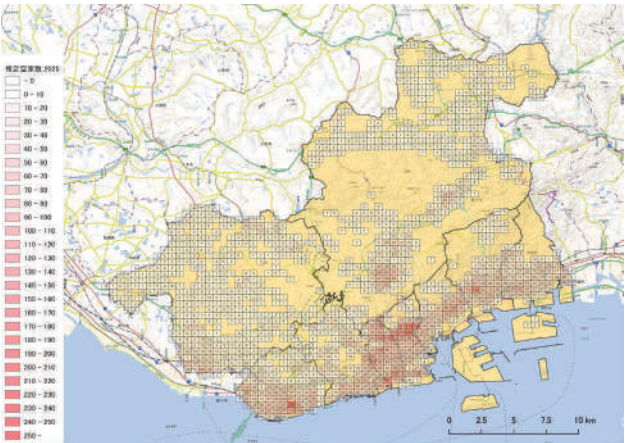


図 4-3 (a) 2025 年の空き家分布推定

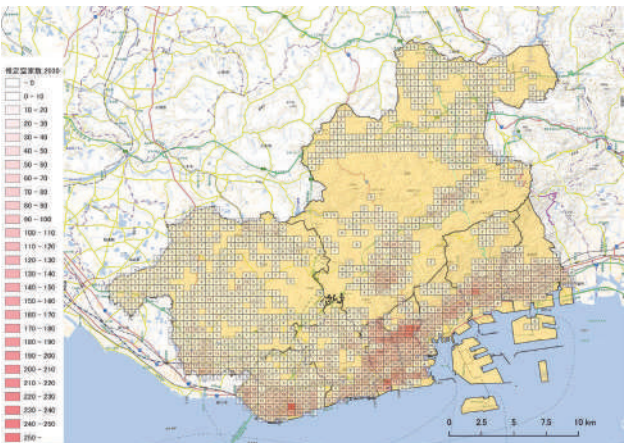


図 4-3 (b) 2030 年の空き家分布推定

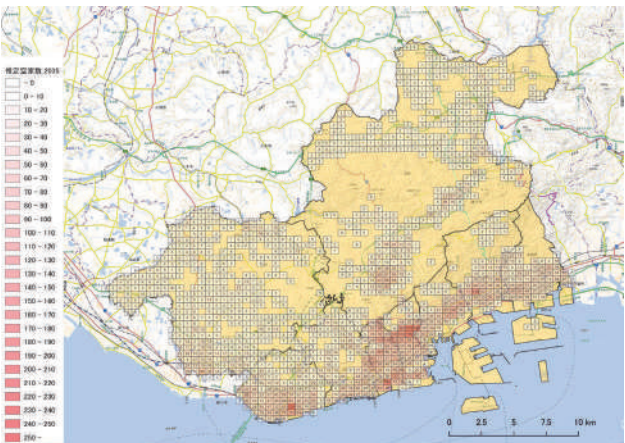


図 4-3 (c) 2035 年の空き家分布推定

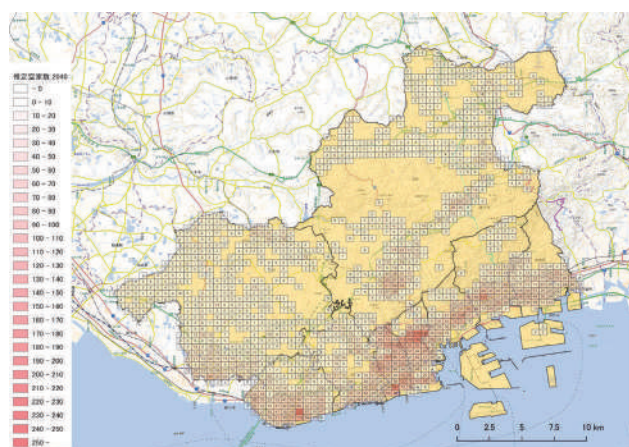


図 4-3 (d) 2040 年の空き家分布推定

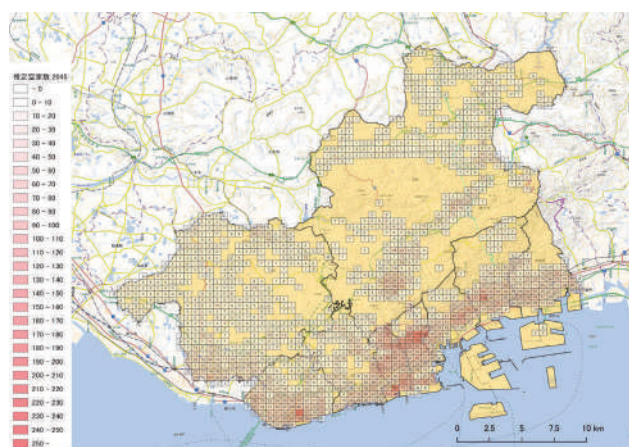


図 4-3 (e) 2045 年の空き家分布推定

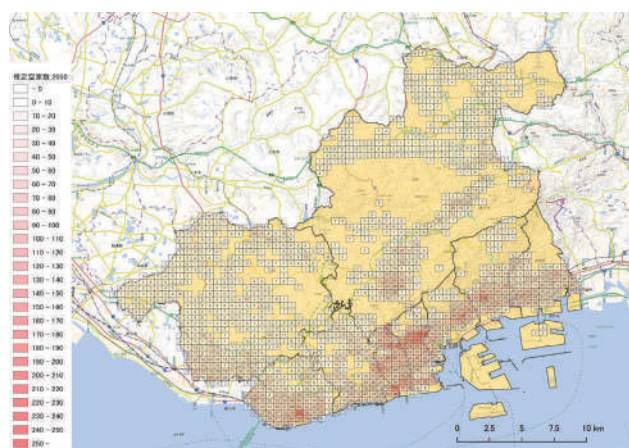


図 4-3 (f) 2050 年の空き家分布推定

図 4-3 2025 年から 2050 年までの空き家分布推定

また、図 4-3 のシミュレーション結果を基に、2020 年の予測値に対する 2050 年の変化率を図 4-4 に示す。

図 4-3 および図 4-4 から、西区周辺で空き家分布の増加率が高いことが分かる。ただし、空き家分布の絶対数は多くは無い。同様に、東灘区、灘区、兵庫区南部周辺で増加率が高く、こちらは絶対数も多い。これにはもとの推定人口が多いことも一因であると考えられる。

また、山間部では増加率が高い地域と低い地域が混在

している。こちらも同様に絶対数は多くない。

このように、将来的な空き家分布を明らかにし特に空き家が増加する可能性の高い精緻なエリアを高い精度で特定することができた。

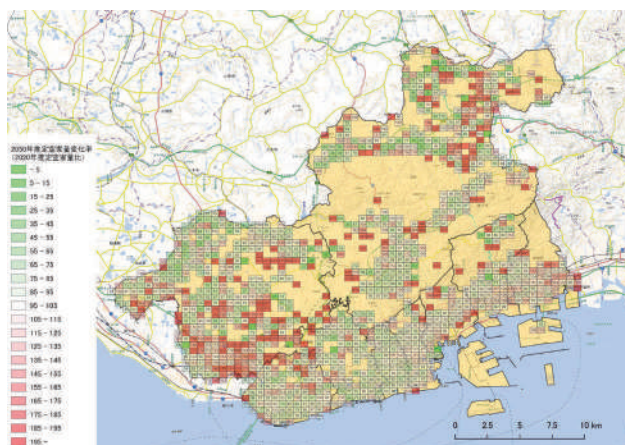


図 4-4 2020 年の予測値に対する 2050 年の変化率

5. 機械学習の解釈手法を用いた実証的分析

5.1 概要

機械学習の解釈手法を活用し、空き家が発生する条件や要因を明らかにする実証的な分析を行う。

機械学習の高度なモデルでは、膨大なデータから複雑なパターンを学習できるが、その決定過程が直観的に理解しにくくどのようにして特定の予測に至ったか説明することが難しくなる。この問題に対処するために本研究では機械学習の解釈手法を用いる。これは解釈可能な機械学習(XAI: Explainable AI)として研究が進んでいる分野であり、機械学習モデルがどのようにして予測や決定を行うかを人間が理解できる形で説明する技術である。これにより、機械学習による予測結果に対する信頼性の向上につながり、自治体での政策策定において明快な根拠に基づいた意思決定を支援することができる。

特に本研究では、機械学習の解釈手法の一つであるSHAP (SHapley Additive exPlanations)を用いて、作成したモデルを解釈し、空き家の分布傾向に対して各変数がどのように寄与しているのか判断する。SHAPは、数学の一分野であるゲーム理論に基づいたフレームワークであり、機械学習モデルの解釈性を高めるために近年活用が広がっている。一般的に解釈手法として用いられる特徴量重要度 (PFI: Permutation Feature Importance) は、ある特徴量の値をランダムに置換してモデルの性能にどのような影響を与えるかを評価する手法であり、シンプルで直感的な理解が可能である一方で、グローバルな重要度しか提供しないため、個々の予測に対する特徴量の寄与を知ることはできなかった。これに対しSHAPは、実装が難しく計算コストが高いものの、特徴量間の相互作用も考慮に入れるため、特徴量が単独でなく組み合わせ

によってどのように影響を及ぼすかを分析でき、特徴量の重要度が変わらない限り解釈の一貫性が保たれる点や、局所的な解釈とグローバルな解釈がどちらも可能である点で優れているため、本研究に適している。

5.2 SHAP によるモデル全体の評価

まず、Summary plot を用いて、特徴量ごとの SHAP 値をプロットし、モデルがどのように機能しているかを概観することで、モデルの予測に対する各特徴量の影響を視覚化する。図 5-1 にこの関数を用いて、上位 20 個の各特徴量に対する SHAP 値の分布を示す。

まず、特徴量は SHAP 値の絶対値の平均に基づいて縦軸に並べ、最も影響力のある特徴量から順に表示する(図 5-1)。これにより、モデルの予測に最も大きな影響を与える特徴量が把握できる。図 5-1 では、特徴量ごとの各データを点で表しており、各点の色はその特徴量の値が大きい小さいかを表し、具体的には、赤色は大きな値、青色は小さな値を表す。また、各点の位置はその特徴量が予測結果を増やすのか減らすのかを表す。具体的には左にあるほど負の影響、右に行くほど正の影響を表す。

ここから、平均築年数が高いエリアや人口が少ないエリアでは空き家と判定されやすい一方で、平均の床面積が小さいエリアや標高が高いエリアほど空き家と判定されにくいことなどが分かる。

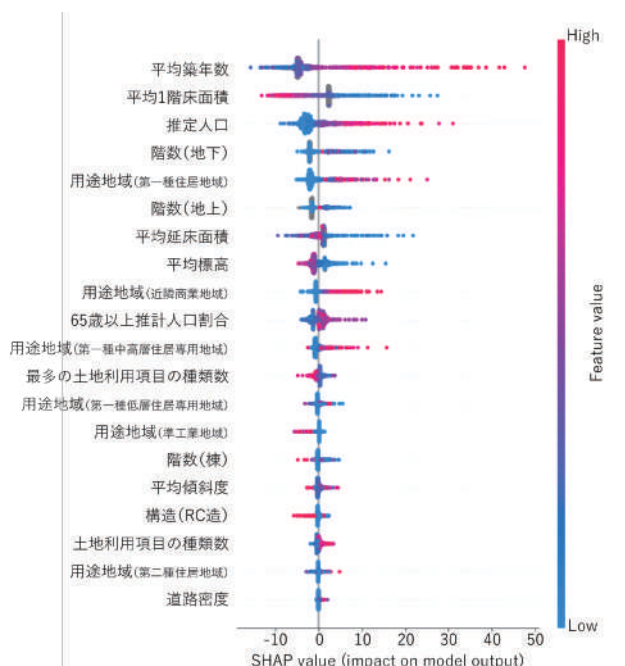


図 5-1 SHAP 値の概要プロット

5.3 SHAP による各変数の詳細な影響

続いて、Force plot を用いてモデルの個々の予測に対する特徴量の寄与を可視化する。これは、特定のデータポイントに対する予測がどのように形成されたかを推測するための可視化手法である。本研究では、計算量の観

点から、すべてのデータセットのうちランダムに 100 個を抽出しプロットを行う。各特徴量についてプロットした結果を上位 10 データについて図 5-2 に示す。

各プロットでは、縦軸の 14.92 が基準値であり、特徴量が一つもなかった場合のモデルの予測値である。ここから各特徴量が横軸の値を取った時の変化量がプロットされている。ここで上方は予測値を増加させる方向に影響を与えたことを示し、下方は予測値を減少させる方向に影響を与えたことを示す。

まず平均築年数では 37 年程度を境にそれより築年数が浅いデータでは予測値を減少させる方向に、築年数が高いデータでは予測値を増加させる方に影響を与えている。ここから、例えば築年数を基準とした空き家政策を検討する場合、閾値として築 37 年程度と設定することで効果的な対策を行えることが考えられる。

同様に、一階の平均床面積については、53 m²程度が境となり、それより床面積が小さいエリアでは予測値を増加させる方に、床面積が大きいエリアでは予測値を減少させる方に影響を与えている。ここから、例えば一階の床面積を基準とした空き家政策を検討する場合、閾値として 53 m²程度と設定することで効果的な対策を行えることが考えられる。

このように、各特徴量について図 5-2 中に示したように政策の判断材料となる値と、その前後の影響を明らかにできた。

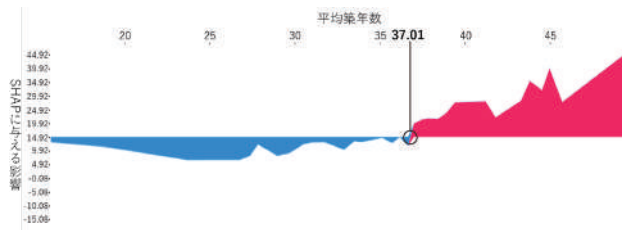


図 5-2 (a) 平均築年数の Force Plot

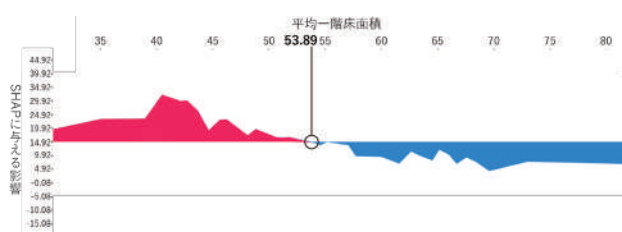


図 5-2 (b) 平均一階床面積の Force Plot

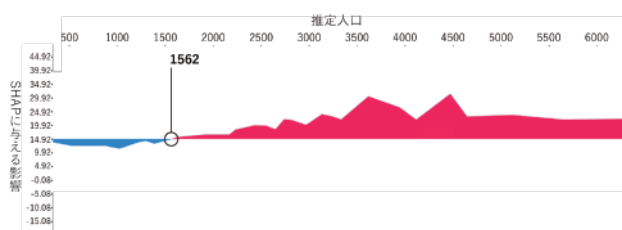


図 5-2 (c) 推定人口の Force Plot

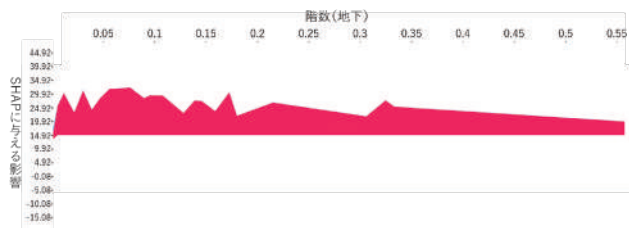


図 5-2 (d) 階数(地下)の Force Plot

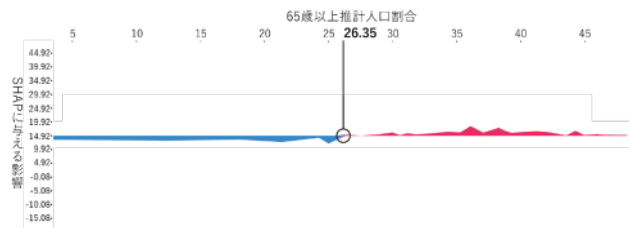


図 5-2 (i) 65 歳以上推計人口割合の Force Plot

図 5-2 各特徴量の Force Plot

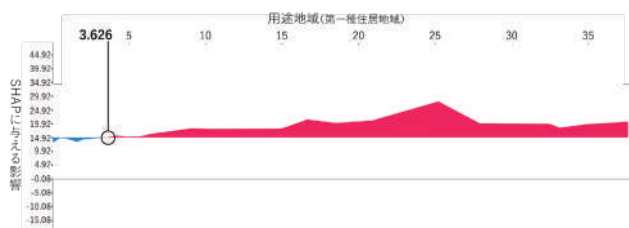


図 5-2 (e) 用途地域(第一種住居地域)の Force Plot



図 5-2 (f) 階数(地上)の Force Plot

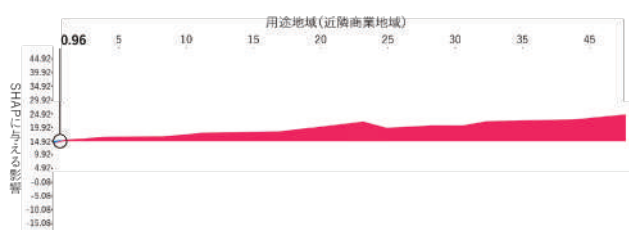


図 5-2 (g) 用途地域(近隣商業地域)の Force Plot

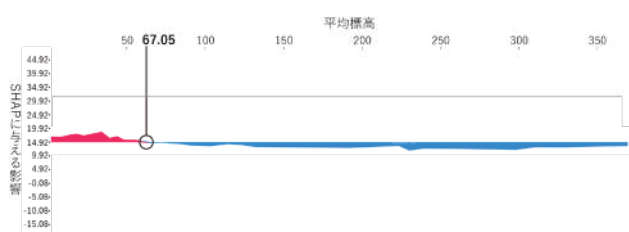


図 5-2 (g) 平均標高の Force Plot

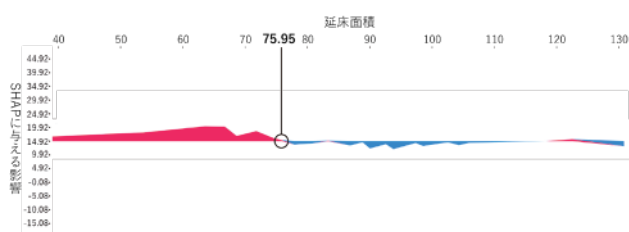
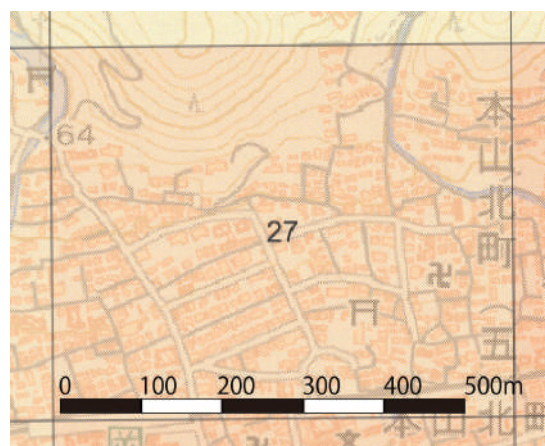


図 5-2 (h) 延床面積の Force Plot

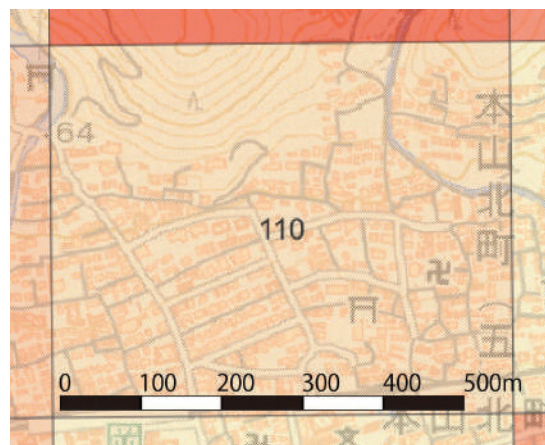
5.4 各メッシュにおける特徴量の影響

続いて、Waterfall plotを用いて、特定のデータポイントに対する予測値を、その予測に至るまでの各特徴量の寄与度を段階的に表示する。これにより、予測がどのように形成されたか、そして、各特徴量が予測にどのように影響を与えているかを明らかにする。

ここでは任意の地域を例にとり解説する。図5-3に東灘区のあるメッシュの予測値 (a) および2050年での増減率 (b) を示す。このエリアは空き家数が27戸、増減率は1, 10%で増加すると予測されている。



(a) 予測値



(b) 2050 年での増減率

図 5-3 東灘区のあるエリアにおける予測結果

続いてこのエリアにおける waterfall plot を図 5-4 に示す。縦軸に示している項目は上から順に影響が大きい特徴量である。10 番目以降の項目は影響力の少なさからひとまとめにしている。横軸は予測値を表す。 $E[f(x)]$ (14.915) は基準値を表し、ここから、各特徴量の SHAP 値を順に加えていき、最終的な予測値がどのように形成されるかを段階的に追跡する。ここでは最終的な予測値は $f(x)=26.617$ となった。プロットの各ステップは、特定の特徴量の影響を示し、特徴の寄与が正の場合は予測値を増加させ、負の場合は減少させる。これにより、モデルの予測が個々の特徴によってどのように形成されるか把握できる。

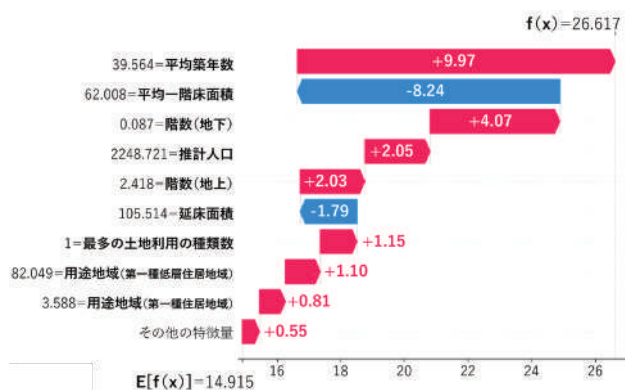


図 5-4 東灘区のあるエリアにおける Waterfall plot

図 5-4 から、このエリアでは平均築年数の高さが予測値を増加させる方に最も影響を与えていることが分かる。ここから、このエリアを対象に何らかの対策を取る場合、築年数を基準として施策を検討することが有効であると考えられる。一方で、一階の床面積の大きさが予測値を減少させる方向に影響を与えていることから、例えば敷地を分割して売却することを規制するなど、建物面積を減少させない施策も一定有効であると考えられる。

6. 結論

本論文では、増加する空き家問題に対処するための新たなアプローチとして、機械学習技術を活用した空き家分布の将来推定とその解釈に焦点を当てた。神戸市を研究対象地域とし、CatBoost モデルを用いて空き家分布の予測を行い、さらに SHAP という解釈手法を用いてモデルの予測結果の解釈を試みた。具体的に得られた成果は以下の通りである。

複数ある勾配ブースティングモデルの中でも CatBoost を用いた予測モデルが有用であることを示し、具体的な分析手法開発とその精度の向上に成功した。そして、人口動態や物件情報を含む様々な特徴量を活用することで、地域ごとの空き家発生の傾向を予測することに成功した。

続いて、ここで作成したモデルを用いて将来の空き家分布を高精度に推定することができた。ここから、具体的に空き家が増加するリスクの高いエリアを推定することができた。

そして、SHAP を用いた解釈手法により、モデルがどのようにして特定の予測結果に至ったかの説明が可能となり、神戸市においては特定の特徴量（例えば、平均築年数や一階の床面積）が空き家判定に大きな影響を及ぼしていることを明らかにすることができた。これにより、空き家対策を策定する際の明確な根拠を提供することが可能となった。

本研究で得られた知見は、自治体が空き家問題に対してより効果的な政策を策定するための基礎となる。特に、将来の空き家分布の予測とその解釈に基づく情報は、地域ごとに異なる空き家発生の傾向に対応した対策の立案を可能にする。

一方で、本研究の枠組みは、あくまで空き家分布の将来予測を通じた政策立案の支援であり、個別の物件に対する現場調査や詳細な管理方策の代替とはなり得ないという限界がある。また、利用したデータやモデルの精度には向上の余地があり、より多様なデータソースや分析手法を取り入れることで、より精緻な空き家発生メカニズムの解明や地域特性に根ざした施策立案が期待される。これらの限界を明確に認識することによって、本研究が自治体に対する政策支援として果たす意義がより明確になると考える。

今後の研究では、さらに多様なデータの収集と分析を行い、空き家対策の精度を高めていくことが望まれる。空き家の将来分布予測をさらに精緻化するためには、自治体や民間セクターとの連携を強化し、新しい情報源の活用や分析手法の改良に取り組むことが重要である。これにより、空き家対策における効果的な政策立案や地域活性化に一層寄与できる可能性があると期待される。

<謝辞>

本研究では神戸市建築住宅局政策課空家空地活用担当にデータの取得にご尽力いただき、分析結果の解釈についても多くの助言をいただいた。この場を借りて御礼申し上げる。

<注>

- 1) 神戸市は都市計画基礎調査を公開していないため、既往研究^{文 12)} で用いられているような建物重心点からの広幅員道路までの距離や部根木施設までの距離、周辺の土地利用といった建物ごとの詳細な情報を入手することができない。特に、便益施設までの距離などの周辺施設要因が不足しているが、ヘドニック法のような考え方を採用し、公示地価で代表させている。
- 2) 図 5-3 に示すようなメッシュでは、基準地域メッシュの形は正方形といわれている一方で、実際にはやや横長の長方形となっている。これは、地図上での経緯度の投影方法により緯度が高くなるにつれてメッシュの東西方

向(経度方向)の間隔が広がるためである。赤道付近ではほぼ正方形に近いメッシュが形成されるが、緯度が高くなるにつれて東西方向(経度方向)の距離が拡大する。日本は緯度が比較的高い位置にあるため、メッシュは横長の長方形になる。

<参考文献>

- 1) 国土交通省：空き家対策について, 2018, <<https://www.mlit.go.jp/common/001290020.pdf>> (参照：2022.07.01)
- 2) 国土交通省：空家等対策の推進に関する特別措置法(平成26年法律第127号)
- 3) 国土交通省：新たな住生活基本計画の概要(令和3年3月19日閣議決定)
- 4) 織田峻央, 森本章倫, 浅野周平：地域特性及び物件属性に着目した空き家の発生状況に関する研究, 都市計画論文集, Vol. 53, No3, pp. 1074-1079, 2018.10,
- 5) 横須賀市：外観目視による空き家調査票, <<https://www.city.yokosuka.kanagawa.jp/4805/tokei/chosei/cdata/documents/s38-4.pdf>> (参照：2022.07.01)
- 6) “熊本市：外観目視による空家等不良度判定マニュアル”, <https://www.city.kumamoto.jp/common/UploadFileDisp.aspx?c_id=5&id=24944&sub_id=10&flid=233620> (参照：2022.07.01)
- 7) 益田理広, 秋山祐樹：昨今の空き家研究の盛衰と手法について-日本における研究状況に着目して-, 地理空間, 第13巻, 第1号, pp. 1-26, 2020.6
- 8) 秋山祐樹, 上田章紘, 大野佳哉, 高岡英生, 木野裕一郎, 久富 宏大：鹿児島県鹿児島市における公共データを活用した空き家の分布把握 自治体の公共データを活用した空き家の分布把握手法に関する研究(その1), 日本建築学会計画系論文集, 第83巻, 第744号, pp. 275-283, 2018.2
- 9) 金森有子, 有賀敏典, 松橋 啓介：空き家率の要因分析と将来推計, 公益社団法人日本都市計画学会 都市計画論文集 Vol. 50 No. 3 2015.10
- 10) 齋藤茂樹, 高倉政寛, 堤拓哉：ベイズ統計を用いた空き家数の推計—ベイズ統計を用いた空き家数の推計—, 第90巻, pp. 261-264, 2017.6
- 11) 秋山祐樹, 馬場弘樹, 大野佳哉, 高岡英生：機械学習による空き家分布把握手法の更なる高度化 自治体の公共データを活用した空き家の分布把握手法に関する研究(その3), 日本建築学会計画系論文集, 第86巻, 第786号, pp. 2136-2021, 2021.8
- 12) 水澤 克哉, 宮本 慧, 田村 将太, 田中 貴宏：機械学習の解釈手法を用いた空き家の発生要因分析 適材適所の空き家対策支援ツールの開発を目指して, 公益社団法人日本都市計画学会 都市計画論文集 Vol. 57 No. 3 2022.10
- 13) 株式会社ゼンリン：ゼンリン住宅地図, <<https://www.zenrin.co.jp/product/category/residentialmap/index.html>> (参照：2022.07.03)
- 14) グリッドデータバンク・ラボ有限責任事業組合空き家対策への電力データ活用可能性の検証, 2019, <<https://www.gdb-lab.jp/wp-content/uploads/2019/10/%E5%88%A5%E7%B4%992.%E5%A0%B1%E5%91%8A%E6%9B%B8%E3%82%B5%E3%83%9E%E3%83%AA%E3%80%90%E4%BD%8F%E5%AE%85%E8%AA%B2%E3%80%91.pdf>> ((参照：2022.07.03))
- 15) 神戸市：神戸市の空き家等の状況とこれまでの取り組み, <<https://www.city.kobe.lg.jp/documents/25499/tai>
- sakukeikaku_2021_besshi2.pdf> (参照：2024.03.19)
- 16) 国土交通省：国土数値情報ダウンロードサイト, <<https://nlftp.mlit.go.jp/ksj/>> (参照：2024.03.19)
- 17) Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. : LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 3149-3157). NIPS'17.
- 18) Dorogush, A. V., Ershov, V., & Gulin, A. (n.d.) : CatBoost: gradient boosting with categorical features support. arXiv:1810.11363. <https://doi.org/10.48550/arXiv.1810.11363>