

# シビックプライド醸成のための 新住民価値の発掘モデルの検証と 住民価値向上のための自治体施策の 具体化に関する研究

2023.9.7

東京国際工科専門職大学 情報工学科

山本 裕 河野 敏也 高田 晃希 黒羽 晟

株式会社 百代

橋本 沙也加 橋本 尚子 岡田 ゆかり



東京国際工科専門職大学

International Professional University of Technology in Tokyo



# 目次

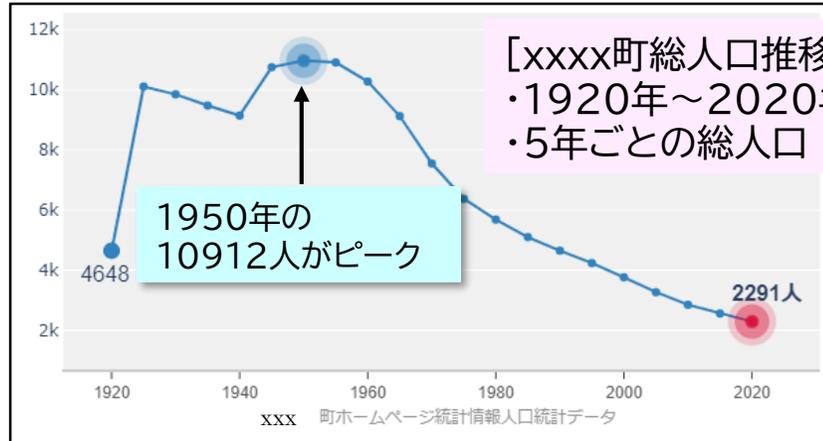
1. 研究の背景と概要
2. 本研究の主なプロセス
3. 新住民価値導出のための特徴量の選択
  - (1)数量化Ⅱ類
  - (2)自由記述回答からの特徴量算出(特徴語抽出)
  - (3)機械学習モデルの適用と特徴量の評価
4. 主成分分析による住民価値の掘り起こし
5. 今後の課題と対応方針

【参考】2022年度アンケート分析について

# 1. 研究の背景と概要 「潜在的住民価値発掘モデル」

## 社会課題

少子高齢化に伴う**地方の過疎化・高齢化** → 地方自治体の課題と対策が急務発掘モデル」を汎用化したい



- ・人口が1950年から2020年まで単調減少。
- ・基幹産業の人口も年々減少。
- ・1990年には老年人口が年少人口を、更に2015年には生産年齢人口を逆転。

## シビックプライドの醸成と研究目的

- ◆ **「市民の当事者意識に基づく都市に対する誇り」**という住民意識を高める**「シビックプライド」**に着目
- ◆ シビックプライド醸成により、地方住民が自治行政を推進し、持続可能なまちづくりが実現できる。  
本研究は、**シビックプライドの醸成のための施策を地方公共団体へ提案**することを目的とする。

## 地方課題の解決を1事例とした**「潜在的住民価値発掘モデル」**

- ◆ ある自治体の事例を研究対象に、シビックプライド醸成施策導出のための**「潜在的な住民価値発掘モデル」**を構築
- ◆ 多くの自治体で訴求を図る**「潜在的な住民価値発掘モデル」**を汎用化したい。

# 1. 研究の概要と背景 課題と研究概要

(1) **新住民価値の明確化**: 「xxxx町第4次総合計画アンケート」結果を分析。

AIを用いてシビックプライド醸成に寄与する**潜在的「新住民価値」の発掘**を行うモデルを創生

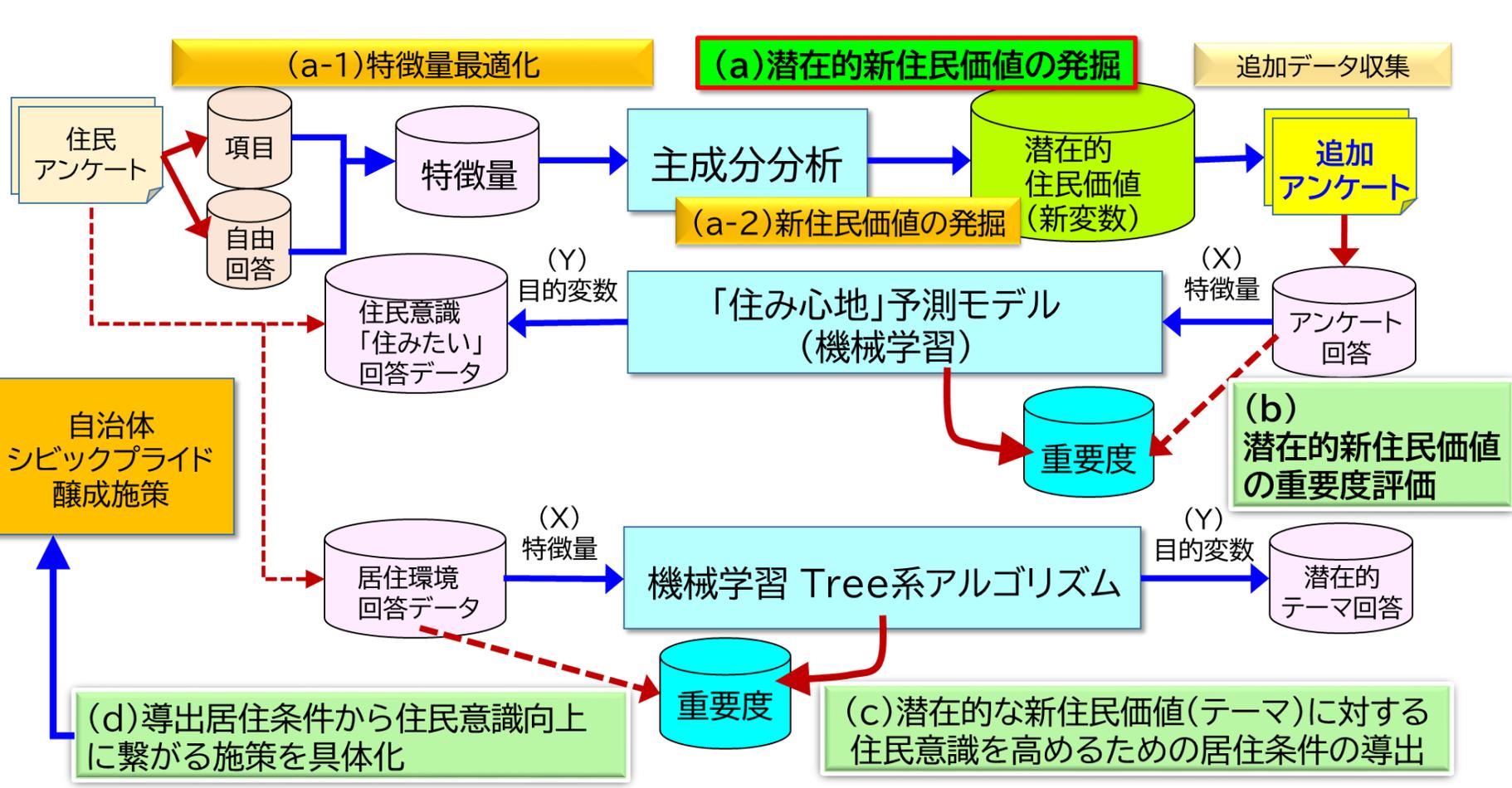
(2) **新住民価値創出のための施策提案**:

- ・発掘した「新住民価値」を向上させるために(アンケートの項目に紐づく)どういう環境要件が重要かを明確化
- ・明確にした環境要件から、具体的な施策案を検討・提案

目的	研究プロセス		概要 <span style="float: right;">今回発表の範囲</span>
(1)新住民価値の明確化	(a) 潜在的な新住民価値の発掘	(a-1)特徴量最適化	アンケート結果から <b>潜在的な「新住民価値」を発掘する主成分分析の精度向上のための特徴量選択</b> (数量化Ⅱ類、相関分析、テキストマイニング、機械学習)
		(a-2)新住民価値の発掘	<b>導出主成分の意味付け(主成分分析)</b> とアンケート追加
	(b)潜在的な新住民価値の重要度評価		追加アンケート回答の分析 追加アンケート回答を(説明変数(X))とした住み心地回答(目的変数(y))に対する(X)の重要度評価
(2)新住民価値創出のための施策提案	(c)「潜在的な新住民価値」に対する住民意識を高めるための居住条件の導出		追加アンケート回答の住民意識を高める居住条件導出 追加アンケート回答を(y)とし他アンケート回答を(X)とした場合の(X)の重要度評価
	(d)導出居住条件から住民意識向上に繋がる施策を具体化		上記(c)の居住条件を満たすような自治体施策を検討提案

# 2. 本研究の主なプロセス

◆前出(a)(b)(c)の課題を対策できるモデルを確立・評価することが本研究目的  
◆今回の発表対象は(a)の範囲(2022年度完了)。(b)(c)は推進中(2023年度完了予定)



(a) 潜在的な新住民価値発掘  
xxxx町住民アンケート  
(2018~22年度)から  
潜在的な新住民価値を発掘

(b) 潜在的な新住民価値の重要度評価  
・導出した「新住民価値」を2022年度のアンケート項目に追加。  
・「住み心地」回答に対する追加アンケート項目回答の重要度を評価

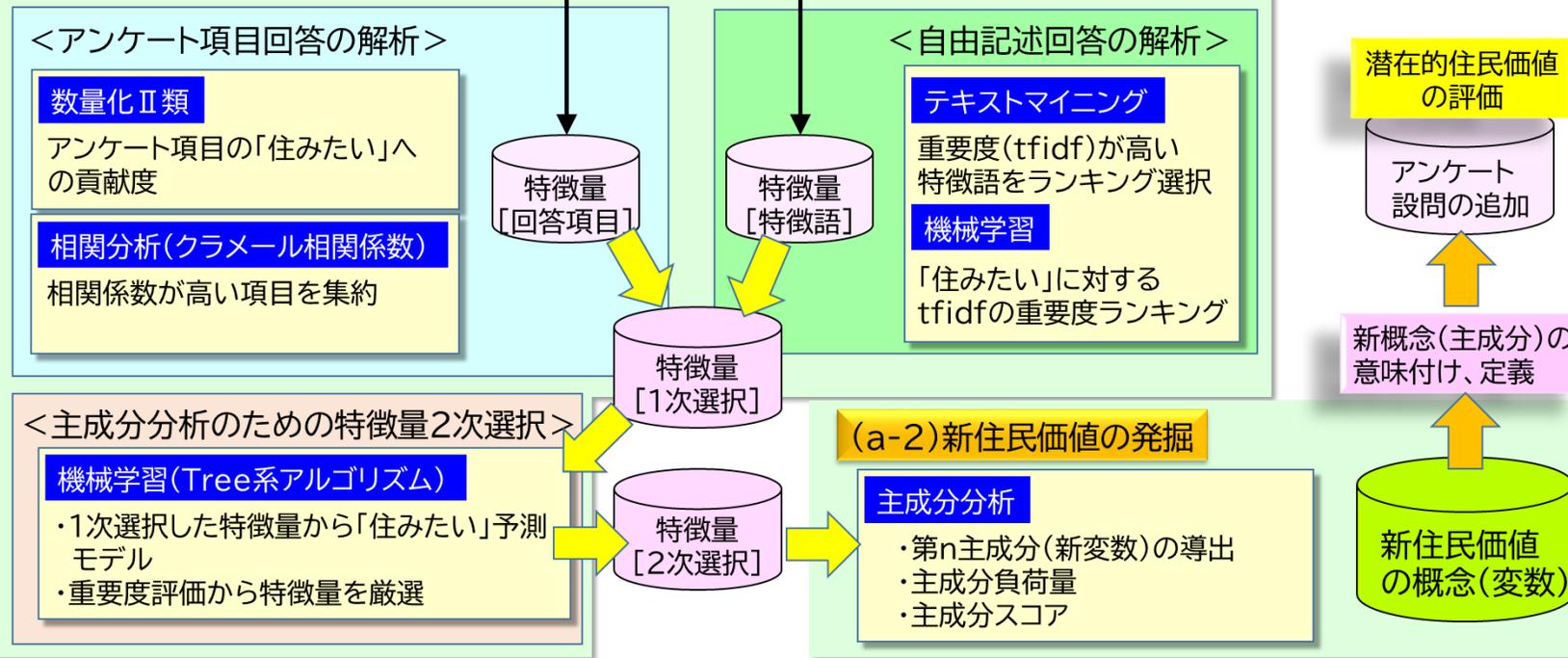
(c)(d) 「潜在的な新住民価値」の住民意識を高める居住条件の導出と施策の具体化  
追加アンケート回答に対する他アンケート項目回答の重要度を評価。重要度が高い居住環境に関して施策を具体化

# 2. 本研究の主なプロセス

## (a)「潜在的新住民価値発掘」(今回発表対象)のプロセス

### (a) 潜在的新住民価値の発掘

#### (a-1) 特徴量の最適化: 「住みたい」回答に重要度が高い特徴量の選択



### (a-1) 特徴量の最適化:

- ・アンケート項目回答からシビックプライド醸成に寄与する回答項目を特徴量選択
- ・選択した特徴量を主成分分析に入力する特徴量とする

### (a-2) 新住民価値の発掘:

- ・選択した特徴量から主成分分析で新住民価値に該当する主成分(新変数)を発掘
- ・新変数の意味付けを行い、対応するアンケート項目を追加し回答を評価する(新変数項目に対する住民の評価)

# 3. 新住民価値導出のための特徴量の選択

## (a-1) 特徴量の最適化プロセス

分析対象	手法	手法の概要
アンケート 項目回答 (特徴量)	数量化Ⅱ類	<ul style="list-style-type: none"><li>・「住み心地」の群の相関比を最大にするアンケート項目(特徴量)の重み(「住み心地」回答に対する寄与率)を評価</li><li>・寄与率ランキングでアンケート回答項目を選択</li></ul>
	相関分析	数量化Ⅱ類で選択した特徴量に対してクロス集計を行い、変数間のクラメール連関係数を評価(多重共線性の排除)
アンケート 自由記述回答 (特徴量)	テキスト マイニング	<ul style="list-style-type: none"><li>・特徴語抽出: KH Coderで抽出した各形態素のTF-IDFを評価</li><li>・特徴語の数量化: 特徴語ごとのTF-IDF値ランキング50を選択(特徴語選択)し特徴語を数量化(ダミー変数化またはTF-IDF)</li></ul>
アンケート 項目回答 + アンケート 自由記述回答	機械学習	<ul style="list-style-type: none"><li>・数量化Ⅱ類およびテキストマイニングで選択した特徴量/特徴語をマージ入力。</li><li>・主成分分析へ入力する特徴量を選択する(特徴量の重要度評価)。</li><li>・精度が高いアルゴリズムを評価選択(PyCaret)。</li><li>・「住み心地」回答変数を目的変数、上記を説明変数とし機械学習モデルで学習/予測/精度評価。特徴量の重要度を評価し50個の特徴量を選択</li></ul>

# 3. 新住民価値導出のための特徴量の選択

## (1) 数量化Ⅱ類

アンケート回答項目を説明変数、アンケート回答項目の選択項目をカテゴリと定義し、モデル式のカテゴリ係数を算出する

### 数量化Ⅱ類の適用例

「住み心地」を判別するモデル式: 「住み心地」のスコア $y$ に影響あるアンケート項目を見出す

$$y = (0.231x_{11} + 0.913x_{12} - 0.872x_{13}) + (1.013x_{21} + 0.447x_{22} - 0.012x_{23} - 0.526x_{24}) + (0.202x_{31} - 0.149x_{32})$$

#### 住民アンケート

1. 住み心地はどうか?  
①良い ②悪い

目的変数

2. 年齢  
①10~30代 ②40~60代  
③70~90代

3. 交通の満足度  
①満足 ②やや満足 ③やや不満  
④不満

4. 一人暮らしか?  
①はい ②いいえ

説明変数	年齢			交通の満足度				一人暮らしか?		
カテゴリ係数	0.231	0.913	-0.872	1.013	0.447	0.012	-0.526	0.202	-0.149	
群	10~30	40~60	70~90	満足	やや満足	やや不満	不満	はい	いいえ	
変数→	$x_{11}$	$x_{12}$	$x_{13}$	$x_{21}$	$x_{22}$	$x_{23}$	$x_{24}$	$x_{31}$	$x_{32}$	
住み心地 良い	1	0	1	0	1	0	0	0	1	0
	1	1	0	0	0	1	0	0	0	1
	1	0	1	0	1	0	0	0	0	1
	1	0	1	0	0	1	0	0	0	1
住み心地 悪い	2	1	0	0	0	1	0	1	1	0
	2	0	0	1	0	0	1	0	0	1
	2	0	0	1	0	0	0	1	1	0

# 3. 新住民価値導出のための特徴量の選択

## (1) 数量化Ⅱ類

### 数量化Ⅱ類結果から特徴量を選択

- ◆特徴量を選択するため、住み心地(目的変数)に対して重要な説明変数を選択する。
- ◆説明変数内において、カテゴリ係数の最大値と最小値の差で与えられるレンジを重要度を示す尺度として利用、レンジ0.15以下の12個の説明変数を削除
- ◆対象は2018年度アンケート全回答(1738名)の全回答項目

### 結果評価

- ◆職業に関して「農業」のカテゴリ係数の絶対値が最も大きい。
- ◆レンジランキング2位の居住地区は地区によって「全員が住み心地が良い」「10%が住み心地が悪い」と意識が異なる。
- ◆3位は訪町者からの協力金で、アンケート回答の自由記述を眺めたところ、世界文化遺産に認定された環境を町としてどう利用するかが、住民の関心事の一部であることが認識できた。

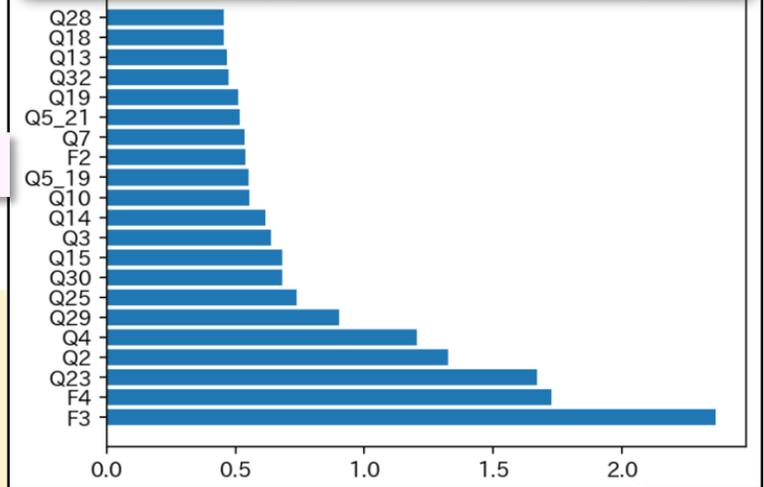
相関比(「住み心地が良い」「悪い」の2群が離れている度合)が、最大になるように各特徴量(アンケート項目)の重み(モデル式のカテゴリ係数)を算出。  
→「住み心地が良い」「悪い」を判別するための寄与率(カテゴリ係数)  
以下はモデル式からレンジを算出する例

$$y = (0.231x_{11} + 0.913x_{12} - 0.872x_{13}) + (1.013x_{21} + 0.447x_{22} - 0.012x_{23} - 0.526x_{24}) + (0.202x_{31} - 0.149x_{32})$$

目的変数:  $y$   
説明変数: 年齢 ( $x_{11}, x_{12}, x_{13}$ )  
説明変数: 交通満足度 ( $x_{21}, x_{22}, x_{23}, x_{24}$ )  
説明変数: 一人暮らしか ( $x_{31}, x_{32}$ )

最大値 = 0.913  
最小値 = -0.873  
レンジ = 0.913 - (-0.873) = 1.786  
= 目的変数に対する年齢の重要度

カテゴリ係数レンジグラフ 1~21位



ランキング詳細は29ページ

# 3. 新住民価値導出のための特徴量の選択

## (2) 自由記述回答からの特徴語抽出

### テキストマイニングの目的

- ◆アンケート項目回答だけでなく、自由記述回答も含めた特徴量から新住民価値を掘り起こす。
- ◆自由記述回答から特徴量として特徴的な語(特徴語)を抽出するためにテキストマイニングを行う。

### テキストマイニングの手法

- ◆自由記述回答から特徴語をマイニングするため、文書群において単語がどのくらい特徴的かを表す指標である **TF-IDF** を利用する。
- ◆文書群の中で特徴的な語ほど、**TF-IDF** の値が高い。

#### TF-IDF

TF : Term Frequency 単語頻度

IDF : Inverse Document Frequency 逆文書頻度

$$TF = \frac{\text{単語}t\text{の出現回数}}{\text{文書内の総単語数}} \quad IDF = \log \frac{\text{総文書数}}{\text{単語}t\text{を含む文書の数}}$$

$$\mathbf{TF-IDF} = TF \times IDF = \text{単語使用頻度} \times \text{単語レア度}$$

[TF-IDFについて](#)

# 3. 新住民価値導出のための特徴量の選択

## (2) 自由記述回答からの特徴語抽出

※KH Coder: テキストマイニングのためのフリーソフト

テキストマイニングの具体的手法(例)

導出形態素の単語出現頻度

TF-IDF算出

回答者	文書群
1	文書1
2	文書2
3	文書3
4	文書4

※KH Coder を使って  
形態素解析

回答者	単語数	単語1	単語2	単語3
1	4	3	1	0
2	21	10	2	9
3	13	0	1	12
4	20	4	10	6

単語1TF-IDF	単語2TF-IDF	単語3TF-IDF
0.21576	0	0
0.13699	0	0.12329
0	0	0.26555
0.05753	0	0.08631

### [TF-IDFの算出例]

・回答者2が使用した単語1のTF-IDFを求める。

総文書数 = 4

回答者2の単語1の使用数 = 10

回答者2の使用単語数 = 21

単語1を含む文書数 = 3

$$TF-IDF = TF \times IDF$$

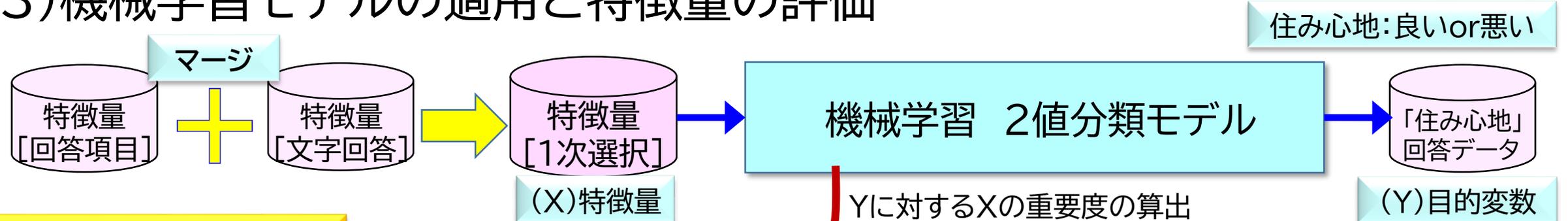
$$= \frac{\text{回答者}_2 \text{単語}_1 \text{の使用数}}{\text{回答者}_2 \text{の使用単語数}} \times \log\left(\frac{\text{総文書数}}{\text{単語}_1 \text{を含む文書数}}\right)$$

$$= 10/21 \times \log(4/3) = 0.13699$$

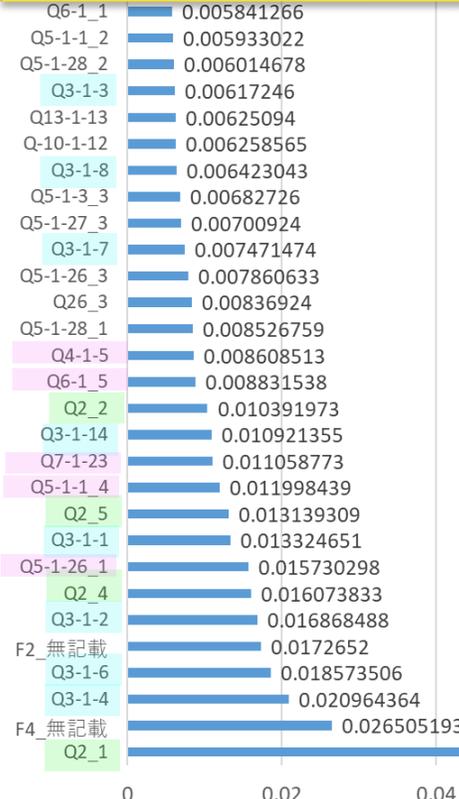


# 3. 新住民価値導出のための特徴量の選択

## (3) 機械学習モデルの適用と特徴量の評価



### 特徴量重要度ランキング



記号	アンケート内容	選択回答
Q2_1	住み続けたさ	#1住み続けたい
F4	居住地区	無記載
Q3-1-4	住み続けたい理由	#4土地と家あり
Q3-1-6	住み続けたい理由	#6自然環境良い
F2	年代	無記載
Q3-1-2	住み続けたい理由	#2ずっと住んでいる
Q2_4	住み続けたさ	#4どちらとも言えない
Q5-1-26_1	日常買い物利便性	#1満足していない
Q3-1-1	住み続けたい理由	#1町に愛着あり
Q2_5	住み続けたさ	#5未回答
Q5-1-1_4	本土との海上交通	#4未回答
Q7-1-1_23	町の課題	#23未回答
Q3-1-14	住み続けたい理由	#14近所付き合いに満足
Q2_2	住み続けたさ	#2将来他に移りたい
Q6-1_5	定住人口問題	#5未回答
Q4-1_5	将来他に移りたい	#5買い物通院に不便
Q5-1-28_1	働く場の確保	#1満足していない
Q26-3	1人暮らしか	#3未回答

### 機械学習モデルによる特徴量評価の目的

◆ 選択項目回答と自由記述回答をマージしたデータから **統合的に特徴量を評価**

### 手法・評価

◆ 住み心地を分類する上で重要度が高い特徴量を算出 **降順にランキング**

◆ ランキングから **重要度0.004以上の50個の特徴量を主成分分析に入力する特徴量として選択した。**

◆ 生活基盤(交通、経済的基盤)の充実が住み心地に大きく **影響**

◆ 居住地区・年代も住み心地に影響あり。居住地区・年代を軸にした掘り下げが課題。

MLモデル実行と特徴量の重要度

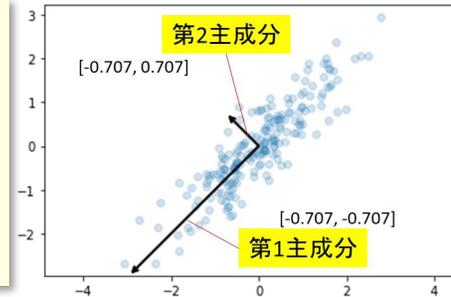
# 4. 主成分分析による住民価値の掘り起こし

## 導出主成分を潜在的な新住民価値とする

選択した前段の**特徴量データの情報を保持したまま、特徴量データを要約できるような新しい軸**を導き出す  
→潜在的な新住民価値

### <主成分分析>

- (1)特徴量データの情報を最大限保持するように軸(ベクトル)を選ぶ(主成分)
- (2)特徴量のデータを主成分に射影したときの**分散が最大**となる軸(主成分)の式を決める
- (3)軸(主成分)の意味付けをする



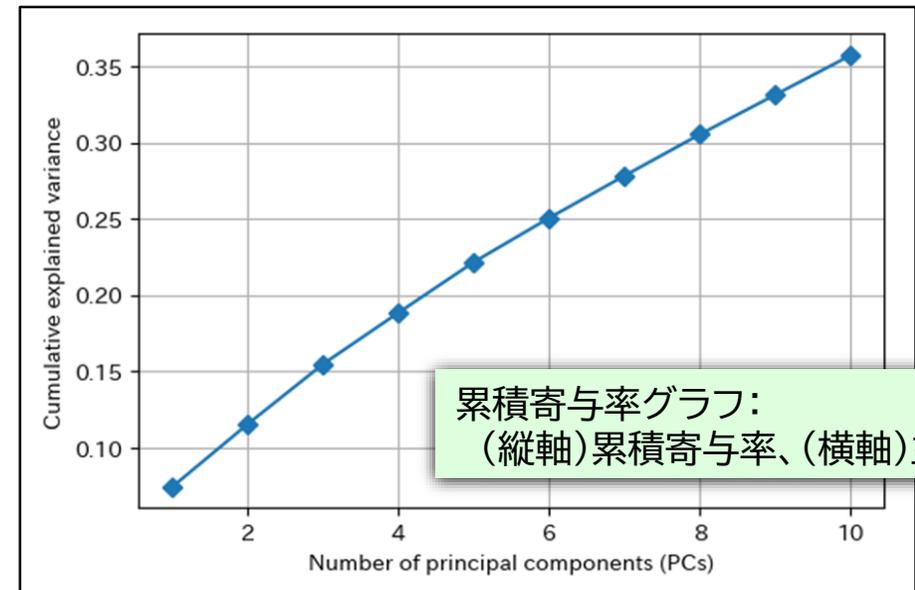
【入力する特徴量】前プロセスでの機械学習にて、Q1に対する**影響度が0.004以上の特徴量(50個)**を選択。  
【主成分の導出】

- F2 から Q26 までのフィールドの値を要素とする 50×1738 の行列を標準化、第10主成分まで算出
- pc1(第1主成分)およびpc2(第2主成分)**を使って**新変量を意味付け**。
- 第1～第4主成分での累積寄与率は20%程度。

主成分	寄与率
pc1	0.074565
pc2	0.041046
pc3	0.039408
pc4	0.033555
pc5	0.033219

寄与率の定義式  
( $\lambda$ は第*i*主成分に対応する分散)

$$\frac{\lambda_k}{\sum_{i=1}^d \lambda_i} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_d}$$



累積寄与率グラフ:  
(縦軸)累積寄与率、(横軸)主成分数



# 4. 主成分分析 主成分内容と意味付け(pc1)

主成分の内容(主成分負荷量の傾向)

pc1の意味付けについて

どのアンケート項目の「主成分負荷量」が、pc1の値を強めるか(貢献度)によりpc1を意味付け

	アンケート項目(説明変数)からの意味付け
合成変量の意味	未来改善(不)志向度
主成分負荷量(正)	<ul style="list-style-type: none"> <li>・現状の生活環境不満</li> <li>・まちづくり参加意欲小</li> <li>・やや消極的</li> <li>・人口・就業・文化遺産活用に一部課題提起</li> <li>・地域特性 (Fエリア、NKエリア、Mエリア)</li> <li>・年代特性 (40~50代、10代~30代)</li> </ul>
主成分負荷量(負)	<ul style="list-style-type: none"> <li>・現状の生活環境満足</li> <li>・まちづくり参加意欲大</li> <li>・積極的</li> <li>・生活インフラ/環境、コミュニティ改善意向</li> <li>・未来の環境に対しての見解意識</li> <li>(コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題)</li> <li>・地域特性 (Hエリア、Yエリア、Zエリア、Kエリア、Oエリア、Nエリア)</li> <li>・年代特性 (60代~90代)</li> </ul>

## ①若手/ベテランの二極化

主成分負荷量の傾向として  
年代特性が現れる(正: 10~50代、負: 60~90代)

## ②若手(不満/消極的)、ベテラン(満足/積極的)

主成分負荷量の傾向として  
年代特性 × 参加意欲(正: 消極的、負: 積極的)

## ③若手はベテランに遠慮している傾向

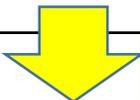
主成分負荷量の傾向として、②の傾向に加えて、  
生活不満を表す負荷量の絶対値大 ×  
参加意欲が低い回答の負荷量大

## ④若手は、古い風習や文化、環境に不満

主成分負荷量の傾向として、②の傾向に加えて、  
昔からの風習や慣習が負担である回答の負荷量大

# 4. 導出主成分の意味付けの検証

主成分の意味付け	追加アンケートによる検証(評価のサマリ)
<p>①若手/ベテランの二極化</p>	<p><b>二極化の住民意識が確認できる:</b></p> <ul style="list-style-type: none"> <li>・町民全体で取り組む意識はそれなりに高い (問10#1 55.5%)</li> <li>・若者の意見や活躍の期待が大きい (問10#4/5: 73~75%)</li> </ul> <p>【世代別】</p> <ul style="list-style-type: none"> <li>・10~40代が比較的「問5#1 町民全体意識」が強そう</li> <li>・「問5#4/#5若者を前面に」意識は総じて若手が高そう(20代が最高、80代も)、問9#12回答(60代、40代、20代)</li> </ul>
<p>②若手(不満、消極的)、ベテラン(満足、積極的)</p>	<p><b>ベテラン(積極的)/若者(消極的)の意識が住民全体に認識されている:</b></p> <ul style="list-style-type: none"> <li>・リーダー/若者育成が必要である認識 (問10#2:63%/問24#5_46%)</li> </ul> <p>【世代別】「問10#2リーダー育成」30、50、60代の意識が高い。50~60代を指導者にした若手育成取組が必要かも。</p> <ul style="list-style-type: none"> <li>・若者が活躍できる職場(問9 44%)</li> </ul> <p>【世代別】問9#12回答(60代、20代、40代)</p> <ul style="list-style-type: none"> <li>・何らかのイベント参加:概ね積極的(73%)</li> </ul> <p>【世代別】問11 30→50→40代、10代は60代より低い、20代は70代より低い</p>



アンケート項目追加



追加アンケート回答の分析

2022年度に追加したアンケート

- 【問10#1】町の将来像の実現に町全体で取り組んで行く
- 【問10#2】xxxx町を支えるまちづくりの推進組織やリーダーの育成を図る
- 【問10#3】xxxx町に関わる外部のあらゆる関係機関・団体・個人と連携・協力体制をとる
- 【問10#4】若者の声を取り入れたまちづくりを推進する
- 【問10#5】若者が活動しやすい環境の形成に取り組む

- 【問9】xxxx町が目指すまちづくりについて、特に優先して取り組むべき項目:  
12 若者が活躍・定着できる魅力的な職場づくり、各職種人材が確保・育成されるまち
- 【問11】xxxx町のまちづくりや社会活動などの取り組みへのあなたの関心について、現在実際に参加しているものや、参加した経験があるものも含めて、あてはまるもの
- 【問24】【商業・にぎわいづくりについて】商業活性化について、特に力を入れるべきだと思う項目  
5 若者の就労支援やまちづくりの担い手育成支援
- 【問28】労働環境における課題: 3 後継者人材がいない、または継承が難しい

# 4. 導出主成分の意味付けの検証

主成分の意味付け	評価のサマリ
<p>③若手はベテランに遠慮している傾向</p>	<p><u>若者はベテランに遠慮している全体感がありそうである</u></p> <ul style="list-style-type: none"> <li>◆若者の意見や活躍の期待が大きい(#4/5: 73~75%)</li> <li>◆若者が活躍できる職場(問9 44%)</li> </ul> <p>【世代別】</p> <ul style="list-style-type: none"> <li>・「#4/#5若者を前面に」意識が総じて若手が高そう(20代が最高、80代も)、問9/#12回答(60代、40代、20代)</li> <li>・一方、②でイベント参加率: 10/20代 &lt; 60/70代</li> </ul> <p>★20代を中心に文字回答の深掘りが必要</p>
<p>④若手は、古い風習や文化、環境に不満を持っている</p>	<p><u>若手を主体に外部との協働体制を求めがちである:</u></p> <ul style="list-style-type: none"> <li>◆若者の意見や活躍の期待が大きい(#4/5: 73~75%)</li> <li>◆若者育成が必要である認識(問24#5_46%)</li> <li>◆全体として外部との連携により新しい習慣・文化を融合の意識(#3: 54%) (10~40代が高め)</li> </ul>



2022年度に追加したアンケート

- 【問10#1】町の将来像の実現に町全体で取り組んで行く
- 【問10#2】xxxx町を支えるまちづくりの推進組織やリーダーの育成を図る
- 【問10#3】xxxx町に関わる外部のあらゆる関係機関・団体・個人と連携・協力体制をとる
- 【問10#4】若者の声を取り入れたまちづくりを推進する
- 【問10#5】若者が活動しやすい環境の形成に取り組む

- 【問9】xxxx町が目指すまちづくりについて、特に優先して取り組むべき項目: 12 若者が活躍・定着できる魅力的な職場づくり、各職種人材が確保・育成されるまち
- 【問11】xxxx町のまちづくりや社会活動などの取り組みへのあなたの関心について、現在実際に参加しているものや、参加した経験があるものも含めて、あてはまるもの
- 【問24】【商業・にぎわいづくりについて】商業活性化について、特に力を入れるべきだと思う項目 5 若者の就労支援やまちづくりの担い手育成支援
- 【問28】労働環境における課題: 3 後継者人材がいない、または継承が難しい

# 5. まとめと今後の課題

## 【まとめ】

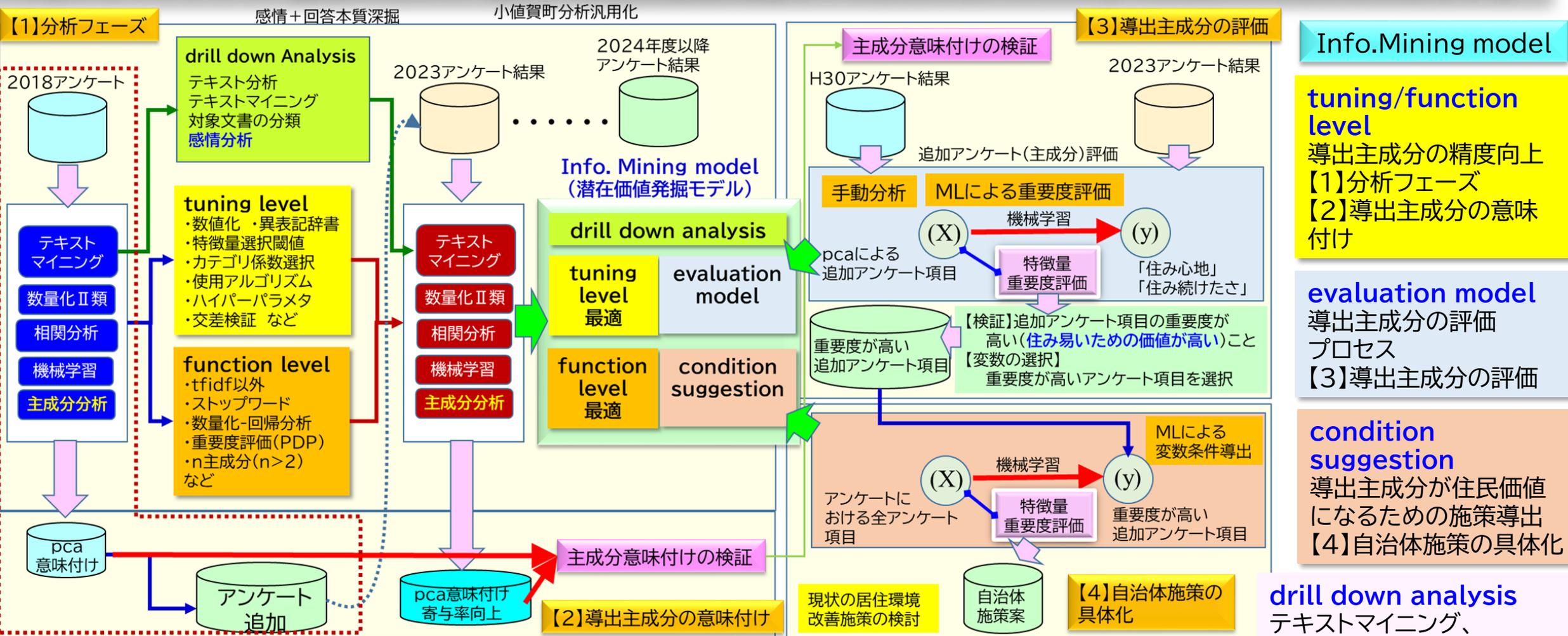
- ◎自治体既存の住民アンケート回答から「潜在的な住民価値」を見える化する多変量解析モデルを構築
- ◎本モデルを適用して導出した「住民価値」に関して、追加アンケート結果から一定の妥当性を検証することができた

## 【今後の課題】

- (1) 主成分分析の寄与率向上のためのモデル精度改善:
  - (a) 主成分分析の説明変数(特徴量)選択の最適化 [機械学習モデル精度向上]
  - (b) 文字回答に関する特徴語選択の最適化 [「住み易い」回答に影響度が高い形態素の選択]
  - (c) 感情分析などを活用した文字回答の深耕分析
- (2) 精度改善モデルを適用した2022年度アンケート回答分析の推進
- (3) 導出した「住民価値」の住民意識を高めるための居住条件の明確化と施策立案支援

# 【参考】モデル汎用化の構想

今後シビックプライド醸成のための「潜在的な住民価値発掘モデル」(Info.Mining model)の汎用化を探求する



Info.Mining model

tuning/function level  
導出主成分の精度向上  
【1】分析フェーズ  
【2】導出主成分の意味付け

evaluation model  
導出主成分の評価プロセス  
【3】導出主成分の評価

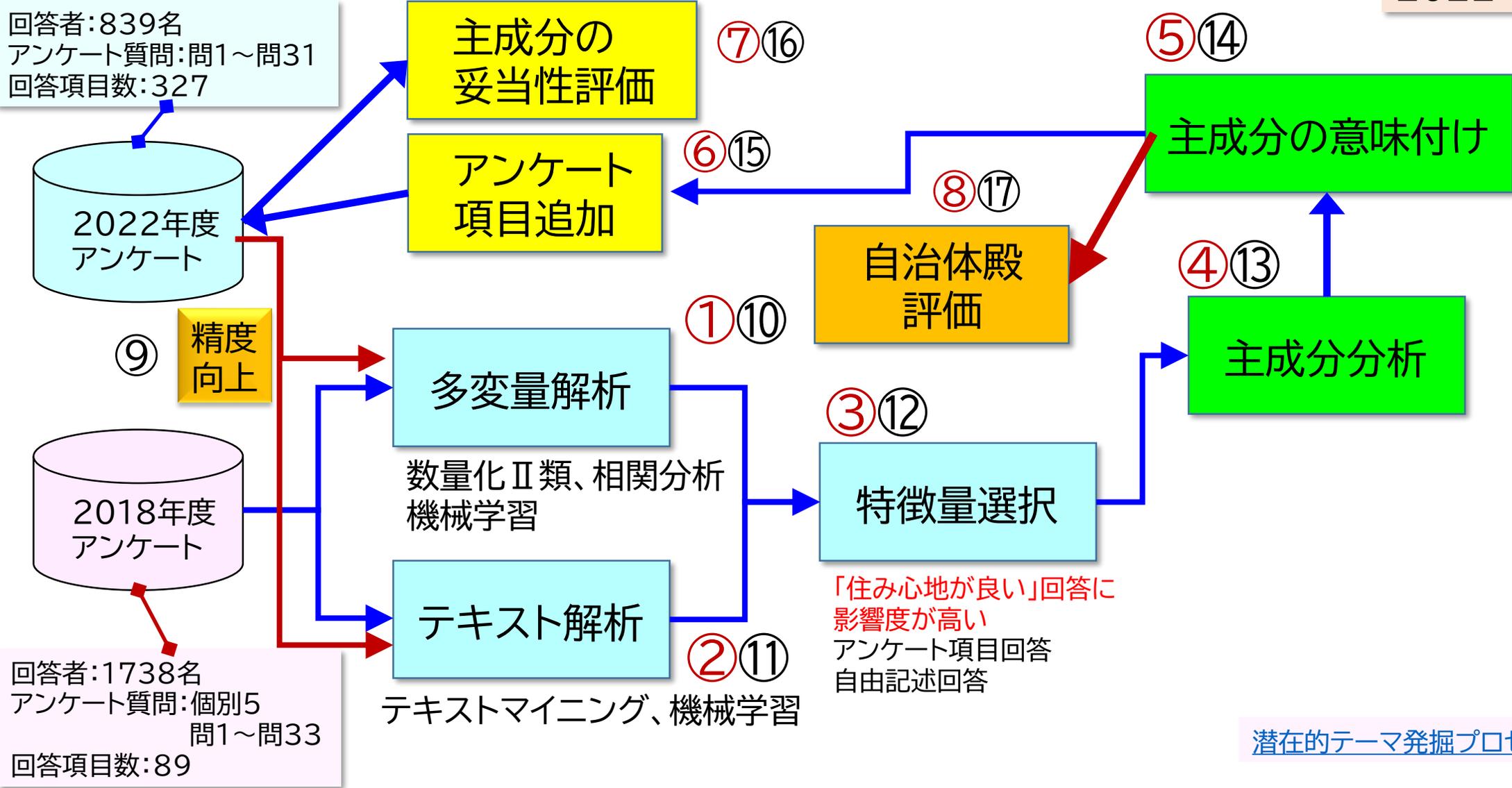
condition suggestion  
導出主成分が住民価値になるための施策導出  
【4】自治体施策の具体化

drill down analysis  
テキストマイニング、感情分析などを活用してアンケート文書の真意を深掘

# 【参考】2022年度アンケート分析について

対象アンケート:  
 2018年度①～⑧  
 2022年度⑨～⑰

## (a) 潜在的な新住民価値の発掘(2018～2022年アンケート対象)



潜在的テーマ発掘プロセス一覧

# 【参考】2022年度アンケート分析状況

## (1) 数量化Ⅱ類

### 【分析内容】

「問5 住み心地」に関して「住み易い」かどうかの質問回答に対して、「各アンケート項目」の影響度を数量化(スコア化)した

### 【制約条件】

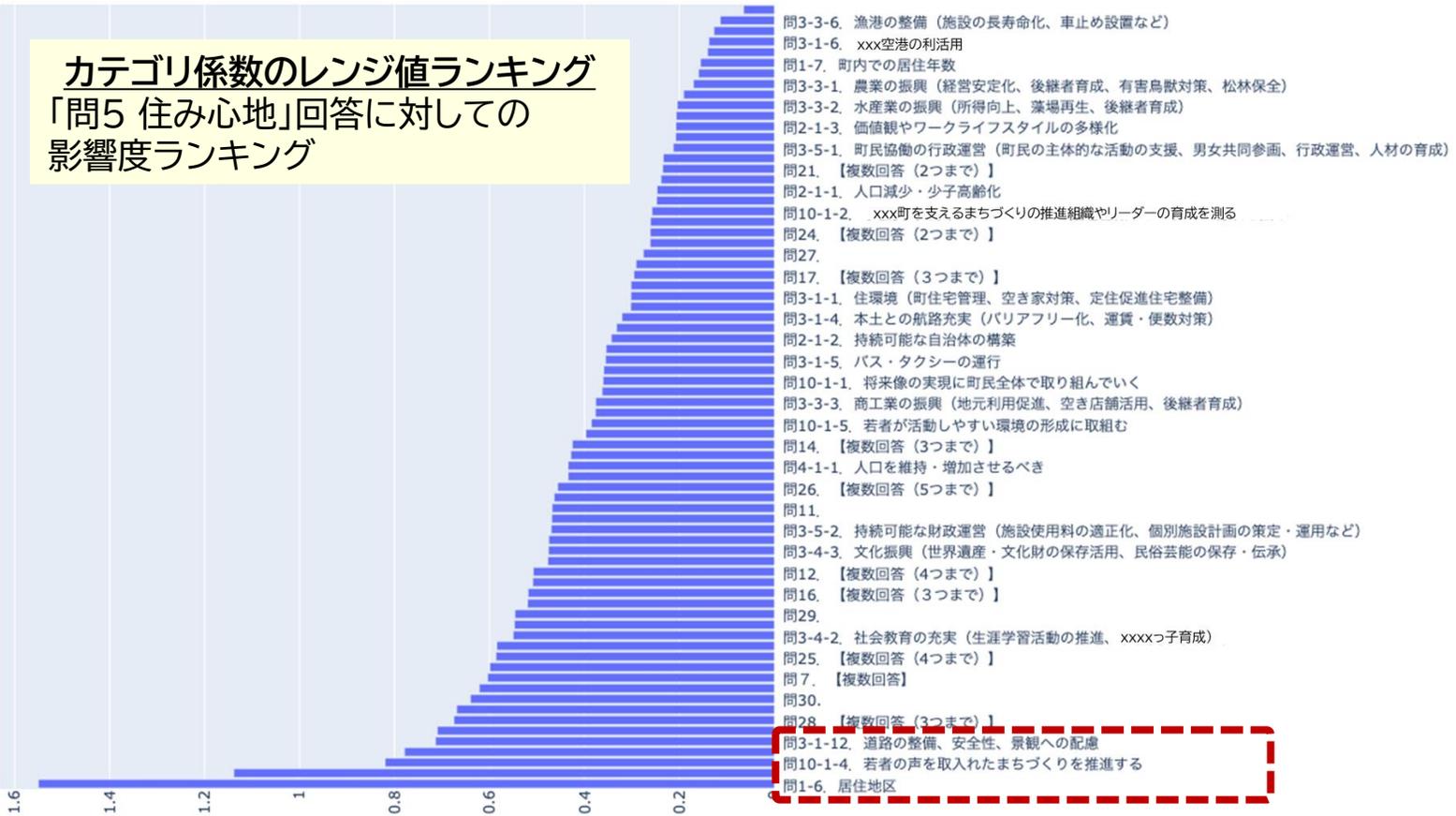
上記の影響度が高いアンケート項目の各回答が「住み易い」「住み易くない」のどちらに寄与したか(影響したか)は本分析では明確になっていない →次(a)(b)ステップで明らかに深耕予定

(a)数量化Ⅱ類(スコアの深耕)

(b)機械学習の重要度評価

# 【参考】2022年度アンケート分析状況 数量化Ⅱ類

カテゴリ係数のレンジ値ランキング  
「問5 住み心地」回答に対しての  
影響度ランキング

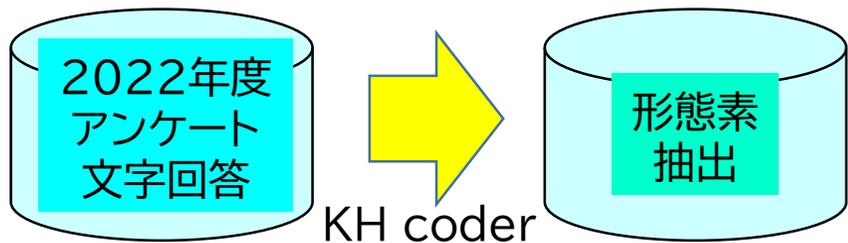


問 項 番	内 容	属 性	ス コ ア
1-6	居住地区	個別	1.550
10-1-4	若者の声 まちづくり	新規	0.820
3-1-12	道路整備、安全性、景観	継続	0.714
1-1	性別	個別	0.709
28	労働環境の課題	継続	0.675
3-1-2	防犯・防災・消防体制充実	継続	0.670
30	xxxx生活での幸せ感	新規	0.640
2-1-6	デジタル社会の到来	継続	0.621
9	まちづくり 優先取組項目	新規	0.600
25	産業振興で優先取組項目	新規	0.586
1-5	所帯人数	個別	0.584
3-4-2	社会教育の充実	継続	0.550
3-4-1	学校教育の充実	継続	0.547
29	xxxxの魅力	新規	0.546
3-1-8	ごみ・し尿対策	継続	0.520
16	子育て環境	継続	0.519
1-4	職業	個別	0.509
12	デジタル環境整備の項目	継続	0.507

- ◆「住み心地」は「居住地区」の依存度が最高。(前回はスコア2位)
- ◆事実上最も影響度が高い特徴量(項目/カテゴリ)として「若者の声を取り入れた街づくり」であり、若者/ベテランの乖離と世代交代要望の認識の高さがうかがえる
- ◆生活環境に関する項目でトップの重要度が「道路の整備、安全性、景観への配慮」  
前回主成分分析でも負荷量が高い項目

# 【参考】2022年度アンケート分析状況

## (2) テキストマイニング



- ◆形態素抽出
- ◆形態素の出現頻度からTF-IDFを算出

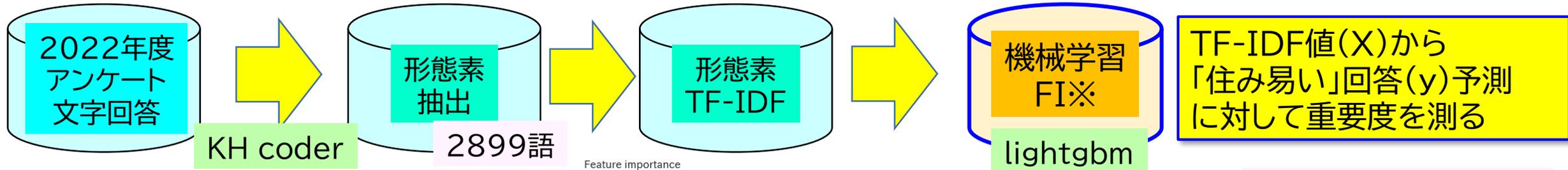
**TF-IDF:**  
 文書群において単語がどのくらい特徴的かを示す指標

形態素	出現頻度
(町名)	155
人口	154
町	85
海	72
生活	67
減少	63
高齢	57
子供	54
住む	51
維持	48
増加	48
仕事	44
自然	43
町民	37
幸せ	37
若い	36
若者	35
家族	35
存続	35
将来	34

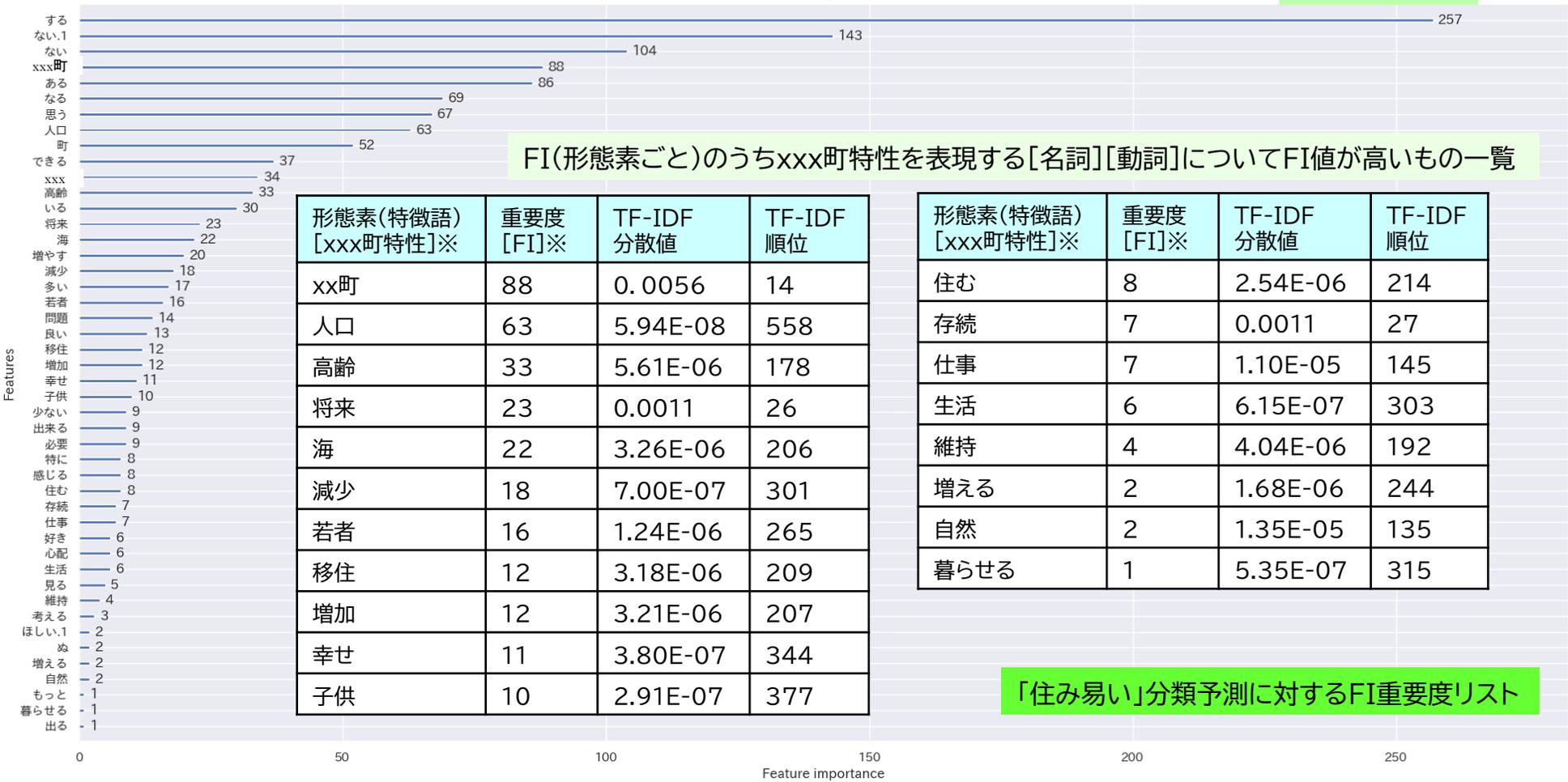
形態素	TF-IDF分散
yyyy町	56.6633
夕日	1.5982
プライバシー	0.1576
人材	0.0474
新田	0.0281
不足	0.0250
松原	0.0199
姫	0.0199
痛ましい	0.0074
開発	0.0062
(町名)	0.0056
発展	0.0037
過疎	0.0025
白浜	0.0025
後継	0.0022
不安	0.0020
事故	0.0016
導入	0.0012
専門医	0.0012
交通	0.0012

形態素	TF-IDF合計
yyyy町	170.3299
夕日	28.6260
プライバシー	8.9832
人材	4.9546
松原	4.6930
姫	4.6930
新田	3.7898
不足	3.5933
痛ましい	1.9404
(町名)	1.8852
開発	1.7827
発展	1.5477
後継	1.4427
過疎	1.3377
白浜	1.1229
タクシー	1.0950
不安	1.0862
事故	0.9132
存続	0.9101
活性	0.8658

# 【参考】TF-IDFの重要度ランキング(vs「住み易い」)



TF-IDF値(X)から「住み易い」回答(y)予測に対して重要度を測る



#X\_train\_accuracy accuracy: 0.9363  
#X\_test\_accuracy accuracy: 0.7484

[FI]※  
・明示的な変曲点から重要度「1」以上の**40変数**を選択ランキング  
・ストップワード含む暫定

「住み易い」分類予測に対するFI重要度リスト

※ xxx町特性: xxx町特性を表す語  
FI: Feature importance

# 【参考】TF-IDFの説明

## TF-IDF

- ◆TF-IDFは自由記述回答から特徴語をマイニングするため、文書群において単語の特徴度合を表す指標である
- ◆**文書群の中で特徴的な語ほど、TF-IDFの値が高い。**

情報理論において確率 $p$ で生起する事象が起こったことを知ったときに得られる情報量を $I(p)=-\log p$ で表す。IDFは情報量と同じ意味付けである。

IDFの場合すべての文書が同様に確からしいとは限らない。しかし、TF-IDFは理論的な基礎ははっきりしないが、有用なため、テキストマイニングや検索エンジンなどで幅広く使われている。

## TF-IDF

TF : Term Frequency 単語頻度

それぞれの文書について、その単語が出てくる程度

IDF : Inverse Document Frequency 逆文書頻度

全体の文書のうち、その単語を含む文書の程度(の逆数)、  
複数の文書に出現する単語ほど特徴的でない

$$TF = \frac{\text{単語}t\text{の出現回数}}{\text{文書内の総単語数}} \quad IDF = \log \frac{\text{総文書数}}{\text{単語}t\text{を含む文書の数}}$$

$$\mathbf{TF-IDF} = TF \times IDF = \text{単語使用頻度} \times \text{単語レア度}$$

# 【参考】状況居住地区ごとの「住み易い」意識

	2022年度			2018年度		
	今回 住み易い	住み易くない	合計	前回 住み易い	住み易くない	合計
Aエリア	314	118	432	563	242	805
	72.7%	27.3%		69.9%	30.1%	
Bエリア・Cエリア	1	0	1	3	0	3
	100.0%	0.0%		100%	0%	
Dエリア	85	32	117	161	71	232
	72.6%	27.4%		69%	31%	
Eエリア	46	15	61	72	32	104
	75.4%	24.6%		69%	31%	
Fエリア	34	13	47	72	26	98
	72.3%	27.7%		73%	27%	
Gエリア	51	13	64	125	45	170
	79.7%	20.3%		74%	26%	
Hエリア	26	11	37	75	36	111
	70.3%	29.7%		68%	32%	
Kエリア	12	6	18	31	9	40
	66.7%	33.3%		78%	23%	
Lエリア	17	9	26	38	9	47
	65.4%	34.6%		81%	19%	
Nエリア	7	4	11	3	9	12
	63.6%	36.4%		25%	75%	

# 参考文献

- [1] 読売広告都市生活研究局(著):シビックプライド-都市のコミュニケーションをデザインする
- [2] Hotelling, H. : “Analysis of a complex of statistical variables into principal components”. *Journal of Educational Psychology* 24: 417-441, 498-520,(1933).
- [3] 菅 民郎, 藤越 康祝(著): 質的データの判別分析 数量化Ⅱ類.
- [4] 佐藤浩輔:町根大学人間科学部2019.07.13,応用心理学研究Ⅰ,テキストマイニング講義資料,  
<https://www.slideshare.net/cos039840935/ss-155407947>
- [5]“PYCARET“. <https://pycaret.org/>, (参照 2023-06-27).
- [6] 涌井良幸, 涌井貞美, ”実習多変量解析入門”,技術評論社(2011)
- [7]相澤彰子, “語と文書の共起に基づく特徴度の数量的表現について“. 情報処理学会論文誌, Vol.41, No.12, pp.3332-3343.
- [8]塚本邦尊他, “東京大学のデータサイエンティスト育成講座“ マイナビ出版(2019)
- [9]高田 晃希,黒羽 晟,山本 裕,橋本 沙也加,橋本 尚子,岡田 ゆかり, ”シビックプライド醸成に繋がる住民価値の掘り起こしと貢献度の検証に関する研究”,1ZH-03 情報処理学会第85回全国大会論文誌.
- [10]江崎雄治, “居住環境から見た住民の価値意識”, 地理学評論, 68A-3 168-179 1995.
- [11]” 地域の情報化への取組と地域活性化に関する調査研究報告書”, 総務省 情報通信政策局 総合政策課 情報通信経済室(平成20年3月)
- [12]森田哲夫,諸岡峻一,塚田伸也,橋本隆, ”テキストマイニングを用いた自由記述データの有効活用に関する研究”,  
[http://library.jsce.or.jp/jsce/open/00039/201306\\_no47/pdf/398.pdf](http://library.jsce.or.jp/jsce/open/00039/201306_no47/pdf/398.pdf)
- [13]森田哲夫,入澤覚,長塩彩夏,野村和広,塚田伸也,大塚裕子,杉田浩, ”自由記述データを用いたテキストマイニングによる都市のイメージ分析”,土木学会論文集D3(土木計画学),Vol.68, No.5(土木計画学研究・論文集第29巻),I\_315-I\_323,2012.
- [14]金井茂樹, ”テキストマイニングによる「市民の声」の分析”, [https://www.jstage.jst.go.jp/article/jichitaigaku/28/2/28\\_42/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jichitaigaku/28/2/28_42/_pdf/-char/ja)
- [15]”平成17年度国土施策創発調査「地域プライド創発による地域づくりのあり方に関する調査」【調査報告書】”,文部科学省生涯学習政策局,文化庁文化庁文化財部,国土交通省都市・地域整備局(平成18年3月).
- [16]”143 都市環境はいかにシビックプライドを高めるか -今治市を事例とした実証分析”,公益社団法人 日本都市計画学会 都市計画論文集 Vol.52 No.3(2017年10月)
- [17] scikit-learn公式,sklearn.ensemble.RandomForestClassifier: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

ここから参考資料

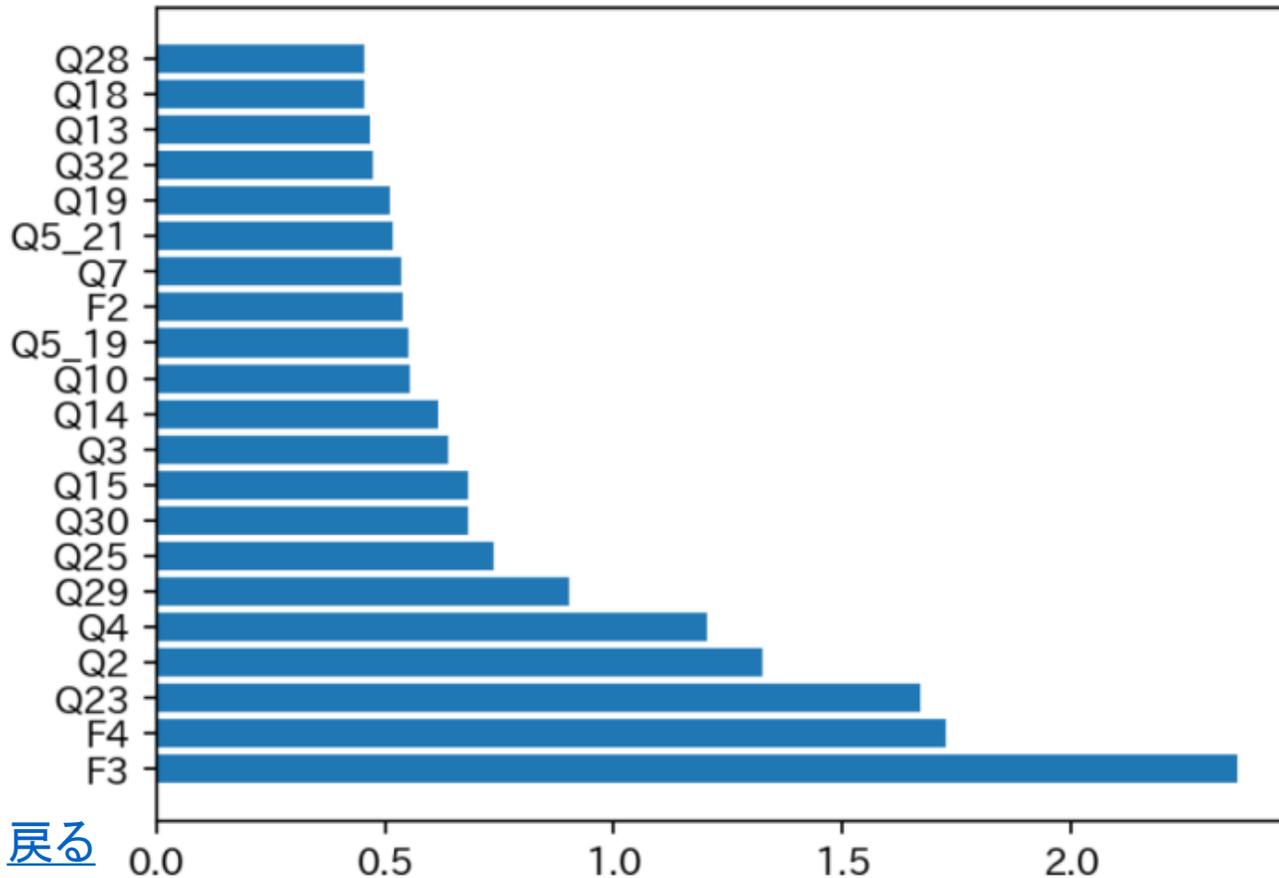
# 【参考】説明変数レンジランキング(数量化Ⅱ類)

【対象】2018年度アンケート回答

説明変数レンジランキング 1~21位

職業が最も重要度が高い

レンジグラフ



[戻る](#)

項目ID	内容	説明変数レンジ
F3	職業	2.364
F4	居住地区	1.726
Q23	訪町者からの協力金	1.671
Q2	住み続けたさ	1.325
Q4	町を離れる理由	1.203
Q29	配食サービス	0.903
Q25	宿泊施設	0.737
Q30	配食サービス	0.681
Q15	土地の利用	0.681
Q3	住み続けたい理由	0.639
Q14	公共施設設備	0.615
Q10	産業振興	0.553
Q5_19	小中学校の教育	0.552
F2	年代	0.539
Q7	交通	0.537
Q5_21	生涯スポーツ	0.517
Q19	世界文化遺産登録	0.509
Q32	配食の金額	0.474
Q13	まちづくり、ボランティア	0.468
Q18	インターネット環境	0.456
Q28	自分で料理できるか	0.454

# 【参考】新住民価値導出のための特徴量の選択

[戻る](#)

## 特徴量選択の意義とプロセス

### 【特徴量選択の目的】

- ・新住民価値の導出方法として主成分分析を活用する。
- ・新住民価値はシビックプライド醸成に寄与するベクトルになることが必要なため、入力する説明変数(特徴量)は「住み心地」回答に影響度・重要度が高いアンケート項目および自由記述回答を選択する。

目的/プロセス	手法	手法の概要
アンケート項目回答の解析(特徴量選択) (多重共線性排除)	数量化Ⅱ類	「住み心地」の群を最も分離するアンケート項目(特徴量)の重み付け、特徴量の重み(「住み心地」に対する寄与率)評価、寄与率ランキングでアンケート回答項目を選択(カテゴリウェイト/レンジが0.15以上)
	相関分析	数量化Ⅱ類で選択した特徴量に対してクロス集計を行い、変数間のクラメール連関係数を評価
アンケート自由記述回答の解析(特徴語選択)	テキストマイニング	形態素解析[KH Coder]で、特徴語のTF-IDFを評価(コーパスごとに)、特徴語ごとのTF-IDF値ランキング50を選択。選択した特徴語を数量化(One-hot encoding)
主成分分析のための特徴量選択	機械学習(教師あり分類)	・数量化Ⅱ類で選択した特徴量とテキストマイニングで選択した特徴語をマージ。機械学習に入力する特徴量を生成。[1次選択] ・精度が高いアルゴリズムを評価選択(PyCalet)。 ・「住み心地」回答変数を目的変数、上記を説明変数とし機械学習モデルで学習/予測/精度評価。特徴量の重要度を評価し50個の特徴量を選択[2次選択]

# 【参考】多変量解析 → 主成分分析での新変数の創出

## 主成分分析

- ・教師なし学習の一つ。ある多変量データからそのデータを要約してくれる潜在的なベクトル(主成分)を導出できる。
- ・主成分分析は、ベクトルにデータの分散を射影し、分散を最大化するようにそのベクトル(主成分)を決定する
- ・**主成分**は、データ全体の情報(分散)を豊富に表現されたベクトルであり、多変量データをそのベクトルで要約可能。
- ・最も大きい分散が射影されたベクトルが**第一主成分**、次に大きい分散が射影されたベクトルが**第二主成分**である。

## 特徴量の選択

## 特徴量の絞り込み

## 数量化Ⅱ類

- ◆ 群データで与えられる目的変数と質的データで与えられる説明変数との関係を**モデル式**で表し、**モデル式**によって、説明変数と目的変数との関連性を明らかにする手法。
- ◆ アンケート調査の分析などに用いられてきた解析手法である。

## クラメール連関係数

質的データと質的データの相関を求める尺度。  
クラメール連関係数はクロス集計表の関連性の度合いから算出

## 機械学習

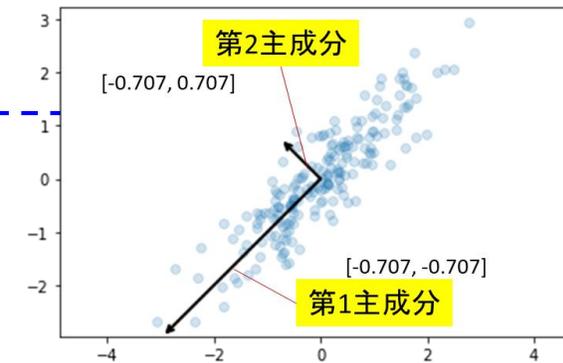
- ◆ 教師あり(分類)の手法を用いて目的変数に該当する変数を分類予測するモデル
- ◆ 入力する説明変数(特徴量)は分類予測の精度を高めるために最適選択する
- ◆ 目的変数の予測に必要な説明変数の重要度を数値で算出する

# 【参考】主成分分析活用の意味付け

- ◆新住民価値を導出方法として「主成分分析」を活用。  
主成分分析は多変量データの情報を最大限保持するように、軸をうまく選ぶ手法
- ◆主成分分析で導出される軸(主成分)は、射影されたデータの分散を最大化される軸。  
つまり、主成分分析に入力する「特徴量」の情報を失うことなく、全体の分布を最もよく表す軸(主成分)を見つけて、軸の定義を行う(主成分の意味付け)
- ◆主成分分析に入力する変数(特徴量)として、事前の多変量解析で「住み心地が良い」アンケート回答に影響度が高いアンケート項目回答(変数)を選ぶ。つまり「シビックプライドの醸成」に影響度が高いアンケート項目回答を、特徴量として主成分分析に入力。
- ◆上記特徴量データの情報を保持したまま、特徴量データを要約できるような新しい軸を導き出す。→潜在的な新住民価値

- (1)左記特徴量のデータの情報を最大限保持するように軸(ベクトル)を選ぶ(主成分)
- (2)特徴量のデータを主成分に射影したときの分散が最大となるように軸(主成分)の式を決める
- (3)軸(主成分)の意味付けをする

例えば2つのアンケート項目:(x)日常の買い物の便利さ (y)住み易さ



# 【参考】導出主成分の意味付けの検証

pc1 未来改善(不)志向度: [正]参加意欲小、不満、消極的、若手(10~50代)

[負]参加意欲大、満足、積極的、ベテラン(60~90代)、インフラ・福祉への関心高い

主成分の意味付け	検証評価の基準	評価のサマリ
①若手/ベテランの二極化	以下の問10の回答の比率 #1(1重要である)、#4(重要である)、 #5(重要である)	<b>二極化の住民意識が以下より確認できる:</b> ◆町民全体で取り組む意識はそれなりに高い(#1 55.5%) ◆若者の意見や活躍の期待が大きい(#4/5: 73~75%) 【世代別】 ・10~40代が比較的「#1町民全体意識」が強そう ・「#4/#5若者を前面に」意識は総じて若手が高そう(20代が最高、80代も)、問9/#12回答(60代、40代、20代)
②若手(不満、消極的)、 ベテラン(満足、積極的)	(a)問10の回答 #2(1重要である)、#4(重要である)、 #5(重要である)  (b)問9 #12の選択回答をしている (c)問24 #5を選択している (d)問11 #1~11に参加しているか 【積極性を測る】	<b>ベテラン(積極的)/若者(消極的)の意識が住民全体に認識されている:</b> ◆リーダー/若者育成が必要である認識(#2:63%/問24#5 46%) 【世代別】「#2リーダー育成」30、50、60代の意識が高い。50~60代を指導者にした若手育成取組が必要かも。 ◆若者が活躍できる職場(問9 44%) 【世代別】問9/#12回答(60代、20代、40代) ◆何らかのイベント参加:概ね積極的(73%) 【世代別】問11 30→50→40代、10代は60代より低い、20代は70代より低い

対応するアンケート項目:

- 【問10#1】町の将来像の実現に町全体で取り組んで行く
- 【問10#2】xxxx町を支えるまちづくりの推進組織やリーダーの育成を図る
- 【問10#3】xxxx町に関わる外部のあらゆる関係機関・団体・個人と連携・協力体制をとる
- 【問10#4】若者の声を取り入れたまちづくりを推進する
- 【問10#5】若者が活動しやすい環境の形成に取り組む

【問9】xxxx町が目指すまちづくりについて、特に優先して取り組むべき項目:

- 12 若者が活躍・定着できる魅力的な職場づくり、各職種人材が確保・育成されるまち
- 【問11】xxxx町のまちづくりや社会活動などの取り組みへのあなたの関心について、現在実際に参加しているものや、参加した経験があるものも含めて、あてはまるもの
- 【問24】【商業・にぎわいづくりについて】商業活性化について、特に力を入れるべきだと思う項目  
5 若者の就労支援やまちづくりの担い手育成支援
- 【問28】労働環境における課題: 3 後継者人材がない、または継承が難しい

# 【参考】導出主成分の意味付けの検証

pc1 未来改善(不)志向度: [正]参加意欲小、不満、消極的、若手(10~50代)

[負]参加意欲大、満足、積極的、ベテラン(60~90代)、インフラ・福祉への関心高い

主成分の意味付け	検証評価の基準	評価のサマリ
③若手はベテランに遠慮している傾向	(a)問10の回答 #4(重要である)、#5(重要である)  (b)問9 #12の選択回答をしている	<b>若者はベテランに遠慮している全体感がありそうである</b> ◆若者の意見や活躍の期待が大きい(#4/5: 73~75%) ◆若者が活躍できる職場(問9 44%) 【世代別】 ・「#4/#5若者を前面に」意識が総じて若手が高そう(20代が最高、80代も)、問9/#12回答(60代、40代、20代) ★今少し深掘りが必要。文字回答データに関して、20代を中心に深掘りをする
④若手は、古い風習や文化、環境に不満を持っている	(a)問10の回答 #3(1 重要である)、#4(重要である)、#5(重要である)  (b)問9 #12の選択回答をしている (c)問24 #5を選択している (d)問28 #3を選択しているか	<b>若手を主体に外部との協働体制を求めがちである:</b> ◆若者の意見や活躍の期待が大きい(#4/5: 73~75%) ◆若者育成が必要である認識(問24#5 46%) ◆全体として外部との連携により新しい習慣・文化を融合の意識(#3: 54%) (10~40代が高め)

対応するアンケート項目:

- 【問10#1】町の将来像の実現に町全体で取り組んで行く
- 【問10#2】xxxx町を支えるまちづくりの推進組織やリーダーの育成を図る
- 【問10#3】xxxx町に関わる外部のあらゆる関係機関・団体・個人と連携・協力体制をとる
- 【問10#4】若者の声を取り入れたまちづくりを推進する
- 【問10#5】若者が活動しやすい環境の形成に取り組む

- 【問9】xxxx町が目指すまちづくりについて、特に優先して取り組むべき項目:  
12 若者が活躍・定着できる魅力的な職場づくり、各職種人材が確保・育成されるまち
- 【問11】xxxx町のまちづくりや社会活動などの取り組みへのあなたの関心について、  
現在実際に参加しているものや、参加した経験があるものも含めて、あてはまるもの
- 【問24】【商業・にぎわいづくりについて】商業活性化について、特に力を入れるべきだと思う項目  
5 若者の就労支援やまちづくりの担い手育成支援
- 【問28】労働環境における課題: 3 後継者人材がいない、または継承が難しい

ここから補足説明資料

# 既存研究との関連性について -研究背景と課題

◆類似研究の調査による本研究の新規性：アンケート(選択回答+自由記述回答)から潜在的な住民価値を発掘し、施策を導出する事例は少ない。

## 【研究の背景】

各自治体において、シビックプライドの改善の取り組みが活発になっている昨今、現在から未来へ、**新しい住民価値を創出する各種検討**が推進されている。取組の殆ど、**住民アンケートをベースに地域の特性(衣食住、医療、福祉、生活基盤、産業など)を把握して、各自治体の担当者が回答データを分析して施策を企画・立案する実情が多い。**

## 【研究の対象となる課題】

居住に関する既存の住民アンケート(**アンケート項目選択+自由記述の形式**)から、シビックプライドに繋がる住民意識・価値を高めるための**潜在的なテーマを導出する、分析・対策プロセスが確立できていない**という課題を認識する。

本研究では、**既存の住民アンケート結果(アンケート項目+自由記述の形式)**から、**潜在的な新しい「住民価値を高める」テーマを導出し、そのテーマが住み心地意識をどう改善でき、意識改善するための具体的な居住環境の改善とは何なのかを、導出・評価するプロセスを明確にする。**また、そのプロセスの汎用化に繋げる研究活動とする。

### (a) 選択回答アンケート項目分析

住民アンケートをベースに傾向分析を実施し、傾向把握のための分類などを目的とした、主成分分析を活用し導出した新しい軸を使ったデータ解析などの例がある。(2)(3)

選択式アンケート回答結果の多変量解析方法において、潜在的な住民価値と施策を導くような分析手法は明らかにされていない

### (b) 選択回答+自由回答分析

選択式アンケート項目と自由記述回答の関連性を評価するテキストマイニング事例がある。(4)(5)(6)

選択式アンケートと自由回答アンケートの両方を重みづけ評価を行い、アンケート全体から潜在的な住民価値を導く分析手法は明らかにされていない

### (c) シビックプライドを高める活動

都市環境に対する価値の住民アンケートからの因子分析などによる地域に対する意識を明確化するなどの例がある。(6)

自治体既存のアンケートを使い、居住環境全般にわたる、住民価値向上の施策を導出するための潜在価値を発見する手法は明らかでない

# 既存研究との関連性について -類似研究の調査(2/4)

## 【(a)選択回答アンケート項目分析】

- (1) 江崎雄治:居住環境から見た住民の価値意識, 地理学評論, 68A-3 168-179 1995,  
[https://www.jstage.jst.go.jp/article/grj1984a/68/3/68\\_3\\_168/pdf/-char/ja](https://www.jstage.jst.go.jp/article/grj1984a/68/3/68_3_168/pdf/-char/ja).  
<概要, 手法> 居住環境の各要素に対して、住民が無意識のうちに与えている「重み」を抽出する。目的変数を居住環境の総合的な満足度、説明変数を個別要素に対する満足度として重回帰分析をし、得られた偏回帰係数を各要素に与えられた「重み」とする。  
<本研究との差異> アンケートから重要度(重み)を抽出する観点は類似しているが、そこから「潜在的な住民価値」を導出する点が異なる。
- (2) 加藤潤三:地域コミュニティに対する住民の価値を測定する—『コミュニティ価値』尺度の作成と検討—, 立命館産業社会論集, 55巻3号, 55 - 66, 2019-12  
<http://doi.org/10.34382/00012886>.  
<概要, 手法> 住民が「コミュニティ価値として重視していると考えられる15の諸要素」を100点配分するアンケート調査(全国11ブロック人口比に応じた651名Web調査)から重要度判断。上記をクラスタ分析で5クラスタに分類、別アンケート項目の「コミュニティに対する態度(コミュニティ意識・コミュニティ感覚、および行動(住民参加))との関連性(相関)、『コミュニティ価値』を量的に測定する尺度を作成。要素項目⇔意識、感覚、行動回答項目との相関分析。  
<本研究との差異> 既存の自治体アンケートを元に住民価値を分析するところ(本研究)が差異。当該研究は、地域コミュニティ価値と意識に関する研究目的のアンケートを元に分析。自由回答とアンケート項目回答を含めた全体アンケート結果から潜在価値を発見する(本研究)ところが差異。
- (3) ICTの活用実態と地域活性化との相関関係の把握への主成分分析の活用  
[https://www.soumu.go.jp/johotsusintokei//linkdata/other034\\_200803\\_hokoku.pdf](https://www.soumu.go.jp/johotsusintokei//linkdata/other034_200803_hokoku.pdf)  
<概要, 手法> 自治体ごとのICT分野別活用指標(アンケート)から、ICTシステム間の関係の強さなどの構造を示す新たな指標を、主成分分析を用いて導出する(新指標:福祉軸、地域活性化軸)  
→ICTシステムがどの分野に使われている傾向が強いかを、新しい軸を設定して傾向分析する  
<本研究との差異> 当該研究はICT活用実態に特化しているものの主成分分析を用いて新たな指標を導出しているアプローチは同じで、主成分分析で導出する価値の精度を高めるために、主成分分析の説明変数を選択している。  
(機械学習)部分と、自由回答をアンケート項目回答を併せてアンケート結果の重要度を測っているところが差異。

# 既存研究との関連性について-類似研究の調査(3/4)

## 【(b)選択回答+自由回答分析】

(4)テキストマイニングによる「市民の声」の分析

[https://www.jstage.jst.go.jp/article/jichitaigaku/28/2/28\\_42/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jichitaigaku/28/2/28_42/_pdf/-char/ja)

[https://www.jkk-labo.jp/wp-content/uploads/2016/11/textmining\\_shiryo.pdf](https://www.jkk-labo.jp/wp-content/uploads/2016/11/textmining_shiryo.pdf)

(5)自由記述データを用いたテキストマイニングによる都市のイメージ分析

[https://www.jstage.jst.go.jp/article/jscejpm/68/5/68\\_I\\_315/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/jscejpm/68/5/68_I_315/_pdf/-char/en)

(6)テキストマイニングを用いた自由記述データの有効活用に関する研究

[http://library.jsce.or.jp/jsce/open/00039/201306\\_no47/pdf/398.pdf](http://library.jsce.or.jp/jsce/open/00039/201306_no47/pdf/398.pdf)

### <概要・手法>

- ・コレスポネンス分析とクラスタ分析や共起度を活用し、アンケート項目と自由記述回答の関連性を評価するテキストマイニング事例(4)(5)(6)
- ・アンケート項目に関してコレスポネンス分析で絞り込んだ課題に関連する「改善」に関する特徴的な語をキーにして課題を深掘りするアプローチ(4)
- ・コレスポネンス分析+クラスタ分析の結果と、共起度分析の結果により、テキストマイニングの結果導出した語の重要度の定量評価をするアプローチ(5)
- ・アンケート項目回答結果の分析プロセスにおいて、テキストマイニングで補完するアプローチ(6)

### <本研究との差異>

アンケート項目と自由記述回答全体から、回答における重要度が高い特徴量(特徴語)を選択、主成分を導出することで潜在的な住民価値を見つけ出す手法を明らかにしようとしている部分が差異

# 既存研究との関連性について -類似研究の調査(4/4)

## 【(c)シビックプライドを高める活動】

シビックプライドの意義としては「地域プライド」として2005年に国土交通省と文部科学省、文化庁が共同まとめた『地域プライド創発による地域づくりのあり方に関する調査』(平成18年3月)の中で定義されている

[https://www.mlit.go.jp/kokudokeikaku/souhatu/h17seika/6pride/06\\_syu.pdf](https://www.mlit.go.jp/kokudokeikaku/souhatu/h17seika/6pride/06_syu.pdf)

(7) 143 都市環境はいかにシビックプライドを高めるか -今治市を事例とした実証分析,公益社団法人 日本都市計画学会 都市計画論文集 Vol.52 No.3 2017年10月

### <概要・手法>

アンケート調査の結果に対して、因子分析と共分散構造分析を用いて、シビックプライドの因子及び、都市環境の評価がシビックプライドに

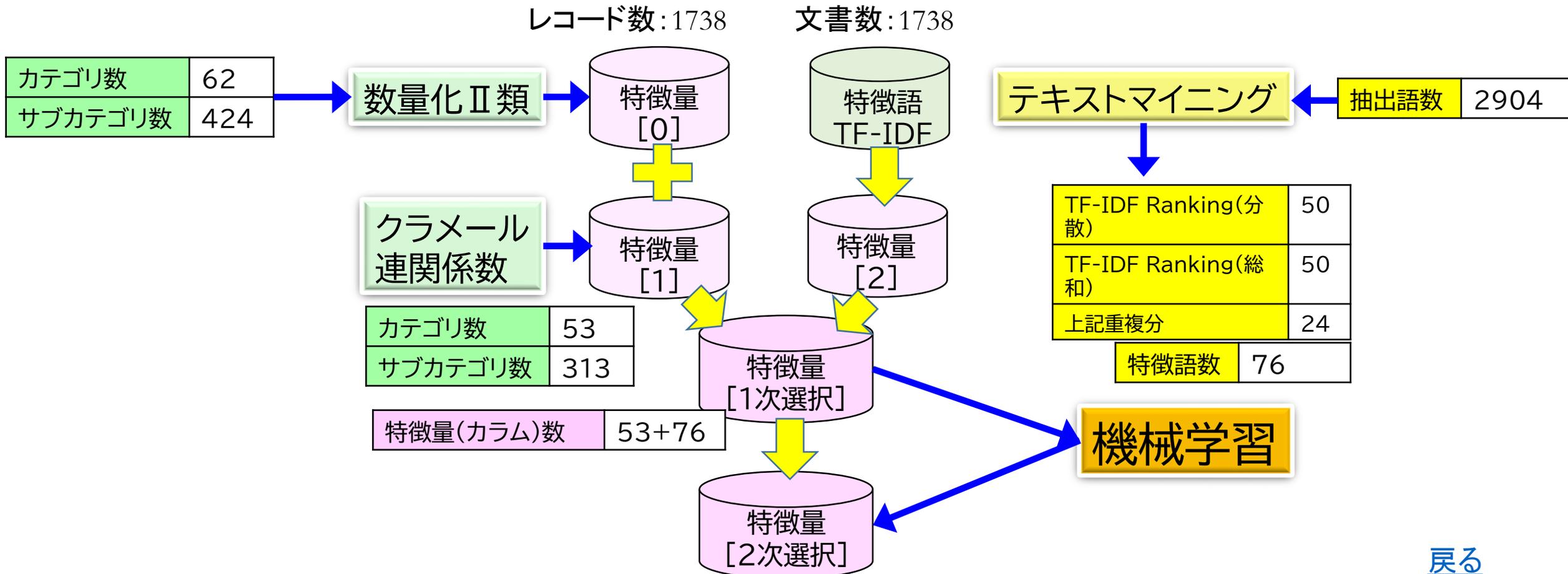
及ぼす影響構造を明らかにする。都市が構築した環境(XXX公園、XXX商店街など)に対する価値の住民アンケートからの因子分析による、住民の潜在意識から地域に対する気持ちの明確化を実施。各因子の各アンケート項目(地域愛着(選好)、地域愛着(感情)などのシビックプライド尺度)に対する因子負荷量評価で意味付けを実施。そこからの潜在意識の推定を実施。

### <本研究との差異>

一部の都市環境に関して、地域に対する意識・感情についてのアンケート結果を因子分析して潜在意識の推定を行うアプローチは同様のものであるが、既存のアンケートから潜在的な価値をあぶり出し、その価値に対する意識の評価と意識を高めるための分析手法は明らかにされていない。本研究では、既存アンケートから主成分分析を用いて、住民価値に繋がる潜在的な住民価値をあぶり出し、評価を行って、当該価値に対する住民意識が高まるような居住条件を導出する手法を明らかにするものである。

# 特徴量1次選択 特徴量の絞り込み

[数量化Ⅱ類] (a)特徴量(入力): カテゴリ62項目/サブカテゴリ424項目、(b)特徴量(絞込): **カテゴリ53項目**  
 [クラメール連関係数] 削減する特徴量(マルチコ排除): カテゴリ9項目、(c)特徴量(マルチコ排除後): **サブカテゴリ313項目**  
 [テキストマイニング] 特徴語を 2904(形態素解析の結果) → **特徴語 76語**を選択 (TF-IDF値ランキングより)



[戻る](#)

# 特徴量1次選択 数量化Ⅱ類/相関分析

戻る

- ・「住み心地」回答への寄与率が高い、その他アンケート項目回答(特徴量)の寄与率(カテゴリスコア)を算出 [数量化Ⅱ類]  
→寄与率が低い特徴量を削減
- ・特徴量の説明力を上げ、予測精度を高めるため、マルチコ(多重共線性)の変量を集約 [相関分析/クラメール連関係数]  
→クラメール連関係数が高い特徴量を集約

## 数量化Ⅱ類結果から特徴量を選択

- ◆**相関比**(「住み心地」「住みたくない」の2群が離れている割合)が、**最大**になるように**各特徴量(アンケート項目)の重み(寄与率:カテゴリウイト)**を算出。  
→「住み心地」「住みたくない」を判別するための寄与率(カテゴリウエイト)
- ◆寄与率ランキングでアンケート項目を選択:  
特徴量選択: カテゴリウエイトのレンジ(max-minの差)が0.15以下を削除 [寄与率が低い]
- ◆削除したカテゴリ(寄与率が低いアンケート項目): 9項目

アンケート#	内容サマリ
Q5-1-2	町内交通(バス・渡船)
Q5-1-8	防火対策
Q5-1-9	街灯など、夜間歩行対策
Q5-1-15	こども園サービス
Q5-1-17	診療所の医療サービス

アンケート#	内容サマリ
Q17	自宅のインターネット利用状況及び今後の利用意向
Q20	あなたは町内の世界遺産のあるエリアを過去3年間を過去3年間の内に何回訪れたことがありますか
Q21-1	問20で「1 ない」と回答した方。このエリアを訪れたことがない理由は
Q24	観光客の増加が見込まれる。宿泊環境の充実に関してどう考えるか。

# 特徴量1次選択 数量化Ⅱ類/相関分析

- ・「住み心地」回答への寄与率が高い、その他アンケート項目回答(特徴量)の寄与率(カテゴリスコア)を算出 [数量化Ⅱ類]  
→寄与率が低い特徴量を削減
- ・特徴量の説明力を上げ、予測精度を高めるため、マルチコ(多重共線性)の変量を集約 [相関分析/クラメール連関係数]  
→クラメール連関係数が高い特徴量を集約

## 相関分析(クラメール連関係数)結果から特徴量を選択

- ◆マルチコの排除のため、アンケート項目間の相関性が高いものを排除する。
- ◆アンケート全項目同士のクロスマトリクスから、項目同士の**クラメール連関係数**を算出
- ◆**クラメール連関係数が0.5以上のもの**をマルチコの変量と認識する
- ◆集約した変量は9変量(他アンケート項目と相関が強い項目を削除、代替変数があるため)

アンケート#	内容サマリ
Q5-1-4	水道・下水道の整備
Q5-1-7	がけ崩れや危険箇所対策
Q5-1-19	小学校・中学校の教育内容
Q5-1-21	生涯スポーツ
F 職業複数	F自営業-農業(複数回答型)

アンケート#	内容サマリ
Q5-1-25	歴史・文化や自然景観など、町の資源活用
Q16-1	xxxx町域の一部は国の重要文化的景観に選定。地域の景観を守るためにどんな協力ができるか？
Q23-1	問22で訪町する方から協力金、税金を設定する場合に、1名あたり徴収する妥当な金額
Q27	あなたは配食サービスを利用していますか

# 特徴量2次選択 MLモデル実行と特徴量の重要度

- ・「住み心地」に対する寄与率が高いアンケート項目[数量化Ⅱ類/クラメール連関係数]と、自由記述回答のうち重要な特徴語[テキストマニング]をマージし、機械学習モデルの入力とする
- ・「住み心地」回答の予測を行う機械学習モデルを構築、学習・予測・特徴量の重要度評価を行う。[特徴量の第2次選択]
- ・選択した特徴量を「主成分分析」へ入力する

## 機械学習モデル実行

- ・アンケート項目のうち「住み心地」に対する寄与率が高い項目[数量化Ⅱ類/クラメール連関係数]と、自由記述回答のうち重要な特徴語[テキストマニング]を選択してマージした結果を特徴量として、「住み心地」を予測する機械学習モデルを構築する
- ・構築した機械学習モデルでの予測結果、重要度が高い特徴量を選択して、「主成分分析」への入力とする [特徴量の第2次選択]

### [実行プロセス]

#### (1) pycaretの実行:

- ・説明変数からQ1(住み心地)を分類する上で最も精度が高いモデルをPyCaretで自動選択。

#### (2) 最も精度が高いモデルで予測を実行した結果、特徴量の重要度評価を実施(feature\_importance、SHAP)

#### (3) 特徴量の重要度ランキング(50程度)から、特徴量を2次選択。

### PyCaret:

・Pythonのオープンソースのローコード機械学習ライブラリ。仮説から考察までのサイクルタイムを短縮することを目的としている。

・主要な学習モデルの精度を比較、精度が高い順に並べてくれる。  
ランキングはpycaretが自動で出力。  
accuracyやf値、適合率、rocなどの指標の%が大きい順に自動でソート

# 特徴量2次選択 MLモデル実行と特徴量の重要度

## 機械学習モデルでの予測

(1) Pycaretでの精度評価:

- ・ランダムフォレストがベストモデル
- ・accuracy(正解率)、適合率、再現率、f値、AUC共に8割を超える精度を確認。

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
<b>rf</b>	Random Forest Classifier	0.8018	0.8628	0.9239	0.8056	0.8602	0.5247	0.5419	0.557
<b>ada</b>	Ada Boost Classifier	0.7961	0.8421	0.8814	0.8227	0.8502	0.5311	0.5377	0.224
<b>et</b>	Extra Trees Classifier	0.7952	0.8578	0.9027	0.8098	0.8532	0.5173	0.5281	0.562
<b>lightgbm</b>	Light Gradient Boosting Machine	0.7944	0.8544	0.8852	0.8184	0.8501	0.5239	0.5299	0.215
<b>lr</b>	Logistic Regression	0.7903	0.8290	0.8665	0.8247	0.8446	0.5220	0.5254	0.465
<b>gbc</b>	Gradient Boosting Classifier	0.7895	0.8451	0.8815	0.8147	0.8465	0.5128	0.5182	0.632
<b>ridge</b>	Ridge Classifier	0.7771	0.0000	0.8540	0.8164	0.8341	0.4938	0.4972	0.033
<b>lda</b>	Linear Discriminant Analysis	0.7688	0.8067	0.8465	0.8116	0.8277	0.4754	0.4794	0.143
<b>svm</b>	SVM - Linear Kernel	0.7624	0.0000	0.8364	0.8198	0.8206	0.4587	0.4782	0.064
<b>knn</b>	K Neighbors Classifier	0.7393	0.7669	0.8404	0.7822	0.8097	0.3972	0.4012	0.240
<b>dt</b>	Decision Tree Classifier	0.7196	0.6930	0.7769	0.7942	0.7850	0.3819	0.3831	0.048
<b>dummy</b>	Dummy Classifier	0.6595	0.5000	1.0000	0.6595	0.7948	0.0000	0.0000	0.022
<b>nb</b>	Naive Bayes	0.5535	0.7574	0.3916	0.8530	0.5346	0.2069	0.2678	0.030
<b>qda</b>	Quadratic Discriminant Analysis	0.3454	0.4914	0.0337	0.5848	0.0630	-0.0119	-0.0386	0.131

(2) ランダムフォレスト実行

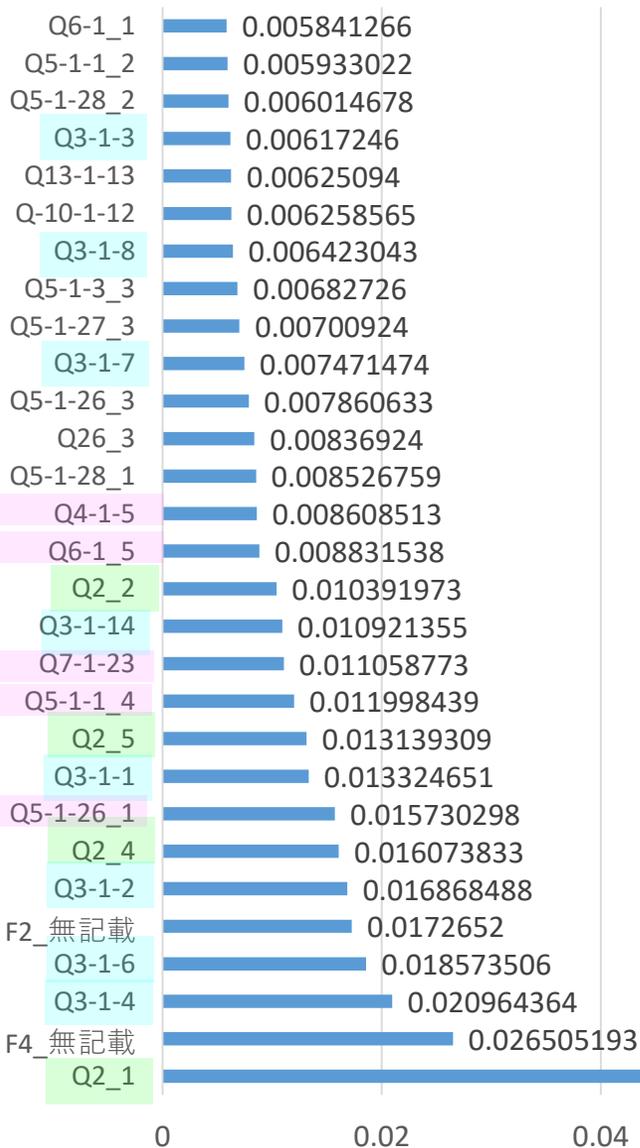
```
x_train, x_test, y_train, y_test =  
train_test_split(x, y, test_size=0.3, shuffle = True)  
clf = RandomForestClassifier  
(n_estimators = 100 ,max_depth=10)
```

# 特徴量2次選択 MLモデル実行と特徴量の重要度

特徴量の重要度 feature\_importance

ランダムフォレスト実行での特徴量重要度

[戻る](#)



importances

記号	アンケート内容	選択回答
Q2_1	住み続けたさ	#1住み続けたい
F4	居住地区	無記載
Q3-1-4	住み続けたい理由	#4土地と家あり
Q3-1-6	住み続けたい理由	#6自然環境良い
F2	年代	無記載
Q3-1-2	住み続けたい理由	#2ずっと住んでいる
Q2_4	住み続けたさ	#4どちらとも言えない
Q5-1-26_1	日常買い物利便性	#1満足していない
Q3-1-1	住み続けたい理由	#1町に愛着あり
Q2_5	住み続けたさ	#5未回答
Q5-1-1_4	本土との海上交通	#4未回答
Q7-1-1_23	町の課題	#23未回答
Q3-1-14	住み続けたい理由	#14近所付き合いに満足
Q2_2	住み続けたさ	#2将来他に移りたい
Q6-1_5	定住人口問題	#5未回答
Q4-1_5	将来他に移りたい	#5買い物通院に不便
Q5-1-28_1	働く場の確保	#1満足していない
Q26-3	1人暮らしか	#3未回答

「日常の交通の利便性」の重要度が高い

# 5. 主成分分析 主成分内容 (pc1/pc2)

## 導出した主成分 (pc1, pc2) の内容

どのアンケート項目の「主成分負荷量」が、pc1, pc2の値を強める貢献度が高いかによりpc1およびpc2を意味付けた

		pc1	pc2
合成変量の意味		<b>未来改善(不)志向度</b>	<b>家族居住/居住継続意向</b>
合成変量の傾向	主成分負荷量(正)	<ul style="list-style-type: none"> <li>・現状の<b>生活環境不満</b></li> <li>・まちづくり<b>参加意欲小</b> ・やや<b>消極的</b></li> <li>・人口・就業・文化遺産活用に一部課題提起</li> <li>・地域特性 (Fエリア、NKエリア、Mエリア)</li> <li>・<b>年代特性</b> (40~50代、10代~30代)</li> </ul>	<ul style="list-style-type: none"> <li>・2人以上で居住</li> <li>・住み続けたい意向が強い</li> <li>・地域特性 (Zエリア、Yエリア)</li> <li>・年代特性 (40~50歳代)</li> </ul>
	主成分負荷量(負)	<ul style="list-style-type: none"> <li>・現状の<b>生活環境満足</b></li> <li>・まちづくり<b>参加意欲大</b> ・<b>積極性</b></li> <li>・生活インフラ/環境、コミュニティ改善意向</li> <li>・未来の環境に対するの見解意識 (コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題)</li> <li>・地域特性 (Hエリア、Yエリア、Zエリア、Kエリア、Oエリア、Nエリア)</li> <li>・<b>年代特性</b> (60代~90代)</li> </ul>	<ul style="list-style-type: none"> <li>・1人暮らし</li> <li>・居住環境への課題提起多い (インフラ、福祉)</li> <li>・地域特性 (Mエリア、Oエリア)</li> </ul>

# 主成分分析(5) 主成分の内容(pca1)

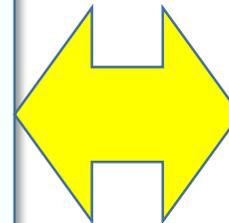
## 主成分1(pc1): 未来改善(不)志向度

### 【合成変量(正/大)の傾向】

- ・現状の生活環境不満、まちづくり参加意欲小
- ・消極的
- ・人口・就業・文化遺産活用に一部課題提起

#### <詳細>

- ・他に移りたい志向が強い40～50代、10～30代
- ・現状生活環境への不満・要望高い
- ・若い世代、一人暮らし
- ・住民まちづくりは、やや消極的
- ・居住地区:Fエリア、NKエリア、Mエリア



### 【合成変量(負/小)の傾向】

- ・現状の生活環境満足、まちづくり参加意欲大
- ・積極的
- ・生活インフラ/環境、コミュニティ改善意向
- ・未来の環境に対しての見解意識  
(コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題)

#### <詳細>

- ・住み続けたい志向が強い60代以降
- ・現状生活環境に満足、提案型、家族/親類と同居
- ・日々の生活基盤を大切にしたい
- ・高齢化や、高齢に伴う障害、障がい者福祉の懸念
- ・積極的な自治体活動への参画を志向
- ・居住地区:  
Hエリア、Yエリア、Zエリア、Kエリア、Oエリア、Nエリア
- ・対人関係は良好

# 今後の課題と対応方針

課題1：主成分分析における第2主成分までの累積寄与率が11.6%  
「潜在的な新住民価値」となる導出主成分の寄与率が低い

課題(1)主成分の寄与率向上のための改善案

課題	プロセス	対策方式	状況(2022アンケート分析)
「住み易い」判別(正例)に寄与する特徴量の見極めと抜粋	数量化Ⅱ類	カテゴリ係数の値とカテゴリスコアとの関係から該当特徴量を選択する	カテゴリ係数値を評価中
多重共線性排除のための特徴量の集約の最適化	相関分析	クラメール連関係数により削除する特徴量として目的変数との相関が強い方を選ぶ	相関分析実施中
特徴語選択方法の最適化	テキストマイニング	TF-IDFの高い特徴語と「住み易い」回答との相関を評価する(機械学習での重要度評価)	文字回答からTF-IDF評価中
機械学習の精度向上(特徴量の選択精度向上)	機械学習	交差検証法の活用	計画中
自由記述回答の深耕	テキストマイニング	感情分析辞書の活用	研究計画中

# 今後の課題と対応方針

- ・課題2：主成分分析で発掘した潜在的テーマが「新住民価値」として重要度が高いかどうかを追加アンケートで検証する(機械学習の適用)
- ・課題3：発掘した潜在的テーマに関して、住民意識を高めるための居住環境条件の改善提案

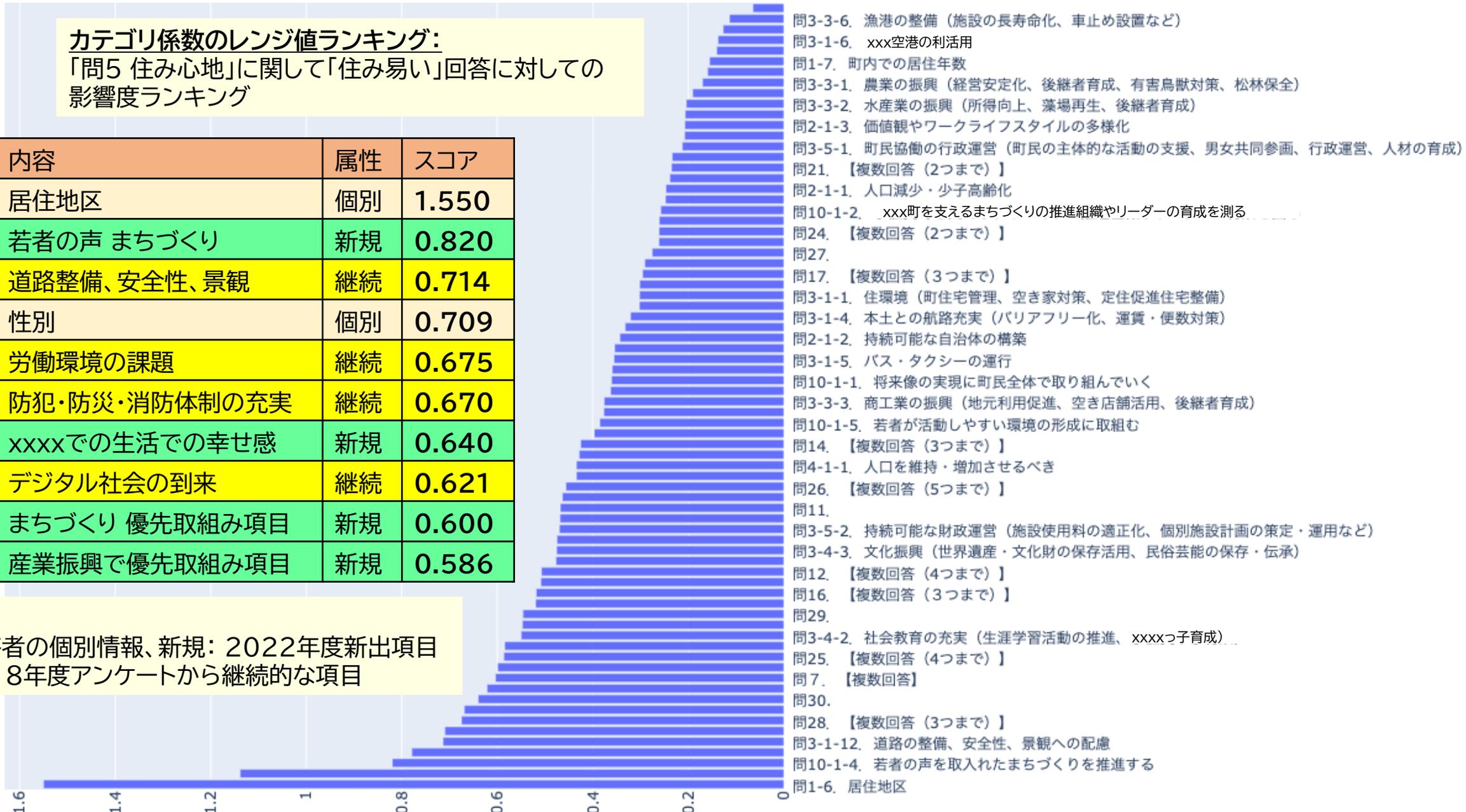
目的	研究プロセス		概要
(1)新住民価値の明確化	(a)潜在的テーマの発掘	(a-1)特徴量選択	アンケート結果から潜在的な住民価値発掘の「ための主成分分析の精度向上のための特徴量選択。 数量化Ⅱ類、相関分析、テキストマイニング、機械学習
		(a-2)新住民価値の概念発掘	主成分の意味付け(主成分分析)とアンケート追加と検証
	(b)潜在的テーマの評価 課題2		追加アンケート結果の分析【機械学習】 (追加アンケート結果を(説明変数(X))とした住み易い(目的変数(y))に対する重要度)
(2)新住民価値創出のための施策提案	(c)潜在的テーマに対して住民意識を高める居住環境条件の明確化 課題3		追加アンケート結果の住民意識を高めるような、居住環境条件を導出する (追加アンケート結果を(目的変数(y))とし他アンケート回答を(説明変数(X))とした場合の(X)の重要度評価)

# 【参考】7. 2022年度アンケート結果の分析状況 数量化Ⅱ類

カテゴリ係数のレンジ値ランキング:  
「問5 住み心地」に関して「住み易い」回答に対しての  
影響度ランキング

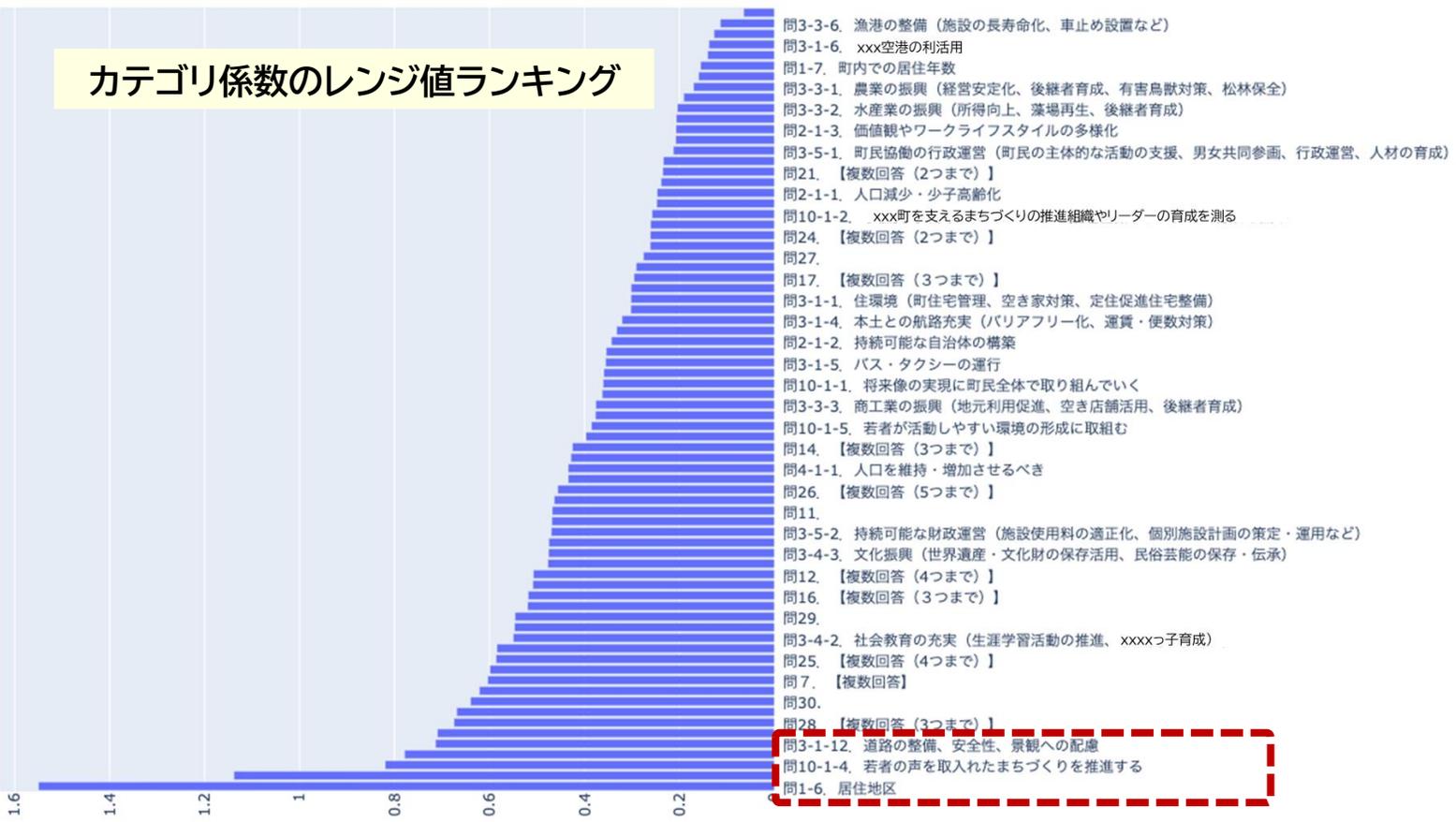
問 項 番	内容	属性	スコア
1-6	居住地区	個別	1.550
10-1-4	若者の声 まちづくり	新規	0.820
3-1-12	道路整備、安全性、景観	継続	0.714
1-1	性別	個別	0.709
28	労働環境の課題	継続	0.675
3-1-2	防犯・防災・消防体制の充実	継続	0.670
30	xxxxでの生活での幸せ感	新規	0.640
2-1-6	デジタル社会の到来	継続	0.621
9	まちづくり 優先取組み項目	新規	0.600
25	産業振興で優先取組み項目	新規	0.586

(※)属性  
個別：回答者の個別情報、新規：2022年度新出項目  
継続：2018年度アンケートから継続的な項目



# 【参考】7. 2022年度アンケート結果の分析状況 数量化Ⅱ類

カテゴリ係数のレンジ値ランキング



問 項 番	内 容	属 性	ス コ ア
1-6	居住地区	個別	1.550
10-1-4	若者の声 まちづくり	新規	0.820
3-1-12	道路整備、安全性、景観	継続	0.714
1-1	性別	個別	0.709
28	労働環境の課題	継続	0.675
3-1-2	防犯・防災・消防体制充実	継続	0.670
30	xxxx生活での幸せ感	新規	0.640
2-1-6	デジタル社会の到来	継続	0.621
9	まちづくり 優先取組項目	新規	0.600
25	産業振興で優先取組項目	新規	0.586
1-5	所帯人数	個別	0.584
3-4-2	社会教育の充実	継続	0.550
3-4-1	学校教育の充実	継続	0.547
29	xxxxの魅力	新規	0.546
3-1-8	ごみ・し尿対策	継続	0.520
16	子育て環境	継続	0.519
1-4	職業	個別	0.509
12	デジタル環境整備の項目	継続	0.507

- ◆「住み心地」は「居住地区」の依存度が最高。(今回はスコア2位)
- ◆事実上最も影響度が高い特徴量(項目/カテゴリ)として「若者の声を取り入れた街づくり」であり、若者/ベテランの乖離と世代交代要望の認識の高さがうかがえる
- ◆生活環境に関する項目でトップの重要度が「道路の整備、安全性、景観への配慮」  
 前回主成分分析でも負荷量が高い項目