2025年参議院選挙の候補者別得票数(選挙区)データを活用した

選挙戦略考案と候補者策定プロジェクト

目次

1. 背景と目的

目的と背景

【目的】

2025年参議院選挙の候補者別得票数(選挙区)のデータを学習し、参議院選挙(都道府県別選挙区)での

候補者の当選確率を予測する

【課題】

年齢、性別、元職・現職・新人、党派、職業といった属性が当落にどの程度影響しているのかを

定量的に把握できておらず、効果的な選挙戦略の策定が難しい。

そのため、どのような候補者が当選しやすいのかを明らかにし、選挙戦略や候補者選定に活用する。

【ビジネスへの応用案】

小売・EC：客の購買履歴・アクセスチャネル・季節要因が 購入確率にどう影響するかを定量分析

→ 在庫管理の精度向上、プロモーション最適化

2. 開発の動機

開発の動機

選挙の当落予測は、実際のデータを扱うことができ、予測結果が現実の社会現象と結びついていることから、

データ分析が現実世界でどのように活用できるかを示す例になると考え、題材を選びました。

3. サイトマップ

サイトマップ

2025_upperhouse_election_constituency_system_predictor

```
│
├── images
│   └── （画像データ格納）
│
├── Python Scripts
│   ├── data_analysis_age.py                （年齢分析）
│   ├── data_analysis_basic_statistics.py （基礎統計分析）
│   ├── data_analysis_gender.py             （性別分析）
│   ├── data_analysis_job.py                （職業別分析）
│   ├── data_analysis_party.py              （政党別分析）
│   ├── data_analysis_seats.py              （議席数分析）
│   ├── data_analysis_status.py             （元現新分析）
│   ├── data_cleaning.py                    （データ前処理）
│   ├── data_exploration.py                 （探索的データ分析）
│   └── data_machine_learning.py            （機械学習モデル）
│
├── README.md(プロジェクトの概要を記載したファイル)
│
└── Data
    ├── 2025_upperhouse_election_constituency_system_cleaning.xlsx  （前処理済データ）
    └── 2025_upperhouse_election_constituency_system.xlsx           （元データ）
```

4. 衆議院選挙・参議院選挙の選挙制度概説

衆議院選挙・参議院選挙の選挙制度

	衆議院	参議院
定数	465人	248人
任期	4年（ただし解散あり）	6年（解散なし、3年ごとに半数改選） →今年は124＋1(東京で欠員)=125
選挙方法	小選挙区比例代表並立制	都道府県単位の選挙区制 →今年は75名の改選 全国単位の比例代表制 →今年は50名の改選
選挙区	全国289の小選挙区で各区1人を選ぶ	都道府県ごとの45選挙区 （鳥取と島根県、高知と徳島県=合区で一選挙区）
比例代表	全国11ブロック 「拘束名簿方式」	全国1ブロックの 「非拘束名簿方式」
比例の投票方法	政党名を記入	政党名か候補者名を記入
重複立候補	できる	できない

衆議院選挙・参議院選挙の選挙制度

※選挙区候補者の当選確率を対象としており、比例代表の候補者はデータに含まれていません。そのため、党派ごとの比例代表の当選傾向は反映されておらず、分析結果は各都道府県の選挙区での当選傾向に限定されます。

「出典：NHK『参議院選挙 2025』より」

5. 仮説設定・調査分析計画

仮説設定・調査分析計画

【収集データ】

2025年参議院選挙データ(候補者別・得票数)

【仮説】

①年齢(職業経験の長さが政治経験に反映され、年齢が高い人ほど当選確率が高い)

②性別(近年のジェンダーギャップ指数からも、男女で違いがある)※公開情報を元に手入力

③所属政党(政党のブランドや支持基盤、保守やリベラルといった政治的立場が当落に影響する)

④元現新(政治経験が豊富な人は当選率が高い)

⑤職業(選挙費用をかけられれば当選しやすくなるので、資本家や経営者の当選確率は高い)

⑥選挙区(選挙区の人口や議席数によって当落確率が変わる) ※公開情報を元に手入力

仮説設定・調査分析計画

【分析手法】

（１）ロジスティック回帰：いくつかの要因（説明変数）から「2値の結果（目的変数）」が起こる確率を

説明・予測することができる統計手法で、多変量解析の手法の１つ

<選定理由>:どの特徴量が結果にどのように影響するのか解釈がしやすく、基準モデルとして採用

（２）ランダムフォレスト：複数の「決定木」を作成し、それらの多数決や平均で予測。特徴量の重要度を取得できる。

<選定理由>:複雑なデータの関係も捉えられ、精度が高く、特徴量の重要度もわかる

<class_weight='balance'>:

当選クラスと落選クラスでのクラス不均衡に対応し、当選者の予測精度を向上させるため適用

<ハイパーパラメータ　チューニング>:

Grid Search CV を使い、交差検証で予測精度が最も高くなるハイパーパラメータを自動で選択

【成果物】

・予測モデルを作成し、特徴量重要度を可視化。

さらにユーザー入力で当落確率を返す簡易アプリを実装。

6. モデルの評価指標

モデルの評価指標

(1)混合行列:正解・不正解の内訳を表にしたもの。

実際＼予測	落選と予測	当選と予測
実際に落選	True Negative(TN)	False Positive (FP)
実際に当選	False Negative (FN)	True Positive(TP)

モデルの評価指標

(2)精度指標(適合率、再現率、F1スコア、正解率)

指標	数式	意味
Precision（適合率）	$\text{Precision} = \frac{TP}{TP + FP}$	「当選と予測した人のうち、本当に当選だった割合」
Recall（再現率）	$\text{Recall} = \frac{TP}{TP + FN}$	「本当に当選した人のうち、予測でも当選とされた割合」
F1スコア	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	Precision と Recall のバランスをとった指標
Accuracy（正解率）	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$	全体の予測で正しかった割合

7. データの収集・整理

データの収集・整理

【データの整理】

・欠損値処理

・異常値のチェック(全角、半角スペースの削除)

・不要な列の削除

・性別/選挙区別の議席数列の追加、職業の分類

8. データ分析

データに基づく分析

- 平均値、中央値、最大、最小などの統計量を把握

- 相関分析とデータの可視化：変数間の関係性確認

- モデル学習：ロジスティック回帰/ランダムフォレストを使って当落予測

- モデル評価：混合行列、正解率、適合率、再現率、F値

データに基づく分析(基本統計量)

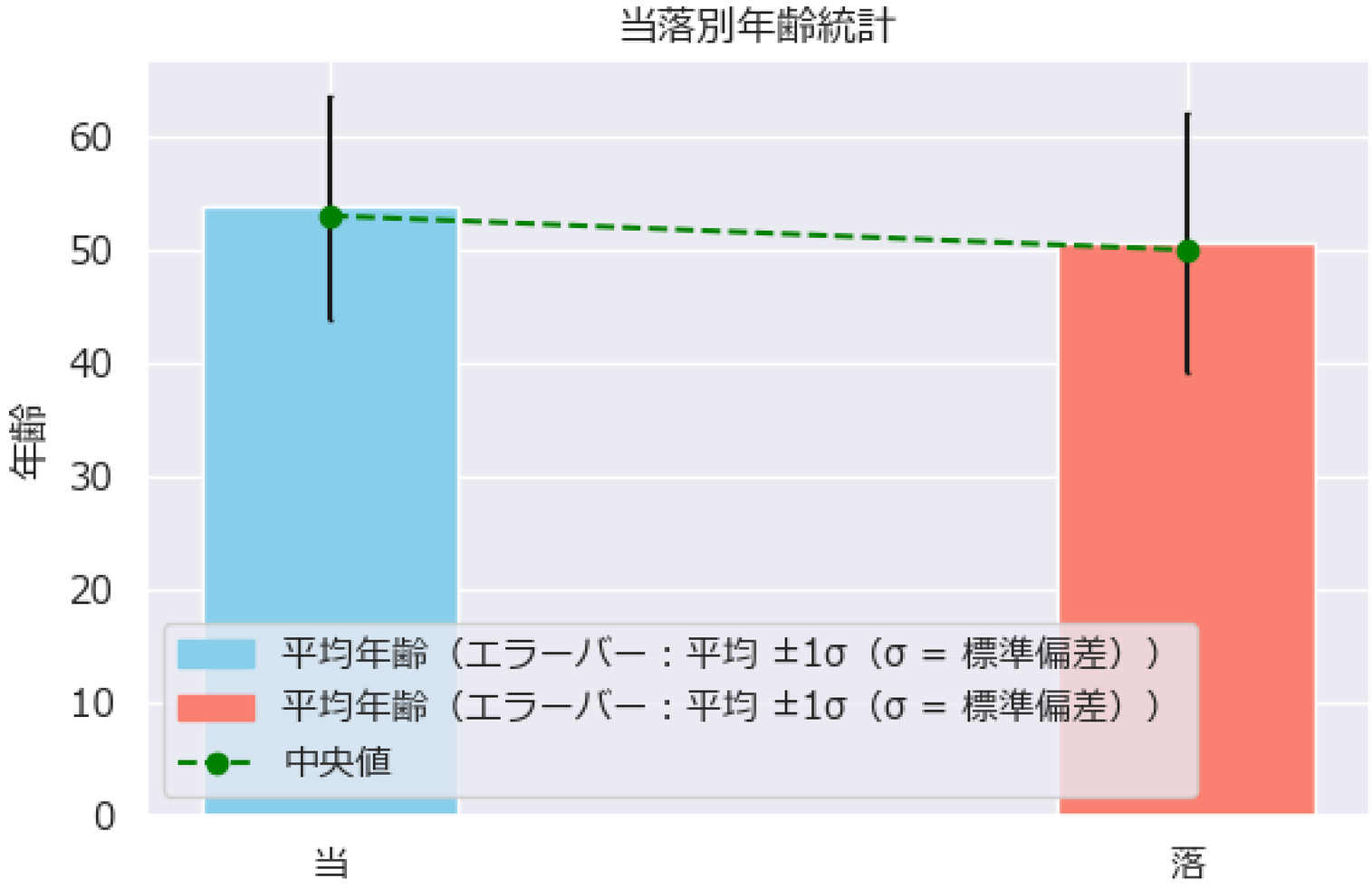
	当落	候補者氏名	年齢	性別	党派	元現新	議席数	職業(分類)
count	350	350	350.0	350.0	350	350	350.0	350
unique	2	350	–	–	27	3	–	22
top	落	眞喜志　　雄　一	–	–	自由民主党	新	–	政治家・政党関係
freq	275	1	–	–	48	287	–	142
mean	–	–	51.27	0.32	–	–	2.15	–
std	–	–	11.27	0.47	–	–	1.54	–
min	–	–	30.0	0.0	–	–	1.0	–
25%	–	–	43.0	0.0	–	–	1.0	–
50%	–	–	51.0	0.0	–	–	1.0	–
75%	–	–	59.75	1.0	–	–	3.0	–
max	–	–	78.0	1.0	–	–	7.0	–

・立候補が最も多い政党は自由民主党で48人である

・元現新の中だと最多は新人で287人である

・職業では政治家、政党関係が最多で、142人の候補者が存在する

データに基づく分析(当落と年齢)



・当選者（平均53.72歳、標準偏差9.84）

→平均 ±1σ = 53.72 ± 9.84 → 43.88歳～

63.56歳

・落選者（平均50.60歳、標準偏差11.56）

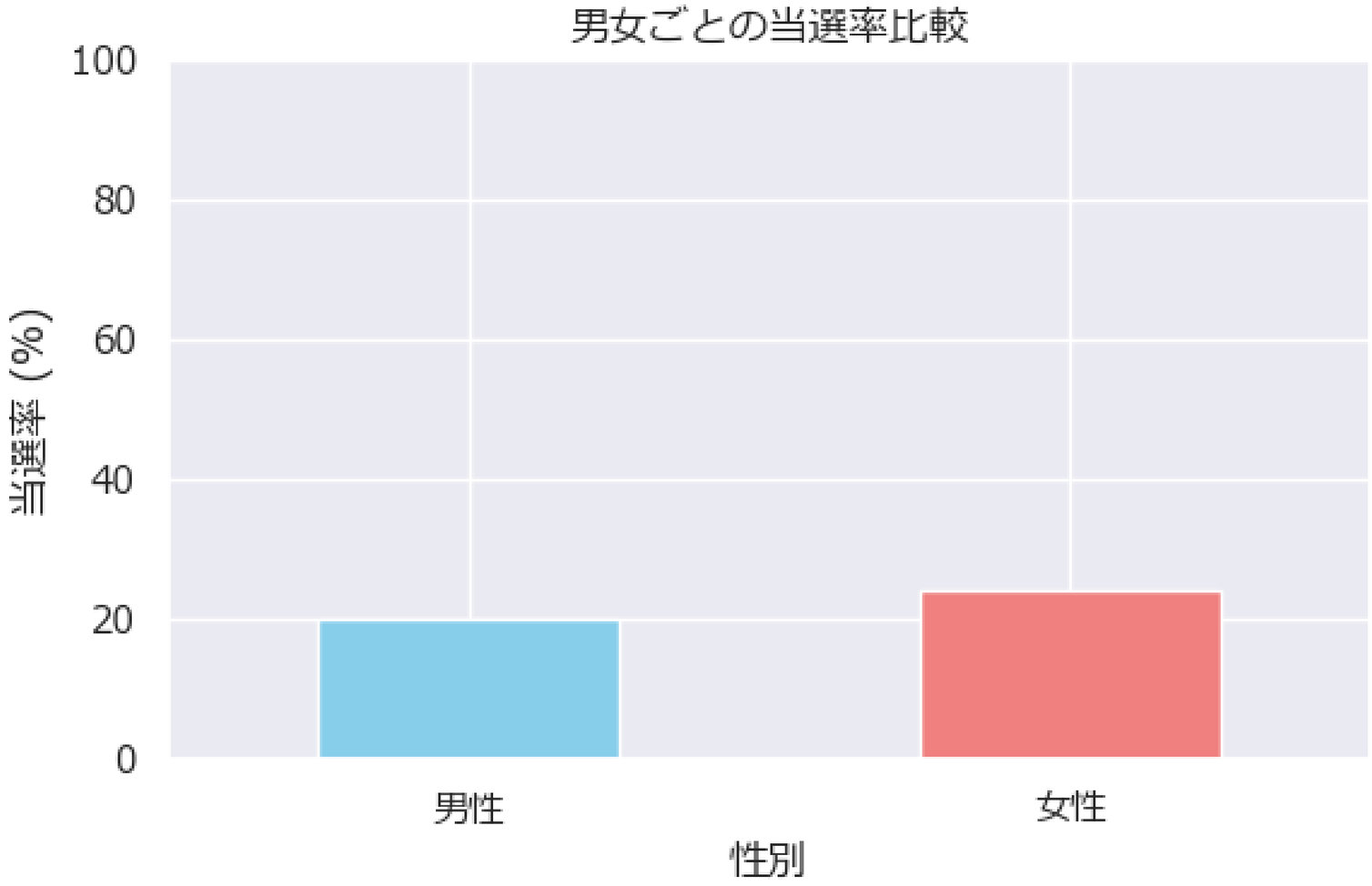
→平均 ±1σ = 50.60 ± 11.56 → 39.04歳～

62.16歳

※正規分布を仮定すると、

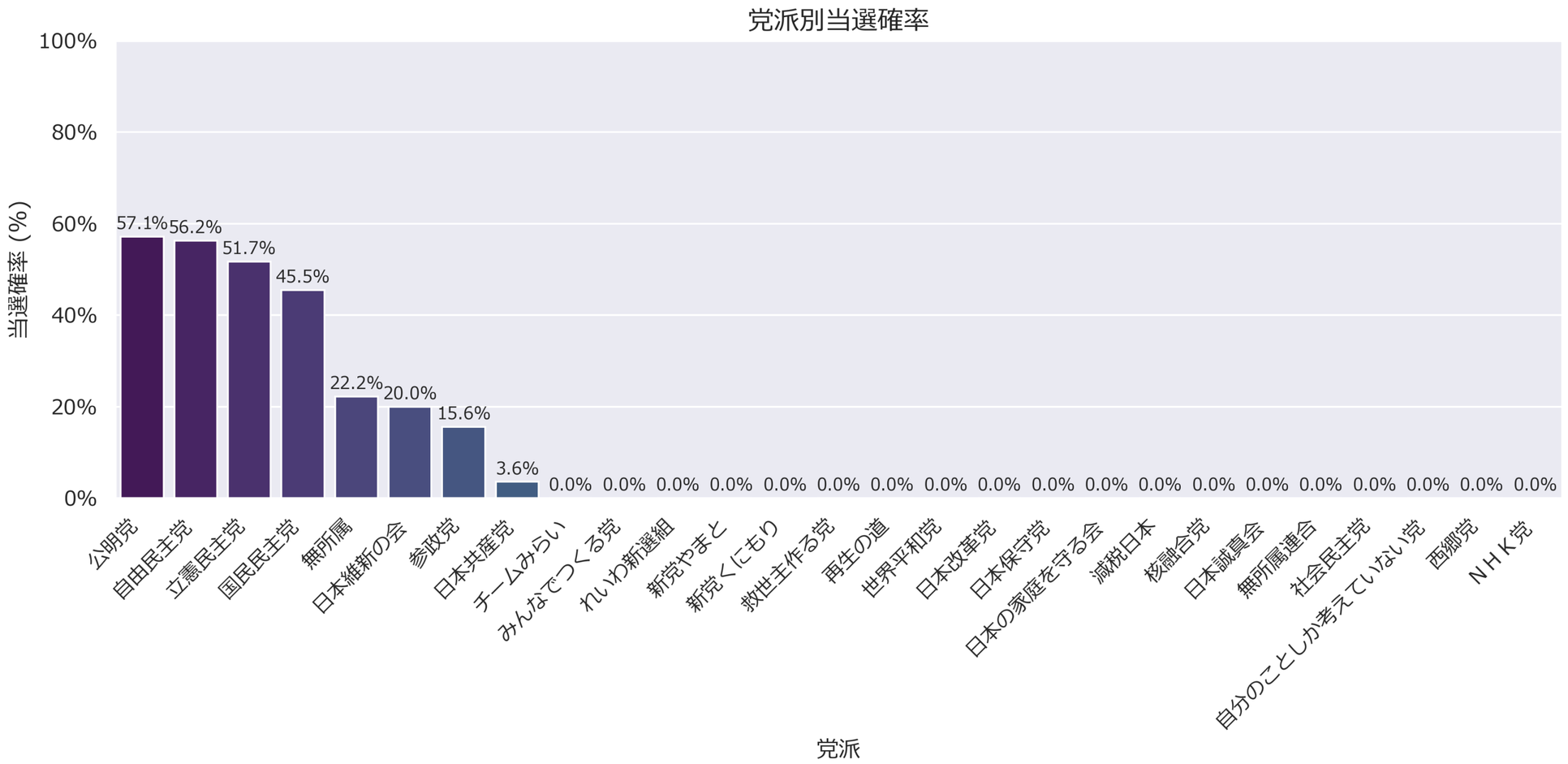
この範囲に 約68%の当選者 が存在する

データに基づく分析（当落と性別）



・男女で当選確率で大きな違いはない

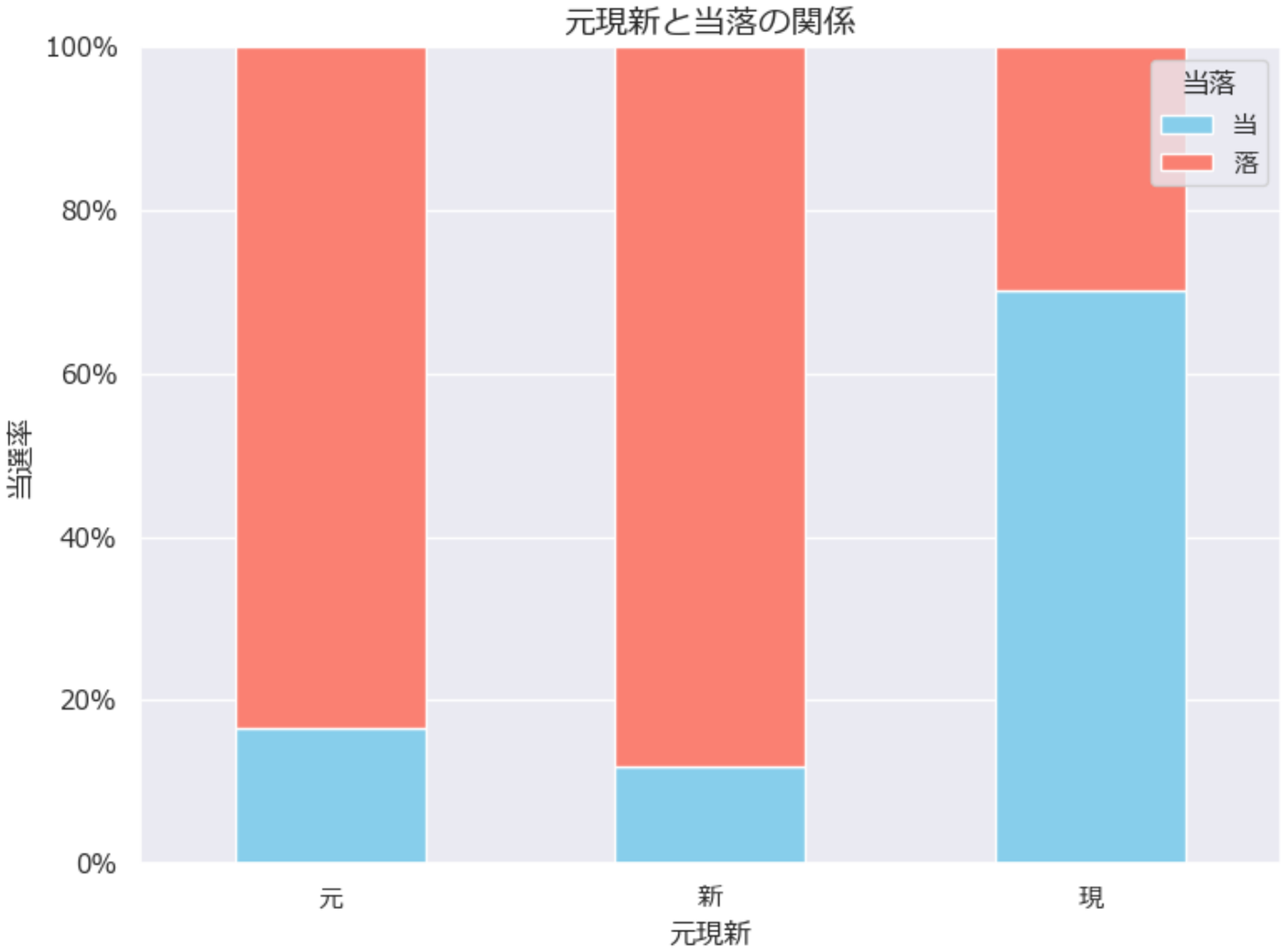
データに基づく分析（当落と党派）



・公明党、自由民主党といった与党の当選確率が高いが、立憲民主党や国民民主党の当選確率も高い

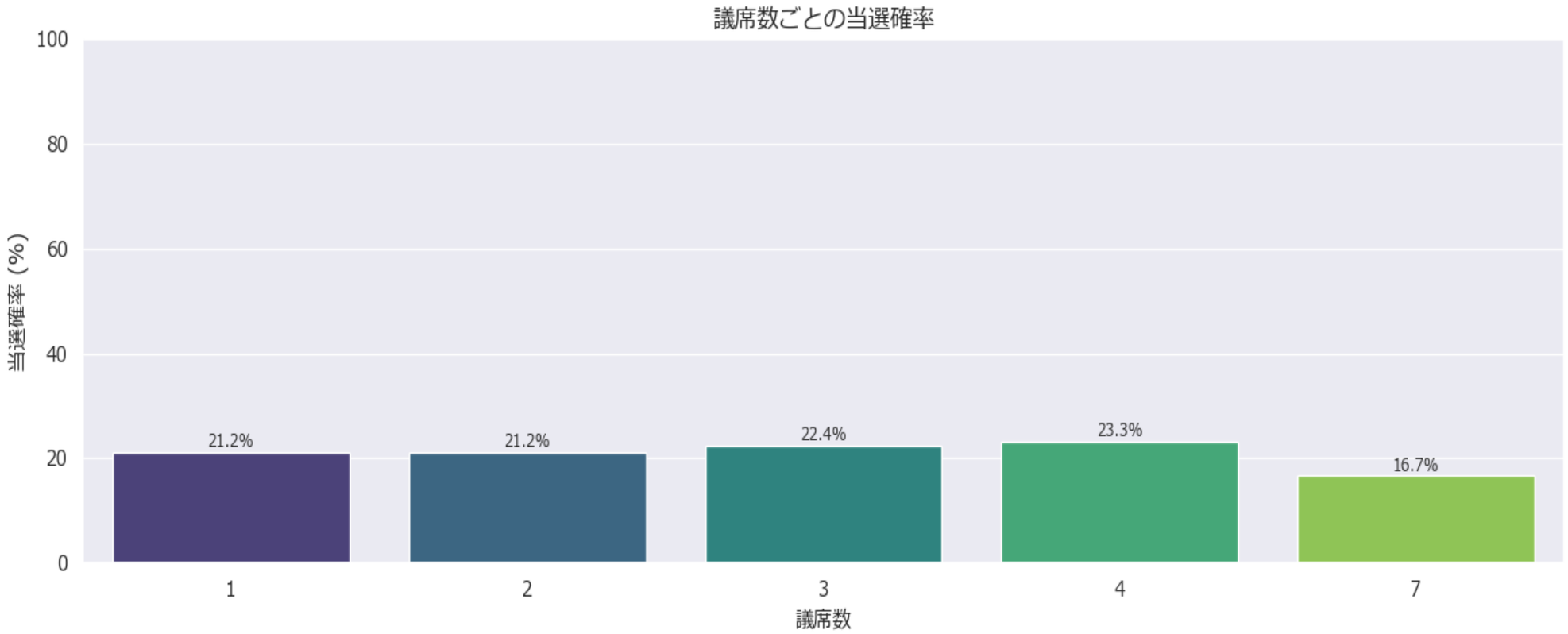
→与党が明らかに有利とまでは言えない

データに基づく分析（当落と元職・現職・新人）



・現職の当選率が高い。新人と元職は同じくらいの当選確率である。

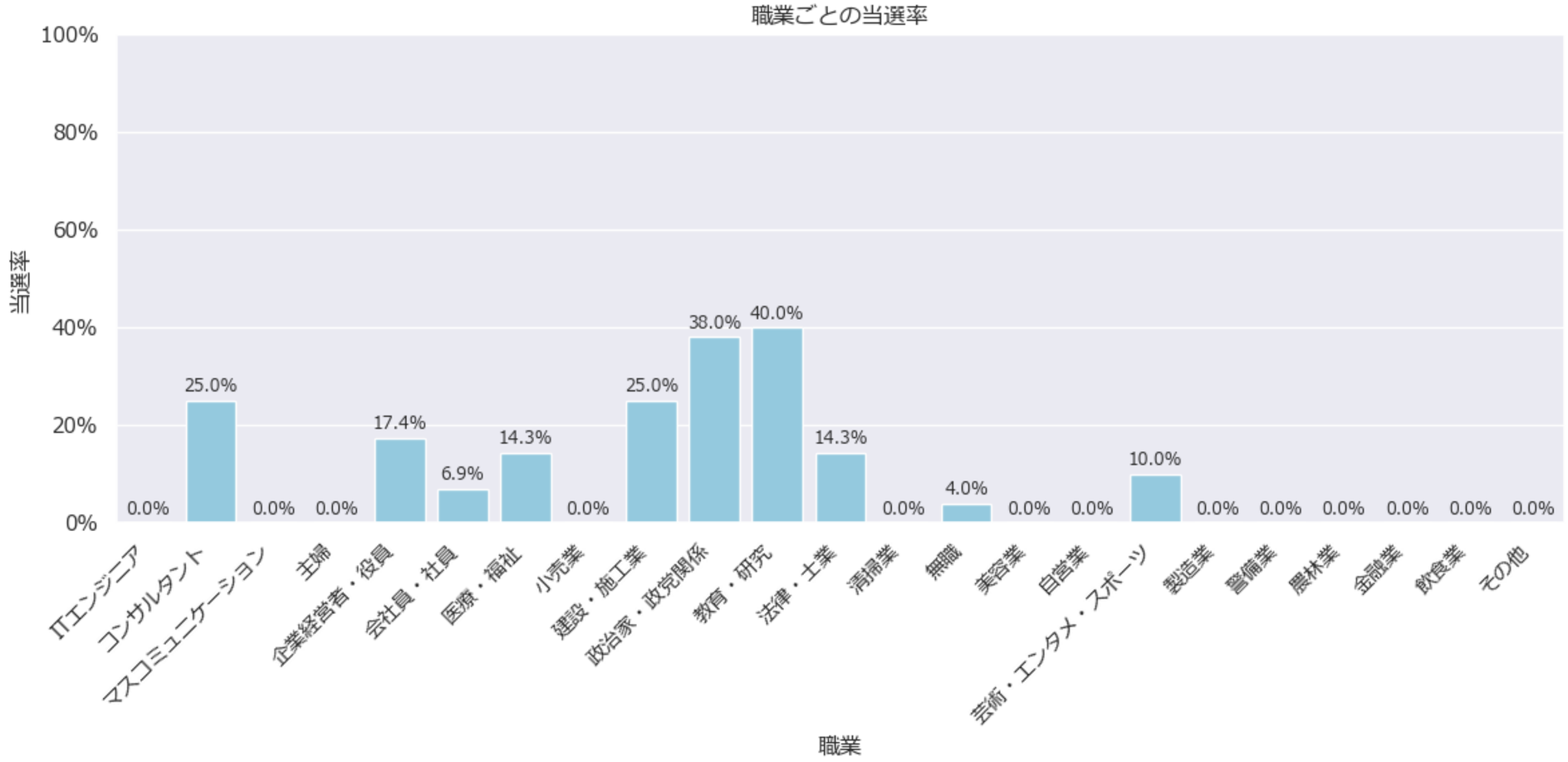
データに基づく分析(当落と議席数)



・1−4議席の選挙区では当選確率は大きくは変わらないが、7議席の東京選挙区では

当選確率がほかの選挙区と比べて低い。つまり若干ではあるものの競争率が高い。

データに基づく分析(当落と職業)



→ (1)教育・研究 (2)政治家・政党関係 (3)コンサルタントと建設・施工業が同率で3位。政治経験よりも教育分野での経験が当落に影響する可能性が示唆されている。

9.モデル評価の解釈

モデル評価の解釈

【評価指標(1):混合行列】

①ロジスティック回帰

実際＼予測	落選と予測	当選と予測
実際に落選	52(TN :True Negative)	3(FP :False Positive)
実際に当選	8(FN :False Negative)	7(TP :True Positive)

・落選と予測した人60人のうち実際に落選した人は52人であり、予測精度は0.866

・当選と予測した人10人のうち実際に当選した人は7人であり、予測精度は0.70

→落選予測の精度は高いが、落選予測と比べて当選予測の精度は低い

モデル評価の解釈

【評価指標(1):混合行列】

②ランダムフォレスト

実際＼予測	落選と予測	当選と予測
実際に落選	54(TN :True Negative)	1(FP :False Positive)
実際に当選	7(FN :False Negative)	8(TP :True Positive)

・落選と予測した人61人のうち実際に落選した人は54人であり、予測精度は0.885

・当選と予測した人9人のうち実際に当選した人は8人であり、予測精度は0.888

→落選と当選の予測どちらも精度が高い

モデル評価の解釈

【評価指標(2):適合率、再現率、F1スコア、正解率】

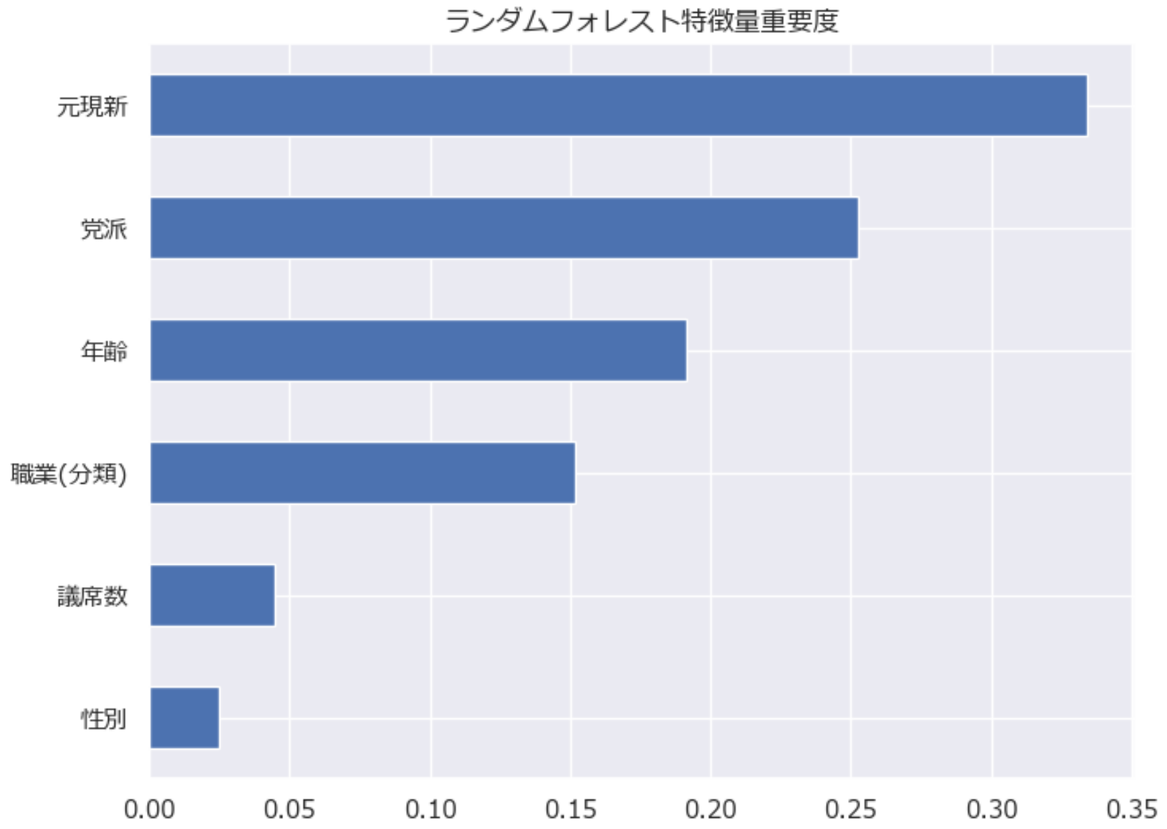
実際・予測	落選と予測	当選と予測
実際に落選	True Negative(TN)	False Positive (FP)
実際に当選	False Negative (FN)	True Positive(TP)

指標	ロジスティック回帰	ランダムフォレスト (グリッドサーチ)	数式	意味
Precision（適合率）	0.700	0.889	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$	当選と予測した人のうち、 実際に当選だった割合
Recall（再現率）	0.467	0.533	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	実際に当選した人のうち、 予測でも当選とされた割合
F1スコア	0.560	0.667	$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Precision と Recall の バランスをとった指標
Accuracy（正解率）	0.843	0.886	$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$	全体の予測で正しかった 割合

落選した人の予測と実際の結果も含めて全体で評価したいため正解率を参照し、予測モデルにはランダムフォレストを採用する

モデル評価の解釈

【評価指標(3):ランダムフォレスト特徴量重要度】



重要度は元現新>党派>年齢>職業>議席数>性別の順番

10. 分析結果の考察と提案

分析結果の考察

①年齢(仮説：職業経験の長さが政治経験に反映され、年齢が高い人ほど当選確率が高いのではないか)

→当選者の平均年齢は53.72歳。標準偏差は9.84

(平均 ±1σ = 53.72 ± 9.84 → 43.88歳～63.56歳。正規分布を仮定すると、この範囲に 約68%の当選者 が集中している)

→ 職業経験の長さが政治経験に反映され、年齢が高い人ほど当選しやすい。

②性別(仮説：近年のジェンダーギャップ指数からも、男女で違いがある)

→男性と女性との間で大きな違いはない。

③所属政党(仮説：政党のブランドや支持基盤、保守やリベラルといった政治的立場が当落に影響するのではないか)

→与党(公明・自民)は強いが、野党第一党である立憲民主党も高い当選率を記録している

分析結果の考察

④元現新(仮説：政治経験が豊富な人は当選率が高い)

→現職が最も当選しやすい。元職でも新人とあまりかわらない当選率になる。

⑤職業(仮説：職業によって有権者の印象や支持基盤が変わる)

→ (1)教育、研究(2)政治家、政党関係(3)コンサルタントと建設・施工業が上位3職種で、教育・研究者が最も当選しやすい。

⑥選挙区(仮説：選挙区の人口や議席数によって当選確率が変わる)

→1,2,3,4議席の当選確率は同水準。7議席(東京選挙区)は他の議席と比べて当選確率が低い。

⑦特徴量の重要度

→ランダムフォレストによる当選確率の寄与度を調べてみると、元現新>党派>年齢>職業>議席数>性別の順番。

分析結果の考察

【まとめ】

・現職＞与党・野党第一党に所属＞40～60代＞教育・研究者の属性を持った候補者の当選確率が高い。

提案

候補者属性(優先度降順)	当選しやすい人への提案	当選しにくい人への提案
元現新	現職は実績をアピールし、安定感を訴求	新人・元職は新しいリーダー像 専門性で信頼を補完
所属政党	与党（自民・公明）または野党第一党 （立憲民主）は既存の組織基盤を活用	小規模政党所属は独自性や地域密着を 前面に出して差別化
年齢	40代から60代の候補者は過去の経験を 前面に押し出すことで、有権者にアピールする	若年層は専門性・新しい視点を強調し、 ベテランとの差別化を図る
職業	多様な声を政策に反映させる必要があるため 様々な職業から候補者を募るようにする	多様な声を政策に反映させる必要があるため 様々な職業から候補者を募るようにする
選挙区	1～4議席区で効率的に当選を狙う	議席数が多い区では小さいコストで始められる オンラインでの広報戦略で認知拡大
性別	男女差が小さいため、ジェンダーに 固執せず政策訴求に集中	男女差が小さいため、ジェンダーに 固執せず政策訴求に集中

11. 予測モデルのデモ操作

予測モデルのデモ操作

Pythonスクリプト data_machine_learning.py を実行すると、ユーザー入力から当落予測ができます。

−入力例− 名前: 山田太郎 年齢: 56 性別: 男性 党派: 自由民主党 元現新: 現職 議席数: 3 職業: 政治家・政党関係

−出力例− 山田太郎さんの当選確率は 90.56% → 予測ラベル: 当選（使用モデル: ランダムフォレスト）

```
C:\Users\frontier-Python\Desktop\2025_upperhouse_election_constituency_system_predictor\python Scripts>python data_machine_learning.py

*** ロジスティック回帰 評価 ***
Confusion Matrix:
[[52  3]
 [ 8  7]]
Accuracy : 0.843
Precision: 0.700
Recall   : 0.467
F1-score : 0.560

*** ランダムフォレスト (グリッドサーチ) 評価 ***
Confusion Matrix:
[[54  1]
 [ 7  8]]
Accuracy : 0.886
Precision: 0.889
Recall   : 0.533
F1-score : 0.667

*** 当選確率予測アプリ ***
名前を入力してください: 山田太郎
年齢を入力してください: 56
性別を入力してください (男性=0, 女性=1) : 0
党派を入力してください ['みんなでつくる党', 'れいわ新選組', 'チームみらい', '世界平和党', '公明党', '再生の道', '参政党', '国民民主党', '救世主作る党', '新
党くにもり', '新党やまと', '日本の家庭を守る会', '日本保守党', '日本共産党', '日本改革党', '日本維新の会', '日本誠真会', '核融合党', '減税日本', '無所属', '
無所属連合', '社会民主党', '立憲民主党', '自分のことしか考えていない党', '自由民主党', '西郷党', 'ＮＨＫ党']: 自由民主党
元現新を入力してください ['元', '新', '現']: 現
議席数を入力してください: 3
職業を入力してください ['ITエンジニア', 'その他', 'マスコミュニケーション', '主婦', '企業経営者・役員', '会社員・社員', '医療・福祉', '小売業', '建設・施工
業', '政治家・政党関係', '教育・研究', '法律・士業', '清掃業', '無職', '美容業', '自営業', '芸術・エンタメ・スポーツ', '製造業', '警備業', '農林業', '金融業
', '飲食業']: 政治家・政党関係

山田太郎さんの当選確率は 90.56% → 予測ラベル: 当選
（使用モデル: ランダムフォレスト）
```

12. 今後の展望・追加データの活用

今後の展望・追加データの活用

・若年層を中心に利用ユーザーが広がっているSNSメディアのフォロワー数を参考に当落の相関関係もみてみたい

・地元出身がどのように当選確率に影響を与えるのか定量的に分析してみたい

・全体的に学習データが少ないので、過去の参議院選挙の結果もデータに追加することを検討したい

13. 苦労したところ

苦労したところ

・比例代表で当選した人のデータセットを準備することが難しく、各都道府県ごとの当落傾向しか見られなかったので、参議院選挙全体の当落予測をすることができなかった。

14. 参照

参照

・森巧尚（2019）『Python2年生 データ分析のしくみ』翔泳社

・「NHK.『参議院選挙 2025』. https://www.nhk.or.jp/senkyo/database/sangiin/2025/san-structure/, （参照 2025年9月14日）」

・総務省(2025)『第27回参議院議員通常選挙開票結果』

URL : https://www.soumu.go.jp/senkyo/senkyo_s/data/sangiin27/index.html