Political Persuasion Campaign: Uplift Modeling Analysis

Comprehensive Analysis of Voter Persuasion using Causal Machine Learning

October 2025

Tanmay Topkhanewale, Aditya Jain, Ayush Lingwal

Contents

Executive Summary

This report presents a rigorous causal analysis of voter persuasion using uplift modeling. Our objective is threefold: (1) quantify the causal effect of campaign contact on individual voter persuasion probabilities, (2) identify characteristics that predict treatment effect heterogeneity, and (3) translate model outputs into operational targeting strategies for field deployment.

We analyzed a validated dataset of **4,052** voters with pre-treatment demographic, geographic, and political features. The main results are summarized below and discussed in greater depth in subsequent sections.

- **Positive uplift:** 98.5% of voters exhibit a positive individual uplift from campaign contact (3,993 out of 4,052).
- **Negative uplift ("backfire"):** Only 1.5% of voters show negative uplift (59 voters).
- **Average uplift:** 2.17 percentage points increase in persuasion probability.
- **Top decile:** Average uplift of 4.05% with a conversion rate of 50.3% (404 voters), representing the highest-return segment.
- **Best model:** Gradient Boosting trained on the top 20 features achieved an AUC of 0.905 on uplift-relevant classification metrics.

Operational recommendation: prioritize targeted outreach to the top 2–3 uplift deciles (approximately 1,203 voters) to achieve a substantially better cost-per-conversion than mass contact. The remainder of this document details data handling, causal methods, model results, segmentation strategy, validation, and practical next steps for deployment.

1. Introduction & Problem Context

Political campaigns today must optimize scarce outreach resources while navigating the complex behavioral responses of voters. Traditional predictive models estimate propensity to support, but they conflate baseline support with persuadability. In resource-limited settings, the critical decision is not merely whom to identify as supporters, but *whom to contact* so that contact changes behavior.

Uplift modeling reframes the problem as a causal inquiry: rather than predicting an outcome, it estimates the *incremental* effect of a treatment (contact) on the outcome (movement toward support). This distinction is operationally meaningful: contacting sure-things yields low marginal return, while contacting persuadables yields high marginal return.

While the statistical underpinnings of uplift models are well-known in marketing and healthcare, their application to political persuasion raises unique considerations: heterogeneous baseline partisanship, potential backfire effects in politically charged contexts, and legal/ethical constraints on outreach. Our analysis explicitly incorporates these considerations by focusing on pre-treatment features, preventing leakage, and auditing for subgroups with potential backfire.

1.1. Scope and Objectives

This analysis aims to provide an end-to-end pipeline that a campaign can operationalize: from data validation through uplift estimation, interpretation, and deployment recommendations. Specific objectives include:

1. Estimate individual-level uplift scores and quantify their distribution.
2. Identify which covariates predict uplift heterogeneity.
3. Evaluate multiple modeling approaches and select a robust, interpretable candidate for production.
4. Translate model outputs into a decile-based targeting plan with expected conversion gains and efficiency calculations.

2. Data Understanding & Preparation

2.1. Data Sources and Validation

The validation set contains 4,052 voter records with the following high-level variables:

- **Demographics:** age, sex, marital indicators, household composition.
- **Neighborhood features:** median age, community codes, education levels, community size indicators.
- **Political attributes:** party registration flags, registration recency (REG_DAYS), eligibility flags.
- **Treatment assignment:** MESSAGE_A (binary indicator whether a voter received campaign contact).
- **Outcome:** MOVED_A (binary indicator capturing movement toward support).

Records were validated for completeness and for plausible ranges (e.g., REG_DAYS nonnegative, age within reasonable bounds). Missingness patterns were analyzed and imputed conservatively (median or mode imputation depending on variable type) only for features intended for modeling while keeping outcome and treatment untouched.

2.2. Leakage Control and Feature Selection

A crucial challenge in causal modeling is leakage: information that would not be available at the time of treatment but correlates with the outcome can produce biased treatment effect estimates. We undertook the following steps to mitigate leakage:

1. Computed correlations between each candidate feature and the outcome. Features with correlation > 0.95 were flagged for review and removed when they reflected alternate encodings of the outcome or post-treatment signals.
2. Restricted features to those plausibly known before any contact (e.g., demographic, registration history, community codes).
3. Used domain knowledge to exclude features derived from subsequent behavior (e.g., post-contact response times).

After these controls, we retained a focused set of features that balance predictive signal with causal validity. Later model comparisons use two feature sets: (A) the *important* top-20 features identified by uplift importance, and (B) the *full* feature set used for benchmarking.

2.3. Feature Engineering

Feature engineering aimed to create stable, interpretable covariates while preserving causal validity. Examples:

- **Marital flags** (M_MAR, F_MAR): explicit binary indicators for married males and females constructed from household and relationship fields.
- **Community granularity** (COMM_609P, COMM_LT10): categorical encodings representing geographic or administrative units, including a flag for communities with fewer than ten households.
- **Education buckets** (ED_4COL, ED_NOCOL): coarse-grained education categories to reduce sparsity.
- **Registration recency (REG_DAYS):** transformed via log1p for stability.

These engineered features both improved model stability and yielded interpretable relationships during SHAP analysis.

3. Causal ML: Uplift Modeling with CausalML

3.1. Conceptual Primer

Uplift modeling estimates the conditional average treatment effect (CATE) at the individual level:

$$\tau(x) = E[Y(1) - Y(0) \mid X = x],$$

where $Y(1)$ and $Y(0)$ denote potential outcomes under treatment and control respectively. Identification rests on the assumption of unconfoundedness conditional on X (no unobserved confounders), which is plausible here because treatment assignment was randomized or pseudo-randomized within campaign constraints. Where randomization is imperfect, we rely on covariate adjustment and careful feature selection to reduce bias.

Practically, uplift models can be implemented via meta-learners (T-learner, S-learner, X-learner), specialized uplift algorithms (Uplift Random Forest, uplift gradient boosting), or causal forests. We evaluated an ensemble of approaches and report results for the highest-performing and interpretable candidates.

3.2. Modeling Approach

Our pipeline includes:

1. **Split:** training, validation, and a held-out test/validation set (the 4,052 records).
2. **Baseline estimators:** logistic regression and propensity score models to establish a baseline.
3. **Specialized uplift models:** UpliftRandomForestClassifier (CausalML), gradient-based uplift learners, and meta-learners implemented with tree ensembles.
4. **Interpretability:** SHAP values computed on uplift-relevant scores and feature importances aggregated across folds.
5. **Robustness checks:** calibration plots, subgroup analyses, and decile lift charts.

3.3. Top Predictive Features (Uplift Importance)

Table 3.1 lists the top 20 features ranked by uplift importance from the causal uplift estimator. These features are not merely correlated with the outcome—they predict differential response to treatment.

Feature	Importance	Interpretation
M_MAR	0.0705	Married males - strong persuadability signal
COMM_609P	0.0567	Community type - geographic targeting
PARTY_R	0.0499	Republican registration
F_MAR	0.0461	Married females
COMM_LT10	0.0418	Small community indicator
MED_AGE	0.0409	Neighborhood median age
ED_4COL	0.0369	4-year college education
REG_DAYS	0.0364	Registration recency (engagement)
E_PELIG	0.0319	Eligibility flag
HH_NR	0.0304	Household non-registration status
VG_08G	0.0293	Voting group 08G
FTR_DK	0.0267	Feature DK flag
ED_NOCOL	0.0255	No college education
OTH_3PT	0.0255	Other 3-point indicator
M_NOCOL	0.0239	Male no-college indicator
COMM_45PT	0.0237	Community 45PT code
NH_WHITE	0.0221	Race flag (White)
PARTY_D	0.0220	Democratic registration
E_ELIG	0.0218	Alternate eligibility flag
VG_06G	0.0212	Voting group 06G

Table 3.1: Top 20 uplift-predictive features from the CausalML uplift estimator.

3.4. Interpretive Notes

The dominance of marital-status indicators (M_MAR, F_MAR) suggests household structure plays a central role in receptivity to campaign messages, potentially through joint decision-making or shared exposure channels. Community codes (COMM_609P, COMM_LT10) indicate geographic clustering: persuadability is localized, pointing to benefits from geographically targeted field operations.

4. Uplift Score Distribution & Analysis

4.1. Summary Statistics

The distribution of uplift scores is summarized in Table 4.1. These aggregate metrics help build a targetable strategy by quantifying central tendency and spread.

Metric	Value
Mean Uplift	2.17%
Standard Deviation	1.07%
Minimum	-0.95%
25th Percentile	1.41%
Median	2.09%
75th Percentile	3.00%
Maximum	4.99%

Table 4.1: Distributional metrics for predicted individual uplift scores.

4.1.1. Interpretation

While mean uplift is modest (2.17%), the upper tail contains voters with substantially higher uplift—up to 4.99%. Targeting based on uplift captures these high-return individuals and improves efficiency relative to blanket contact.

5. Machine Learning Model Performance

5.1. Models and Metrics

We compared 11 algorithms on two feature sets: (A) the top-20 important features and (B) the full feature set. Metrics reported include accuracy, AUC, precision, recall, and F1 score. AUC here measures the model's discrimination power for uplift-relevant ranking (via uplift-specific evaluation or surrogate scoring).

5.1.1. Important Features (Top 20)

Model	Accuracy	AUC	Precision	Recall	F1 Score
GradientBoosting	83.8%	0.905	73.9%	86.2%	79.6%
CatBoost	83.5%	0.903	74.5%	83.3%	78.6%
LightGBM	82.9%	0.899	74.0%	82.0%	77.8%
AdaBoost	83.0%	0.894	71.6%	88.6%	79.2%
XGBoost	81.7%	0.888	73.3%	78.5%	75.8%
RandomForest	82.0%	0.884	73.4%	79.7%	76.4%

Table 5.1: Performance metrics for models trained on the top 20 features.

5.1.2. Full Feature Set (Benchmark)

Model	Accuracy	AUC	Precision	Recall	F1 Score
CatBoost	95.7%	0.993	91.8%	96.8%	94.2%
LightGBM	95.7%	0.992	91.9%	96.8%	94.3%
XGBoost	95.7%	0.992	92.4%	96.3%	94.3%

RandomForest	94.8%	0.984	90.2%	96.3%	93.1%
GradientBoosting	94.4%	0.983	89.5%	96.1%	92.7%

Table 5.2: Performance metrics for full feature set models (used as an upper-bound benchmark).

5.2. Discussion

full-feature models achieve near-perfect AUC, which likely reflects the inclusion of many correlated variables and potential overfitting to training specifics. The 20-feature models present a pragmatic balance: a modest decrease in AUC (0.08) for much greater model simplicity, interpretability, and lower maintenance costs. We recommend deploying the 20-feature GradientBoosting model in production with ongoing monitoring.

6. Uplift-Based Targeting Strategy

6.1. Decile Analysis and Tactical Tiers

We segmented the population into deciles by predicted uplift score and computed average uplift, conversion rate, and counts per decile. Table 6.1 highlights key deciles.

Decile	Avg Uplift	Std Dev	Conversion Rate	Count	Recommendation
0 (Lowest)	0.41%	0.33%	30.5%	406	Low Priority
1	0.97%	0.10%	31.1%	412	Low Priority
2	1.40%	0.14%	35.4%	398	Medium Priority
3	1.73%	0.06%	31.2%	413	Medium Priority
4	1.96%	0.08%	42.8%	397	Medium Priority
5	2.26%	0.12%	33.5%	439	Medium Priority
6	2.61%	0.11%	39.3%	384	Medium-High Priority
7	2.99%	0.11%	33.7%	427	High Priority
8	3.38%	0.12%	38.7%	372	High Priority
9 (Highest)	4.05%	0.32%	50.3%	404	Highest Priority

Table 6.1: Decile-level uplift statistics and recommended priorities.

6.1.1. Operational Tiers

- **Tier 1 (Must Contact):** Deciles 8-9 (top 20%, 776 voters). Highest ROI — contact with strongest messaging.
- **Tier 2 (High Value):** Deciles 6-7 (next 20%, 811 voters). Contact if budget allows; use standard persuasion content.
- **Tier 3 (Moderate):** Deciles 3-5 (30.8%, 1,249 voters). Contact with lower-frequency or automated channels.

- **Tier 4 (Low Priority):** Deciles 0-2 (30%, 1,216 voters). Deprioritize to conserve resources.

6.2. Efficiency Scenarios

Using average uplift and counts, we estimate additional conversions under different targeting scenarios:

1. **Contact Everyone:** Expected uplift = 2.17% $\times$ 4,052 88 additional votes.
2. **Contact Top 3 Deciles (7-9):** Voters = 1,203 (29.7%). Expected uplift 3.5% $\times$ 1,203 42 additional votes (1.6$\times$ efficiency).
3. **Contact Top Decile Only (9):** Voters = 404 (10%). Expected uplift 4.05% $\times$ 404 16 additional votes (1.8$\times$ efficiency).

These calculations are illustrative and assume uniform costs per contact. A real deployment should account for marginal contact cost differences across channels (phone, door-to-door, SMS).

7. Key Findings & Strategic Implications

7.1. Summary of Findings

- **Widespread positive uplift:** 98.5% positive, indicating campaign contact is broadly beneficial.
- **Heterogeneous response:** Uplift varies substantially across deciles, providing a clear signal for targeted resource allocation.
- **Marital status matters:** M_MAR and F_MAR are top predictors, suggesting household-focused messaging opportunities.
- **Geographic clustering:** Community codes strongly predict uplift; localized field strategies are advised.
- **Simplicity vs. performance:** Top-20 features retain most predictive power while reducing complexity.

7.2. Implications for Campaigns

From a strategic perspective, these results allow the campaign to:

- Allocate field canvassing and volunteer time to voters with the highest expected marginal impact.
- Design message variants targeted to demographic and community profiles to maximize lift.
- Reduce unnecessary contact with low-uplift segments, lowering costs and potential backfire risk.

8. Recommendations & Next Steps

1. Deploy uplift scoring and generate contact lists for top deciles
2. Develop tailored messages for married voters and communities with high uplift
3. Conduct A/B tests to validate, compare targeted versus random contact
4. Build real-time scoring dashboard for optimization
5. Incorporate ongoing data and post-election validation for continuous model improvement

9. Limitations & Ethical Considerations

9.1. Limitations

- **Identification assumptions:** The unconfoundedness assumption is plausible but not guaranteed; any non-random treatment assignment can bias estimates.
- **Data scope:** Results are contingent on the validation set; external validity to other populations or future elections requires caution.
- **Modeling choices:** Different uplift algorithms may yield different individual rankings; ensemble approaches and calibration are recommended.

9.2. Ethical considerations

Applying causal models to political persuasion has ethical dimensions:

- **Transparency:** Be transparent with stakeholders about models' limitations and intended use.
- **Respect privacy:** Ensure data handling complies with applicable privacy laws and internal policies.
- **Avoid manipulation:** Targeting should aim to inform and engage, not to exploit vulnerabilities or spread misinformation.

10. Conclusion

This uplift modeling exercise demonstrates that causal machine learning can materially improve campaign efficiency by focusing outreach on those most likely to change behavior in response to contact. The recommended operational approach—deploy the gradient-boosting model trained on the top 20 features and prioritize the top 2–3 deciles—offers a pragmatic path to enhance persuasion with lower cost and risk.