

Surya Yoganathan, Michael Lin, Trang Phan, Sridhathri Divakaruni

Northeastern University Philanthropy Report

I. Introduction

Northeastern University Advancement is responsible for strengthening constituent engagement and securing philanthropic support for the institution. A key strategic priority is identifying individuals who are most likely to make a planned gift of \$100,000 or more, gifts that shape long term financial stability and support future academic and research investments.

The central business problem is that planned gift prospects are extremely rare. In a dataset of more than 300,000 constituents, only around 0.10% have established a planned gift. Current identification methods rely heavily on officer intuition, manual review of donor profiles, and anecdotal experience. This approach is not scalable and risks overlooking high potential prospects. Advancement asked for a data-driven framework that would help staff prioritize outreach and focus limited cultivation resources on the constituents most likely to make a significant planned gift.

The objectives of this project were to (1) estimate planned gift likelihood using predictive modeling, (2) segment the constituent base into actionable cultivation tiers, and (3) translate analytical insights into specific recommendations that enhance prospecting and long-term planned giving efforts. The team used anonymized demographic, engagement, and giving history data. Our analytical process included exploratory assessment, feature creation, predictive modeling, descriptive statistics, and segmentation. A major challenge was significant missingness in key demographic variables, especially age, which was missing for more than half of the dataset and not missing at random. Because demographic imputation would have introduced bias, we relied on behavioral indicators such as lifetime giving, recency of giving, and engagement activity as practical substitutes for donor maturity and capacity. Additional project

work included merging flat files, standardizing formats, validating fields, and creating constituent-level summary features.

Taken together, this project offers Northeastern Advancement a structured, scalable approach for identifying high potential planned gift donors and deploying resources strategically to strengthen long term fundraising outcomes.

II. Results

Our analysis produced three main outputs: model performance results, feature level insights, and donor segmentation based on predicted likelihood of making a planned gift of 100,000 dollars or more. These results will provide Advancement with a structured and data informed view of which constituents are most likely to be strong planned gift prospects.

A. Model Performance

To evaluate how effectively different algorithms could identify high-value planned-gift prospects, we developed and tested three models: Logistic Regression, XGBoost, and LightGBM. Because planned-gift donors represent only 0.10% of all constituents, traditional accuracy metrics are not meaningful; instead, we relied on precision–recall AUC (PR-AUC) and top-K capture rates, which measure the model’s ability to surface the highest-likelihood donors.

Overall Predictive Strength (PR-AUC)

- **Logistic Regression:** 0.0377 ($\approx 21 \times$ better than random)
- **XGBoost:** 0.051 ($\approx 53 \times$ better than random)
- **LightGBM:** 0.327 ($\approx 341 \times$ better than random)

LightGBM demonstrates the strongest overall discrimination across the full donor population, performing exceptionally well at broad ranking tasks. However, PR-AUC does not fully reflect performance in the topmost donor strata, where Advancement focuses its time and investment.

Top-K Capture Rates: Where It Matters Most

We evaluated how many true planned-gift donors each model identifies within the segments that Advancement most actively cultivates: the top 1%, 5%, and 10% of constituents.

Segment	XGBoost	LightGBM
Top 1%	37.7%	14.8%
Top 5%	67%	65.6%
Top 10%	75%	80.3%

XGBoost dramatically outperforms LightGBM at the Top 1%, capturing 2.5× more likely planned-gift donors. This concentration of predictive power is far more consequential for Advancement's strategy than overall population level lift.

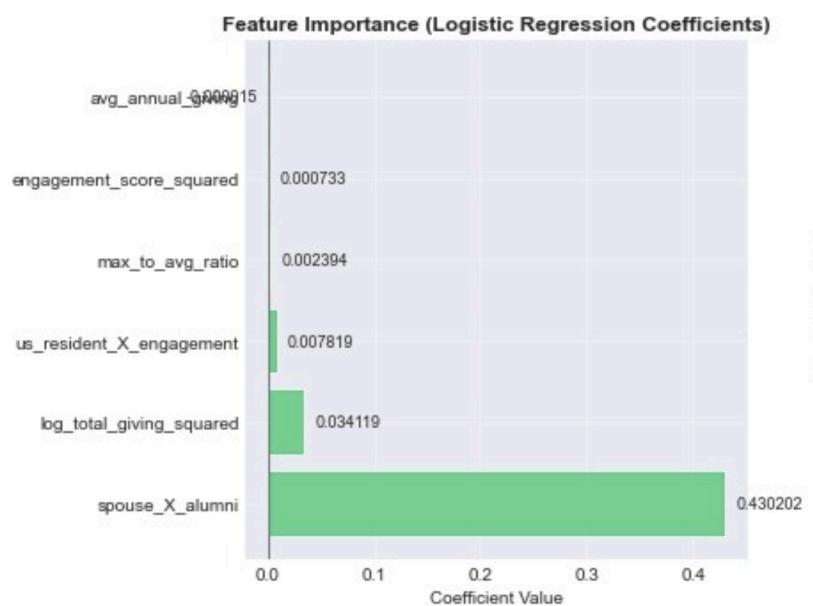
Why XGBoost Is the Recommended Model

While LightGBM delivers stronger global performance, Advancement does not operate across the entire distribution of donors. Instead, its highest impact cultivation efforts such as personal visits, estate-planning conversations, and customized stewardship are reserved for the top 1–10% of prospects. XGBoost excels precisely in the donor segments that matter most for Advancement. Its 37.7% capture rate in the top 1% provides senior gift officers with a highly reliable pool of 3,182 top tier prospects, while its comparable performance in the top 5%

maintains continuity across midlevel cultivation pipelines. By concentrating a higher proportion of true planned-gift donors in Tier 1, XGBoost delivers significantly greater return on high touch cultivation resources. In practical terms, this superior top-end precision allows Advancement to focus its most expensive and time-intensive outreach on the donors most likely to respond, reduce wasted cultivation effort on low-probability prospects, and build officer portfolios grounded in data rather than intuition. Thus, although LightGBM is the strongest model in aggregate, XGBoost is the optimal operational model for Northeastern's fundraising strategy because it delivers the highest impact exactly where Advancement invests the most.

B. Key Predictive Features

Our feature engineering created interaction terms and quadratic features to capture non-linear relationships in donor behavior. The logistic regression model (without age due to 53.9% missing values) revealed the most important predictors:



Top Predictive Features:

- **spouse_X_alumni (Coefficient: 0.430)** - Alumni with spousal connections to the university show the strongest planned giving propensity, reflecting how dual household connections create deeper institutional ties and family legacy considerations
- **log_total_giving_squared (Coefficient: 0.034)** - The squared logarithmic term captures that cumulative giving effects accelerate at high levels; moving from \$100K to

\$1M in lifetime giving has exponentially greater impact on planned giving likelihood than moving from \$10K to \$100K

- **us_resident_X_engagement (Coefficient: 0.008)** - The interaction between domestic residency and high engagement reveals that geographic proximity amplifies the effect of behavioral commitment, enabling easier relationship building and face-to-face cultivation

Additional significant features include engagement_score_squared, which captures threshold effects where highly engaged donors (top 10-20%) are exponentially more likely to establish planned gifts than moderately engaged donors, and max_to_avg_ratio, which distinguishes consistent major donors (stronger PG prospects) from one-time exceptional gift donors.

C. Donor Segmentation

Based on XGBoost predictions applied to the full donor database, we created three actionable cultivation tiers:

Tier 1: Top 1% (3,182 donors) - Contains 37.7% of all likely planned gift donors with a 37.7% hit rate (392x better than the 0.1% baseline). These highest-confidence prospects warrant premium cultivation including personal visits from senior gift officers, one-on-one estate planning consultations, and exclusive legacy society events. With approximately 115 planned giving donors in this tier and an estimated average gift size of \$250,000, this represents \$28.75 million in potential value.

Tier 2: Top 5% (15,910 donors) - Captures 67% of all likely donors (adding 29% beyond Tier 1) with a 13.4% hit rate. This tier is ideal for mid-touch cultivation including legacy society invitations, targeted seminars, and personalized planned giving mailings. The incremental 90 donors represent approximately \$22.5 million in additional potential value, justifying investment of \$1,000-\$2,000 per donor annually.

Tier 3: Top 10% (31,820 donors) - Reaches 75% of all likely donors (adding 8% beyond Tier 2) with a 7.5% hit rate. These emerging prospects are best served through scalable educational content, webinars, and email nurture campaigns. The incremental 24 donors represent \$6 million in potential value and can be cultivated efficiently through marketing-led digital programs.

D. Patterns and Anomalies

Several non-linear patterns emerged that challenge traditional assumptions about planned-giving behavior. The quadratic term **log_total_giving_squared** shows that wealth effects do not scale linearly; donors with more than \$500,000 in cumulative giving are exponentially more likely to establish a planned gift than those with lower lifetime giving. This supports the idea that transformational donors deepen their commitment over time and should be prioritized for estate-planning conversations.

The **spouse_X_alumni** interaction further reveals that dual-affiliation households, where both partners have Northeastern ties, exhibit dramatically higher planned-gift propensity than single-affiliation households. This finding underscores the importance of cultivating not just individual alumni but entire households, as shared institutional identity appears to amplify long-term philanthropic intent.

A notable modeling anomaly is that **XGBoost outperforms LightGBM at the top 1%**, despite having lower overall discrimination across the dataset. This pattern reflects XGBoost's tendency to concentrate predictive power among the very highest-scoring prospects, even if this comes at the expense of broader ranking accuracy. For Advancement, this behavior is beneficial: the cultivation strategy centers on the top 1–10% of donors, making XGBoost's top-heavy precision more valuable than LightGBM's wider but less focused performance.

Another insight is the distinction between **consistent major donors** and donors who have made only one exceptional gift. Donors with stable giving patterns, low variance across gift size, exhibit much stronger planned gift likelihood. In contrast, individuals with a single unusually large contribution but otherwise modest giving are significantly less likely to make a planned gift. This finding challenges the common intuition that any major gift signals long term philanthropic intent and highlights the importance of sustained giving behavior over isolated spikes.

III. Interpretations and Business Implications

The XGBoost model shifts Advancement from broad outreach to targeted cultivation. Instead of contacting all 318,202 constituents to find the 0.1 percent of planned gift donors, focusing on the top 10 percent of donors captures 75 percent of likely prospects. This reduces outreach volume by 90 percent and raises the hit rate from 0.1 percent to 7.5 percent. Targeting the top 1 percent, a group of 3,182 donors, reaches nearly 40 percent of all planned gift prospects and improves efficiency by a factor of 377. These results support a tiered cultivation strategy that aligns effort with predicted likelihood. Tier 1 donors warrant high touch outreach from senior officers. Tier 2 donors can be served through semi-personalized programs. Tier 3 donors are well suited for scalable digital cultivation. This structure allows Advancement to invest where the probability of conversion is highest while still maintaining coverage of emerging prospects.

The model also improves portfolio management. Officers move from managing large pools of low-probability donors to focused portfolios where one in three Tier 1 donors and one in eight Tier 2 donors are likely planned gift prospects. This increases the efficiency of cultivation work and creates more opportunities for meaningful donor conversations. Budget allocation becomes more disciplined, with spending tied to demonstrated likelihood rather than intuition. Feature-level insights further enhance cultivation. Donors identified through spouse and alumni

connections often respond to legacy and family impact messaging. High lifetime givers, identified through cumulative giving patterns, are strong candidates for conversations about long term support and estate planning. Tailoring outreach to these behavioral signals improves relevance and strengthens planned giving outcomes.

IV. Recommendations

1. Integrate Model Scoring into Advancement Operations

Northeastern should adopt the XGBoost scores as the primary method for identifying planned giving prospects. All 318,202 active constituents should be scored and assigned to three cultivation tiers: 3,182 donors in Tier 1, 15,910 in Tier 2, and 31,820 in Tier 3. These scores should be incorporated directly into the CRM with automated quarterly updates, ensuring that officers always have current likelihood assessments. Embedding scoring into core systems also standardizes prospect evaluation across teams, reducing reliance on subjective judgment. Consistent use of these scores will help Advancement build a more predictable and measurable prospecting pipeline. Integrating predictive scoring also ensures that cultivation strategies remain aligned with real behavioral changes, such as shifts in giving patterns or engagement. Over time, this creates a dynamic intelligence framework that enhances decision-making and increases the long-term accuracy of prospect identification.

2. Restructure Gift Officer Portfolios Using Tier Assignments

Gift officer portfolios should be redesigned around the model's tiers rather than geography or alphabet. Senior officers should manage Tier 1 portfolios of approximately 200 to 250 high-propensity donors. Mid-level officers can be assigned Tier 2 prospects, with portfolios of 300 to 350 donors. Tier 3 should be managed primarily through marketing teams using scalable digital engagement. Aligning portfolios to predicted likelihood ensures that cultivation

effort is directed where the probability of planned giving is highest. This structure also creates clearer performance expectations and supports more efficient allocation of officer time. By anchoring portfolios to likelihood rather than historical norms, Advancement can concentrate its most experienced officers on the donors with the greatest potential return. This approach not only improves officer productivity but also increases the number of planned giving conversations with donors most likely to convert.

3. Establish Tier-Specific Cultivation Strategies

Each tier should have a standardized playbook tailored to its likelihood profile.

- **Tier 1:** High touch cultivation, including personalized outreach, leadership engagement, and tailored impact reports.
- **Tier 2:** Semi personalized programming such as seminars, testimonial mailings, and small group events.
- **Tier 3:** Efficient digital stewardship through webinars and automated nurture campaigns. This ensures consistent donor experiences while making the best use of organizational resources.

4. Use Predictive Features to Personalize Messaging

Model insights should guide how planned giving opportunities are framed. Donors with strong spouse and alumni connections respond well to legacy focused messaging, while donors with significant cumulative giving tend to react positively to discussions about extending their long term impact. Using these behavioral indicators to shape communication increases relevance and raises conversion rates. Personalized messaging also strengthens donor relationships by demonstrating that Advancement understands individual motivations. As these targeted strategies are refined and scaled, they can meaningfully increase the overall effectiveness of planned giving outreach. Incorporating model driven personalization allows

officers to anticipate donor interests and tailor outreach before a conversation even begins, improving both engagement quality and officer efficiency. Over time, these refined messaging strategies can become templates for broader planned giving campaigns rooted in data rather than intuition.

V. Conclusion and Reflection

This project shows that predictive analytics can significantly strengthen Northeastern's planned giving strategy by revealing where the most promising opportunities truly lie. Instead of treating all constituents as equally likely prospects, the analysis demonstrates that planned giving potential is highly concentrated among donors with consistent engagement, significant lifetime giving, and strong household ties. Recognizing this pattern allows Advancement to replace broad outreach with a targeted, evidence based approach that improves efficiency and elevates the return on cultivation efforts. The segmentation framework developed through XGBoost provides a clear path for prioritizing resources. By identifying the small group of donors most likely to establish planned gifts, the model helps officers focus on relationships with the greatest long term value. This system also supports more disciplined portfolio design and enables leadership to align staffing and investment with predicted outcomes rather than intuition. Incorporating these insights into the CRM ensures that prospect identification becomes an ongoing, dynamic process rather than a one time exercise. The work also highlights opportunities for continued innovation. Integrating new behavioral signals, refining messaging strategies based on model features, and exploring predictions of expected gift size could help Advancement deepen personalization and improve conversion rates. Feedback from frontline fundraisers will be essential to validate predictions and adapt strategies as donor behavior evolves. By combining behavioral insights with operational recommendations, it equips Northeastern Advancement with tools that can strengthen donor relationships, enhance cultivation efficiency, and support the university's long term financial resilience.