Dataset:

individual_features_engineered.csv:

- 318K rows (individual donors from combined FY1976-2025 transactions)
- Planned gift data merged at individual level
- Target: household_has_pg (1 if ≥$100K planned gift, else 0)

Python Files

Logistic_Regression.py

4 engagement breadth features:

1. unique_campaigns - number of different campaigns donor supported
2. unique_designations - number of different funding areas donor gave to
3. unique_opportunity_types - variety of gift types
4. unique_credit_types - mix of hard credits and soft credits

XGBoost.py and Light_gbm.py

14 features (6 base + 2 quadratic + 6 interactions):

Base features:

1. is_alumni - graduated from NU (1/0)
2. gift_cv - coefficient of variation in gift amounts (inconsistency)
3. years_since_last_gift - recency of giving
4. has_masters - has graduate degree (1/0)
5. avg_annual_giving - average dollars per year
6. age - donor age (53.9% missing, XGBoost learns from pattern)

Quadratic terms (capture accelerating effects): 7. log_total_giving_squared = $\text{log\_total\_giving}^2$ 8. engagement_score_squared = $\text{engagement\_score}^2$

Interaction terms (created automatically): 9. us_resident_X_engagement = is_us_resident × engagement_score 10. male_X_engagement = is_male × engagement_score 11. friend_X_engagement = is_friend × engagement_score 12. spouse_X_alumni = has_spouse × is_alumni 13. max_to_avg_ratio = max_gift ÷ avg_gift_size 14. log_total_giving_x_engagement = log_total_giving × engagement_score