Predict, Pursue, Propel:
Driving Targeted Constituent Engagement

MIT MANAGEMENT BUSINESS ANALYTICS | OPERATIONS RESEARCH CENTER

Oxfam Advisors: John Abdulla, Nivetha Nagarajan, Debbie Medvinsky
Faculty Advisor: Professor Robert Freund


Bhargavi Lanka


Emily Nowak

About Oxfam

Oxfam International is an organization that **fights inequality to end poverty and injustice**. **Supporters, or constituents,** of Oxfam **fall into different categories**, primarily characterized by their giving and action history.

- **Lead** — Took 1 Action in past year
- **Advocate** — Took 2+ Actions in past year
- **Donor** — Gifted in past 3 years
- **Donor / Advocate** — Gifted in past 3 years **AND** Took 2+ Actions in past year

Donor Sub-Categories

Time on File
- Lapsed — Donated > 36 mo. ago
- Donor — Within past 36 mo.
- Consecutive — 2+ years in a row
- Committed — 4+ years in a row

Donor Level — Amount Gifted (over the past 3 years)
- Low-Level — <= $999
- Mid-Level — $1,000 – $9,999
- Major-Level — $10,000+

Problem Statement

Oxfam's constituents are essential to delivering on their mission, making it critical to reach and engage as many of them as possible. Current marketing strategies are not fully data-informed.

Our goal is to **improve the efficacy of constituent outreach by leveraging data to enable targeted marketing efforts** through

- Identifying **who is in Oxfam's constituent base** to broadly guide decision-making
- Determining **who Oxfam should reach out to** across various marketing channels

Data

- **1M+ Constituents**
- **XK Leads**
- **XK Active Donors**
- **600+ Features**
- **400+ Demographic Features**

Methodology

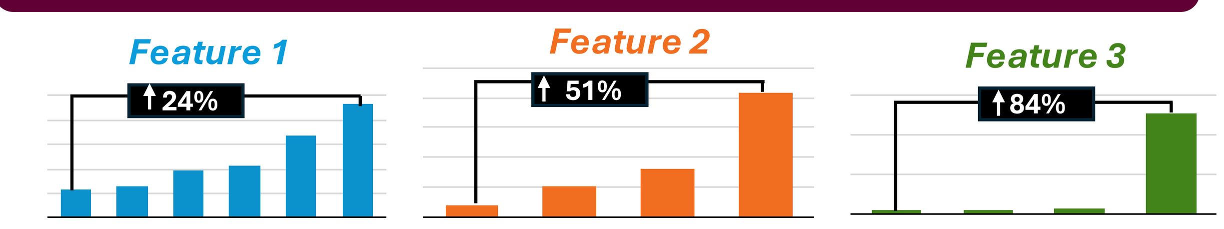
Demographic, Engagement, and Financial Data → *Data Aggregation and Processing → EDA → Demographic Analysis Tables

1. **Constituent Clustering** — Find clusters of constituents to see if natural groupings exist
2. **Supporter Prediction Models** — Predict categories that Leads might fall to build targeted outreach lists for marketing efforts

Target Tool → Targeted Marketing Outreach

*Engagement and Financial data are used to create the various Supporter categories. Models are built exclusively on Demographic data.

Exploratory Data Analysis



EDA confirmed that the data is very homogeneous, so other approaches were needed to understand constituent groupings

EDA resulted in a script to automatically generate **Demographic Analysis Tables**, like this one

Supporter Category by Urbanicity

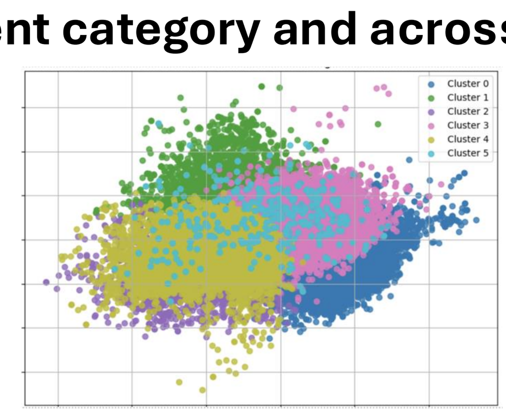
urbanicity	ZZ	Lead	Advocate	Donor	Donor/Advocate
Rural	15%	16.50%	18.90%	13.20%	10.80%
Suburban	57.8%	57.40%	58%	58.20%	55.80%
Urban	27.1%	26.10%	23.10%	28.60%	33.30%

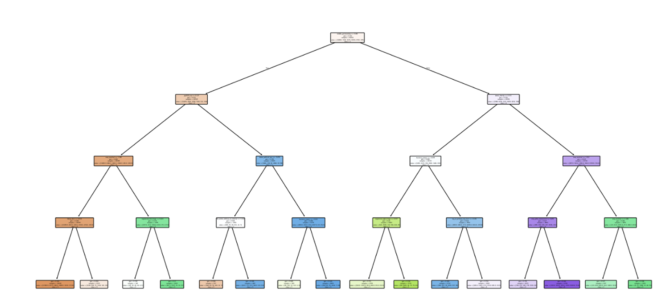
Clustering Constituents

Goal: Cluster constituents by demographics to find natural groupings within the supporter base

For each constituent category and across all constituents (Advocates category shown as example):

400+ Features → **PCA to reduce dimensions** (retaining 80% variance) → 82 Components → **KMeans** (Elbow plot to find K) → K Clusters → **CART to map features to clusters** → Important Features





middle_partisanship (0.235596)
pp2020_rep (0.178982)
emp_medical (0.159299)
donor_charity (0.141543)
business_owner (0.138651)
emp_pilot_commercial (0.036924)
likely_dem (0.033099)
reg_or_primary_rep_ever (0.025948)
registered_right_fringe (0.023678)
donor_political_any (0.015214)

Features that lacked interpretability or relevance consistently ranked as important, limiting value for meaningful segmentation

Supporter Prediction Models

Goal: Predict a constituents behavior given only demographic data

Filter data based on Prediction Outcome → **Pre-process data and test different models** (**Handling Class Imbalance:** BCW, SMOTE, Random Undersampling; **Modeling:** XGBoost, RF) → **Select best model** (Recall and Accuracy prioritized in model selection) → **Outputs**

Demographic Data → Supporter Category, Donor Time on File, Donor Level

SUPPORTER CATEGORY — TEST RESULTS

Model	Accuracy	Macro-Avg Precision	Macro-Avg Recall
Base XGB	0.7232	0.5225	0.6998
RF + SMOTE	0.7216	0.5129	0.7107
Base RF	0.7199	0.4968	0.7338
RF + BCW	0.7126	0.5055	0.7259
XGB + SMOTE	0.6513	0.4059	0.5904
XGB+Random Undersampling	0.6197	0.5838	0.5584
Baseline	0.595	0.149	0.25
RF+Random Undersampling	0.5669	0.7478	0.425

Base XGB
Improvement on Baseline model:
+ 12.8% Accuracy
+ 37.35% M-A Precision
+ 44.98% M-A Recall

TIME ON FILE — TEST RESULTS

Model	Accuracy	Macro-Avg Precision	Macro-Avg Recall
RF + BCW	0.6701	0.4829	0.684
Base RF	0.6651	0.4295	0.6732
Base XGB	0.6622	0.4445	0.6382
RF + SMOTE	0.6587	0.4356	0.6544
XGB + SMOTE	0.5949	0.2847	0.5394
Baseline	0.589	0.147	0.25
XGB+Random Undersampling	0.5202	0.5347	0.5115
RF+Random Undersampling	0.3448	0.6467	0.2759

RF + BCW
Improvement on Baseline model:
+ 8.11% Accuracy
+ 33.59% M-A Precision
+ 43.4% M-A Recall

DONOR LEVEL — TEST RESULTS

Model	Accuracy	Macro-Avg Precision	Macro-Avg Recall
RF + BCW	0.959	0.5315	0.8984
Base RF	0.9589	0.5215	0.9403
Base XGB	0.958	0.5249	0.9266
RF + SMOTE	0.9568	0.5091	0.9001
Baseline	0.945	0.315	0.333
XGB + SMOTE	0.9431	0.4265	0.5562
XGB+Random Undersampling	0.6875	0.6375	0.6028
RF+Random Undersampling	0.6242	0.7043	0.3497

Base RF
Improvement on Baseline model:
+ 1.39% Accuracy
+ 20.65% M-A Precision
+ 60.73% M-A Recall

Outputs generated from each best model create filterable files that make up the **Target Tool**, like this one for Supporter Category

Predicted Supporter Category	Age Bucket	Household Income	Length at Residence (years)	N
Donor	65+	$50k-$74k	15	Yes
Donor/Advocate	65+	$50k-$74k	15	Yes
Donor	65+	$50k-$74k	9	No
Donor	65+	$75k-$99k	0	Yes
Donor/Advocate	65+	$100k-$149k	15	Yes
Donor/Advocate	65+	$50k-$74k	11	Yes
Donor	65+	$50k-$74k	11	No
Donor	65+	$75k-$99k	15	No
Donor	65+	$100k-$149k	4	Yes
Donor/Advocate	65+	$100k-$149k	4	Yes

In this example, the file is filtered to predicted Donors and Donor/Advocates in California that are cat and dog enthusiasts aged 65+. This filtering is useful because it allows marketing teams to find people who may align with the messaging of their outreach and the intended goal of outreach (donation, action, etc.)

End Products

Demographic Analysis Tables
An automatically generated script that provides a current demographic summary of Oxfam's supporter base

Target Tool
Files containing the supporter predictions from the built models along with several demographic features for filtering purposes

Impact

12 Hours Saved Annually by Demographic Analysis Tables

21% Predicted Efficiency Improvement of Marketing Outreach

Future Work
- Run A/B Test to assess impact
- Deploy and scale Target Tool for future outreach