



# Predict, Pursue, Propel: **Driving Targeted Constituent Engagement**





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 3 years

Gifted in past

**Donor** 

**Donor / Advocate** 

Gifted in past 2+ years in a row 3 years AND Took 2+ Committed Actions in past year 4+ years in a row

Time on File **Donor Level Amount Gifted**  Lapsed (over the past 3 years) Donated > 36 mo. ago Donor

**Donor Sub-Categories** 

 Low-Level <= \$999 Mid-Level \$1,000 - \$9,999 Major-Level \$10,000+

#### **Problem Statement**

**EDA** 

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



Constituents





XK



**Active Donors** 



400+ **Demographic Features** 

# Methodology



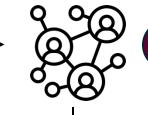
Demographic, Engagement, and

and Processing Financial Data

**Demographic Analysis Tables** 

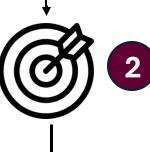
\*Data Aggregation

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



**Constituent Clustering** 

Find clusters of constituents to see if natural groupings exist



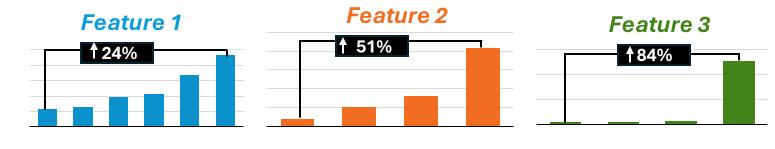
**Supporter Prediction Models** 

Predict categories that Leads might fall to build targeted outreach lists for marketing efforts



Targeted Marketing Outreach

# **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 automatically genera **Demographic Analys** Tables, like this or

to	Supporter Category by Urbanicity					
ite	urbanicity	ZZ	Lead	Advocate	Donor	Donor/Advocate
sis	Rural	15%	16.50%	18.90%	13.20%	10.80%
	Suburban	57.8%	57.40%	58%	58.20%	55.80%
ne	Urban	27.1%	26.10%	23.10%	28.60%	33.30%
<b>→</b>		27.170	_0.1070		20.0070	30.0076

RF+Random Undersampling

XGB+Random Undersampling

RF+Random Undersampling

Model

RF + BCW

Base XGB

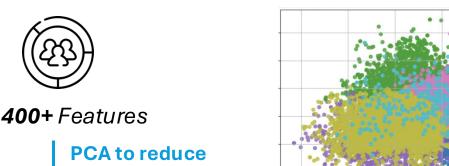
RF + SMOTE

XGB + SMOTE

Base RF

## **Clustering Constituents**

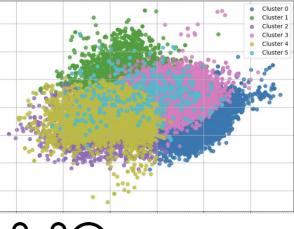
Goal: Cluster constituents by demographics to find natural groupings within the supporter base For each constituent category and across all constituents (Advocates category shown as example):



dimensions (retaining 80% variance)

Components

**KMeans** (Elbow plot to find K)



**CART** to map features to K Clusters clusters Important Features

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)

middle\_partisanship (0.235596)

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

**Outputs** 

# **Supporter Prediction Models**

Supporter

Category

Goal: Predict a constituents behavior given only demographic data



Pre-process data and test different models Handling Class Imbalance: BCW, SMOTE, Random Undersampling Modeling: XGBoost, RF

SUPPORTER CATEGORY **TEST RESULTS** Macro-Avg Precision Model Macro-Avg Recall Accuracy Base XGB 0.7232 0.5225 RF + SMOTE 0.7216 0.7107 0.5129 Base RF 0.7199 0.4968 0.733 RF + BCW 0.7259 0.7126 0.5055 XGB + SMOTE 0.6513 0.4059 0.5904 0.5838 0.5584 XGB+Random Undersampling 0.6197 0.595 0.149

0.5669

TIME ON FILE **TEST RESULTS** 

0.670

0.6651

0.6622

0.6587

0.5949

0.589

0.5202

0.3448

**Accuracy** 

0.7478

Macro-Avg Precision | Macro-Avg Recall

0.6732

0.6382 0.6544

0.5394

0.5115

0.2759

0.25

0.4829

0.4295

0.4445

0.4356

0.2847

0.147

0.5347

0.6467

Demographic Data	Donor Time on File

4	DONOR LEVEL						
	TEST RESULTS						
((\$))	Model	Accuracy	Macro-Avg Precision	Macro-Avg Reca			
	RF + BCW	0.959	0.5315	0.89			
7	Base RF	0.9589	0.5215	0.94			
Donor	Base XGB	0.958	0.5249	0.92			
Level	RF + SMOTE	0.9568	0.5091	0.90			
	Baseline	0.945	0.315	0.0			
	XGB + SMOTE	0.9431	0.4265	0.5			
	XGB+Random Undersampling	0.6875	0.6375	0.60			
	RF+Random Undersampling	0.6242	0.7043	0.34			

Select best model Recall and Accuracy prioritized in

Base XGB

+ 12.8% Accuracy

+ 37.35% M-A Precision + 44.98% M-A Recall

# RF + BCW

- + 8.11% Accuracy
- + 33.59% M-A Precision
- + 43.4% M-A Recall

#### 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

Supporter Category	Age Bucket	Household Income	Residence (years)	٢
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/Advocato	GE+	\$100k \$140k	1	Vac

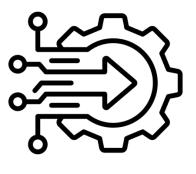
In this example, the file is filtered to predicted Donors and Donor/Advocates in California that are cat and dog enthusiasts aged 65+. This filtering is useful because it allows marketing teams to find people who may align with the messaging of their outreach and the intended goal of outreach (donation,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

Hours Saved Annually by **Demographic Analysis Tables** 

21% Predicted Efficiency Improvement of Markoting Outro of Marketing Outreach

#### **Future Work**

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