

Understanding Venezuelan Migration to Colombia

Multi-Class Classification of Drivers for Migration

Serena Quiroga
Module 5 Project,
Flatiron Data Science Program
February 28th, 2020



Problem Statement

Using Household Survey data, can we classify the primary motivations for migration from Venezuela to Colombia?

Based on question: **What was the principal motive for moving from your place of residence 12 months ago?**



HELP POLICYMAKERS AND THE
INTERNATIONAL COMMUNITY PROVIDE
THE APPROPRIATE DEVELOPMENT
AND HUMANITARIAN INTERVENTIONS



UNDERSTAND THE REASONS FOR
MIGRATION TO COLOMBIA



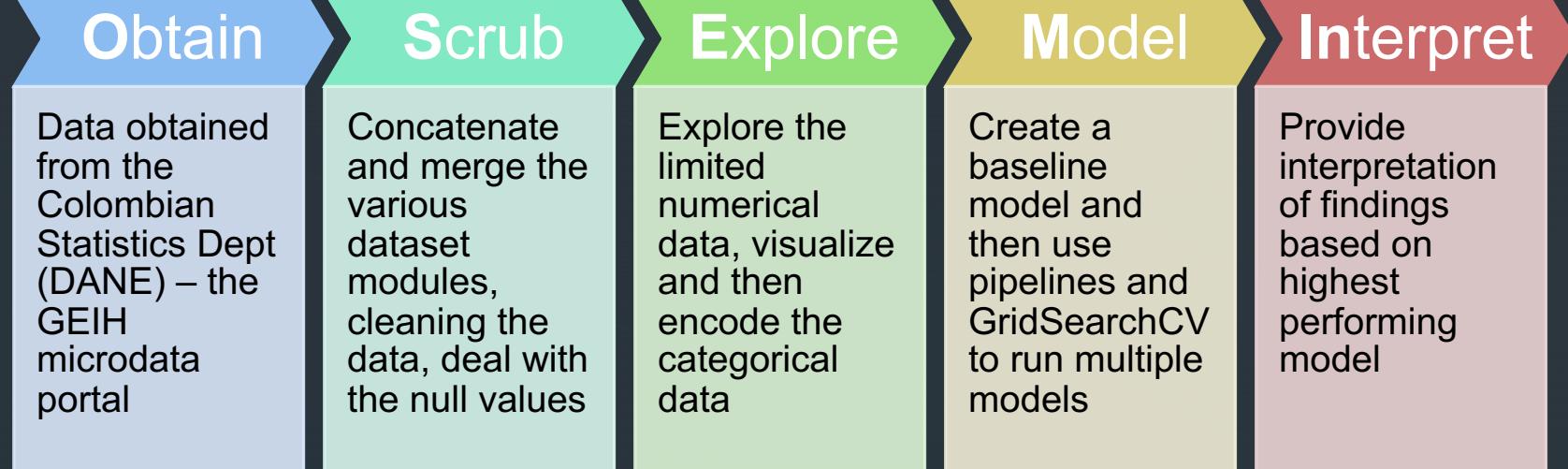
PROVIDE INSIGHTS FOR PROPENSITY
TO RETURN

Policy Value-Add

Methodology

- OSEMiN approach
- Pyreadstat
- Scikit-Learn – Pipelines, Param Grids, and GridSearchCV
- Classification Models
 - Decision Tree Classifier
 - SVM
 - Random Forest Classifier
 - AdaBoost Classifier
- Model Performance Metrics – Weighted F1 Score, Confusion Matrices, Feature Importance

OSEMIn



Overview of Migration Drivers

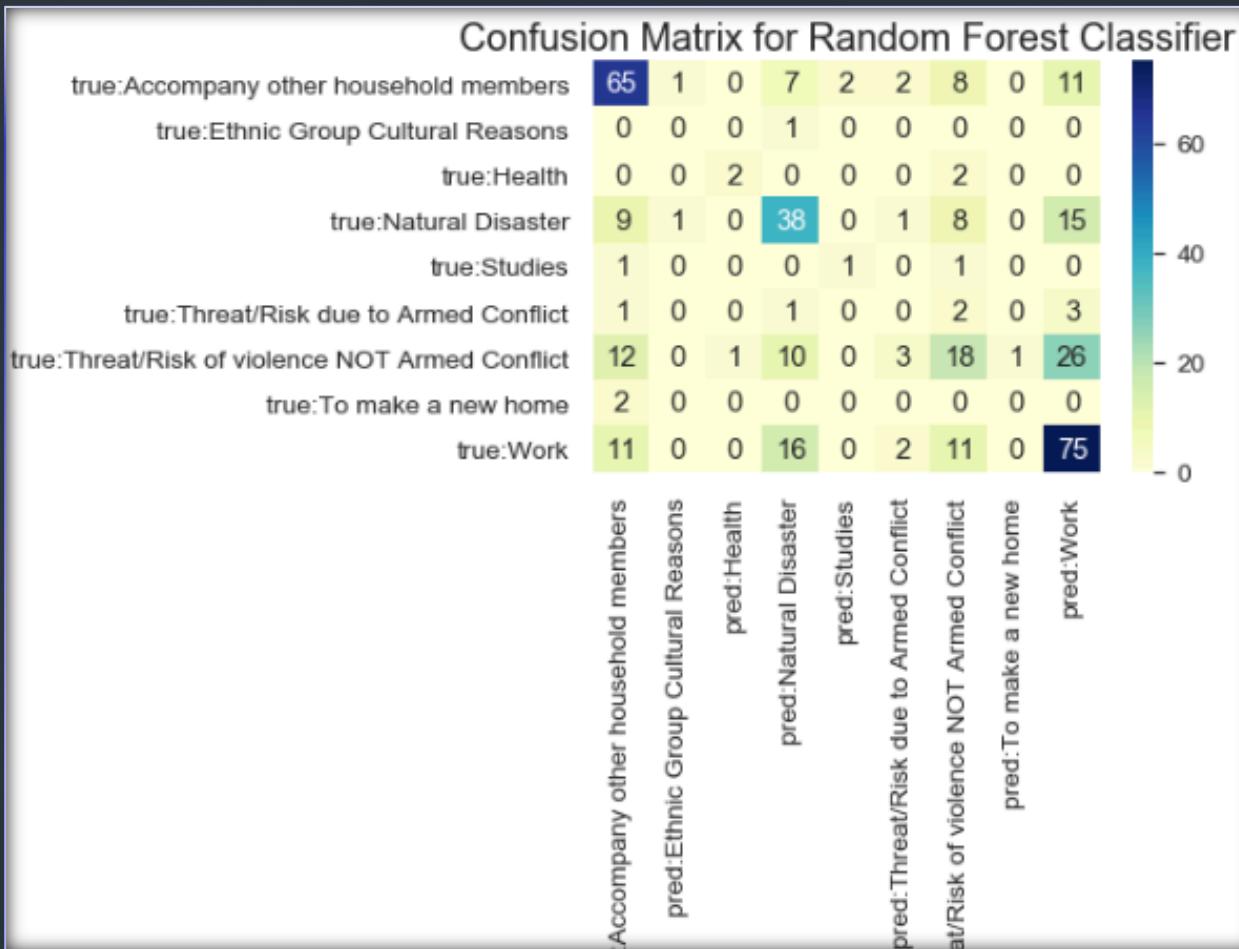
In our training data, ***Work, Accompanying another household member, Natural Disaster, and Threat/Risk of violence NOT due to Armed Conflict*** were the primary drivers in our data.



Findings

- Despite hyperparameter tuning, models perform with suboptimal performance measures.
- Highest weighted performance measure score is nearly 53% (Random Forest Classifier).
- Important features in the data are Age and Number of years of Schooling.

RF Classifier predicts with 52.8% accuracy



- **Work and Accompanying other household members** highest labeling accuracy
- Important features: **Age**, Years of **Schooling**, if currently a student, if currently employed

AdaBoost
Classifier had
52.5%
weighted F1
score

SVM and AdaBoost
Classifiers made nearly
the same correct and
incorrect predictions



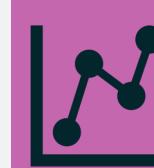
Future Considerations



Labeled data on
Propensity to Return



Extend model to integrate
data from other host
countries



Time Series Analysis over
past few years

Thank you

- **Data Used:** 2019 GEIH
Microdata from National Dept. of Statistics of Colombia (DANE)
- **Data Source:**
<http://microdatos.dane.gov.co/index.php/catalog/599/study-description>

