

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



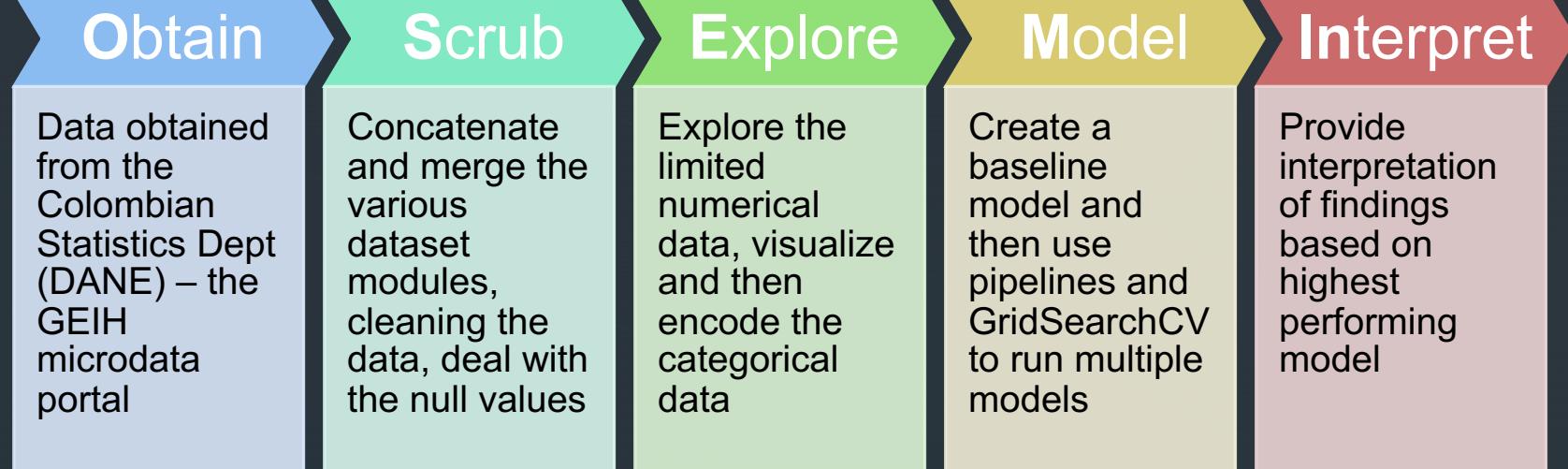
PRODUCE AN INPUT MODEL FOR
UNDERSTANDING TRENDS OVER TIME
AND FORECASTING TRENDS AND
NEEDS OF MIGRANTS

Policy Value-Add

Methodology

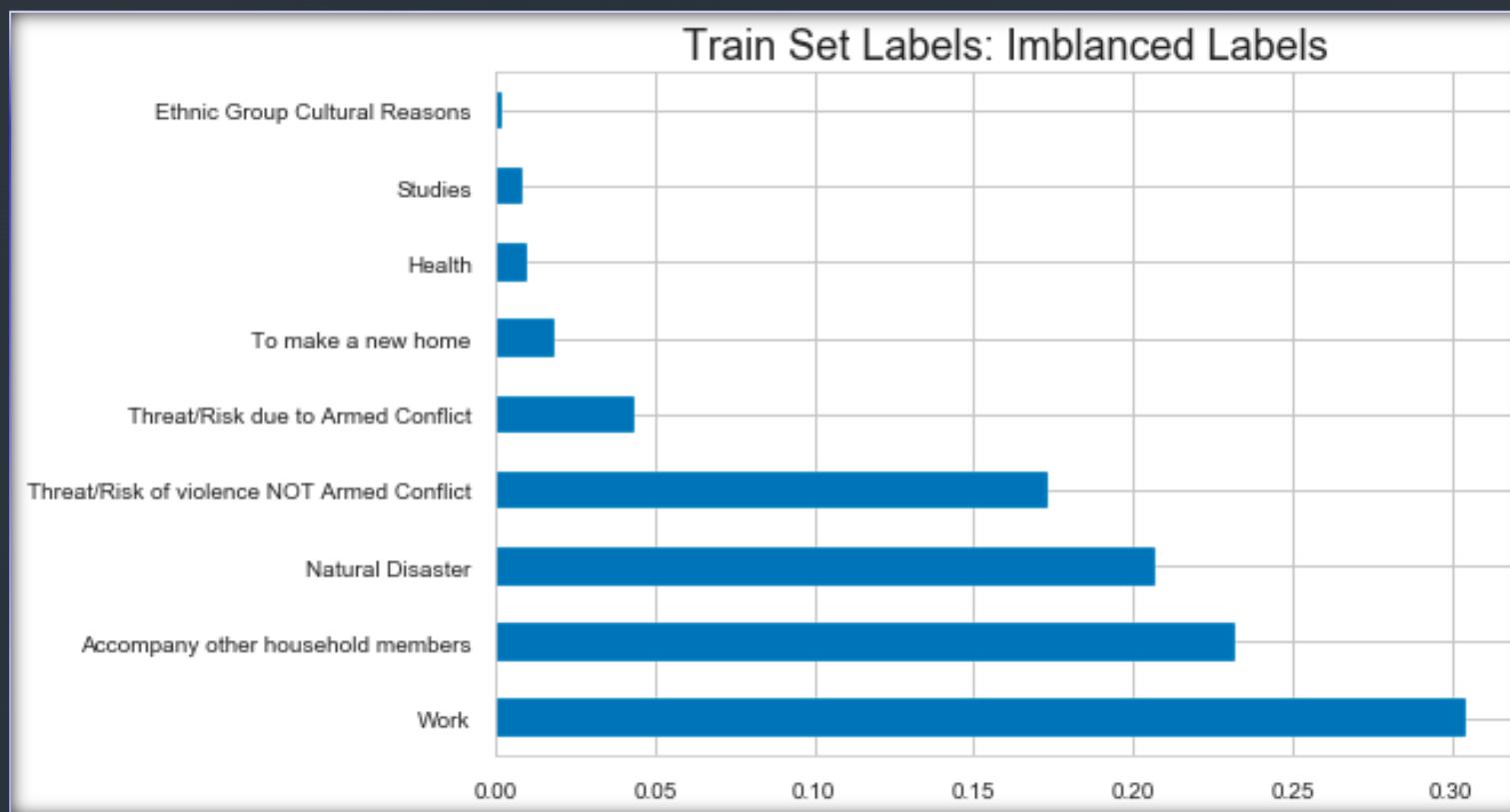
- OSEMiN Approach
- Working with different data file types
- Data cleaning & visualizations
- Machine Learning Algorithms for Classification Modeling
- Model Performance Metrics & Evaluation

OSEMIn



Overview of Migration Drivers

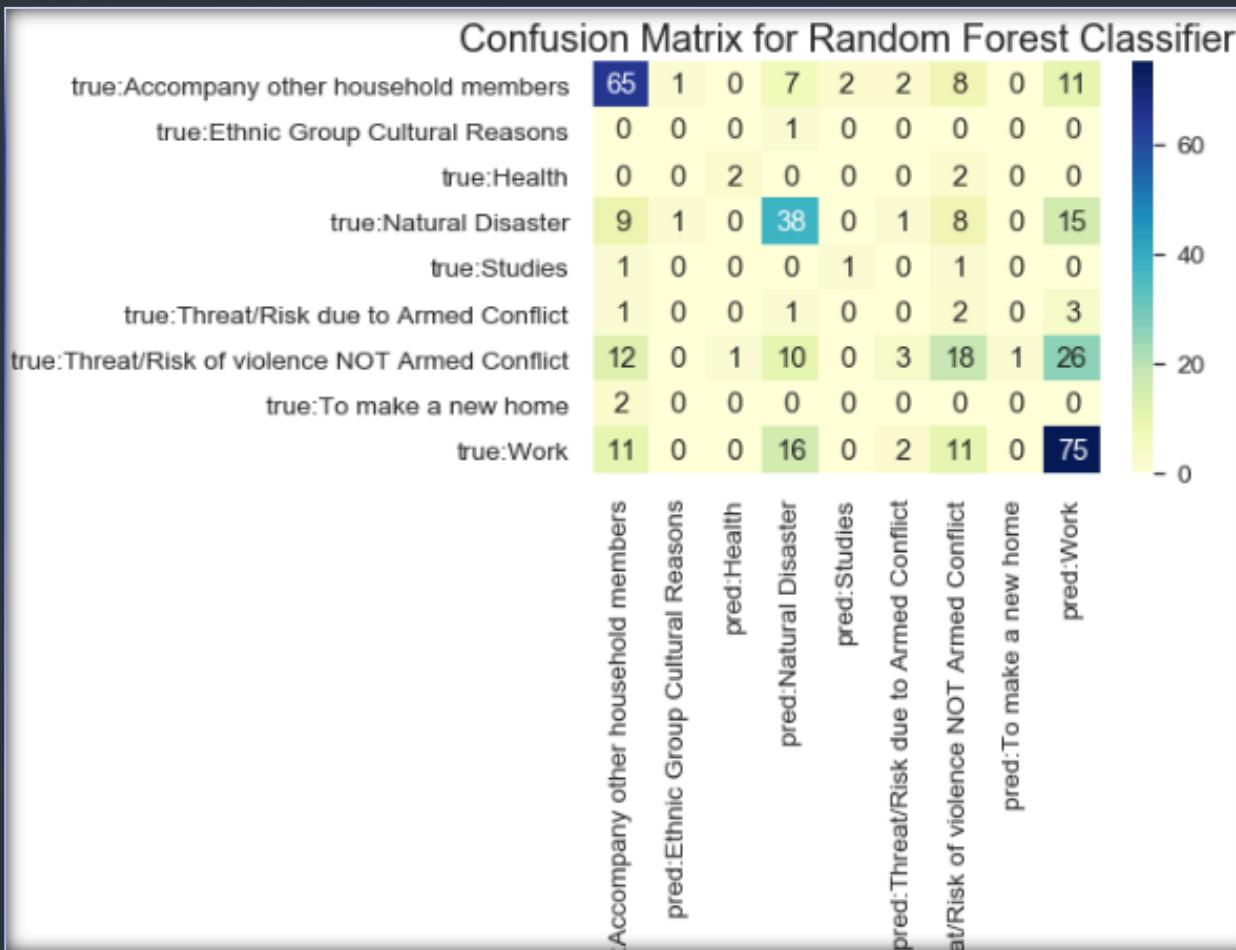
In our training data, ***Work, Accompanying another household member, Natural Disaster, and Threat/Risk of violence NOT due to Armed Conflict*** were the primary drivers in our data.



Findings

- A *Best Guess* ~30% accuracy for classifying the primary reason for migration
- Baseline model scored 43.6% at accurately classifying reasons for migration
- With an improved algorithm and better tuned parameters, we achieved a 52.8% score
- Important features in the data are Age and Number of years of Schooling

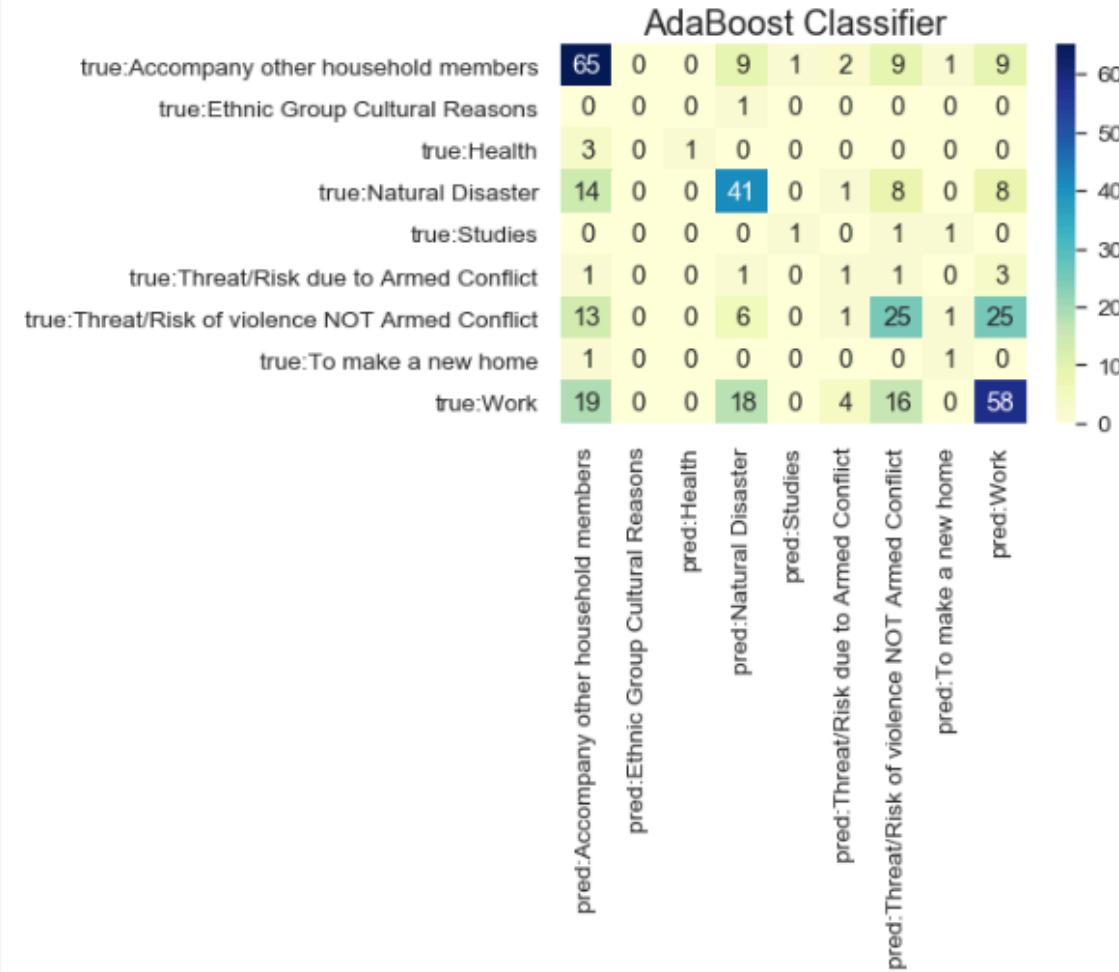
Best model predicts with 52.8% accuracy



- **Baseline score: 43.6%**
- **Work and Accompanying other household members** highest labeling accuracy
- Important features: Age, Years of Schooling, if currently a student, if currently employed

Other models
scored
around 51
and 52%

The 2 other algorithms
made nearly the same
correct and incorrect
predictions



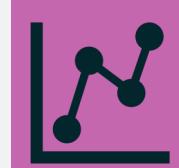
Future Considerations



Incorporate more data and more variables from other survey modules



Extend model to integrate data from other host countries



Time Series Analysis over past few years that survey was conducted

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>

