Data Mining to Help Foster Care Children

Daniel Oldham
Advisor: Professor Berezovski

Problem
Partnership for Strong Families (PSF), a foster care organization based out of Gainesville, FL, collected data regarding foster care children and their parents between 2010 and 2017. This data could prove to be extremely useful in benefitting PSF’s operations and ability to help these children. However, the scale of the data itself, with nearly 250,000 entries over these years, makes it impossible to analyze without proper data mining software and an in-depth approach.

Goals
Leverage the capabilities of modern statistical analysis (R) and data mining (Orange) software to find useful insights into this data.

These insights could include correlations between children being removed from their homes and parental characteristics, age, demographics, foster care history, mental health, and others.

Geographical Data
Total number of cases per zip code
> 4,000 cases
3,000 < cases < 4,000
1,500 < cases < 3,000
200 < cases < 1,500

Distribution of Cases 2010-2017

Data Analysis Process

- Import raw data
- Logic-check data entries, update if needed
- Add columns (year, number of cases, etc.)

Data Cleaning

- Execute R scripts for basic statistical info
- Search for insights / plot distributions
- Plan further analysis based on findings

Statistical Analysis

- Organize / filter dataset into subgroups
- Select characteristics for analysis
- Export subgroup datasets into Orange

Data Wrangling

- Search for correlations, visualize data
- Plan further analysis based on insights
- Run machine learning algorithms

Data Mining

Present insights and recommendations