Data Mining to Help Foster Care Children

Daniel Oldham
Advisor: Professor Berezovski

Problem
Partnership for Strong Families (PSF), a child welfare services organization based out of Gainesville, FL, collected data regarding foster care children and their parents between 2010 and 2017. Some aspects included were: ages, dates, location, case history, service(s) provided, and clinical records. The 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 records 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 and data mining software to find useful insights into this data, specifically, attempt to predict what factors lead to a child re-entering the foster care system, and what can be done to keep them out.

Placement Settings: First & Second Placements
- No Case
- Foster-Rel
- Pre-Adoptive
- Foster-NRel
- Other
- Institution
- Group Home

Analysis of Different Placement Settings
By building programs in R that outputted a child’s placement settings for multiple cases over time, correlations could be seen between different initial placements and corresponding cases after that.

- Foster care with a relative is clearly superior than with a non-relative (Nrel): 50% of relative cases exit system, whereas only 25% of non-relative cases do.
- Institutional placement is least effective at keeping children out of the system (only 9% exit afterwards).
- Pre-Adoptive stage is achieved most often through foster care with a relative (12% move into). Pre-Adoptive is not a significant exit point through any other setting.

Ranking a Child’s Entire Case History
- Rank each type of placement setting in system based on insights from visualization
- Apply to all children’s placements
- Add up all ranks in a child’s case history
- This is the “weight” of their case as a whole

Final Insights: Influences on Child Retention in Foster Care System
By running a Random Forest model on the data and additional factors that we were able to extrapolate, the influential characteristics that would determine the success of a case became clearer.

Ranks are based on retention of child in foster care system
Weight is calculated based on ranks of child’s entire history of placements in system
A successful case would have a lower weight. Which factors of a child’s case influence weight?

Geographic Analysis:
By tracking the number of cases per zip code (top), visualizing problematic areas by comparing these numbers to the population density of these areas to create a per capita relationship (bottom) was fairly simple and provided significant insights into problematic areas for the foster care system in Northern Florida.

Age Analysis:
Different placement settings had different ages of children present in them. By plotting these differences, we were able to see how the children’s ages changed for different kinds of placements. For example, those in adoption had lower ages than those who were placed with foster homes.

Influences of Case Characteristics on Success
...