Predicting Permanent Housing for Homeless Families in the Puget Sound Region

Ariel Rokem, Bryna Hazelton, Kivan Polimis, Kathleen Moore
The University of Washington eScience Institute
It all started in last summer’s DSSG program

- DSSG (Data Science for Social Good): Student fellows work with project leads from non-profit, academic and/or government organizations
- Data-driven projects focused on urban data
- Second installment starts June 13th
Background

- Over **4,000** homeless families in the Tri-county (Snohomish, King, Pierce) area every year.

- BMGF and Building Changes: cut family homelessness **by half** by 2020.

- Make family homelessness **rare, brief, and one-time**.
Data Processing Pipeline

King HMIS extract

Pierce HMIS extract

Snohomish HMIS Extract

Cleaned Data for Households

Cleaned data for Families

Family Enrollments

Mapping of enrollments to episodes

Family Episodes

1 row per family per episode

Aggregating into episodes
Create family/episode aggregate variables
Defining episodes

Enrollment 1

Emergency Shelter

Enrollment 2

Rapid Rehousing

Time

Enrollment 1

Emergency Shelter

Enrollment 2

Rapid Rehousing

Inter-enrollment gap (<30 days; based on Wong et al. 1997)
Decision Trees predict family outcomes

each ‘leaf’ represents a combination of programs:

This tree predicts exit status with approximately 70% accuracy (p<0.05)
How to communicate prediction error?

Error bars for the predictions of decision trees
How to calculate error bars?

Error bars for decision tree algorithms:

- Use “random forests” of trees - a resampling strategy that makes the trees more robust…
- … and provides sub-samples for jack-knifing
- Correct for biases due to limited sampling

Wager et al. (2014)
Example
Open source software

http://uwescience.github.io/sklearn-forest-ci/

- Integrates with scikit learn
- Supports both classification and regression
- Documentation includes usage examples:
  - http://uwescience.github.io/sklearn-forest-ci/auto_examples/plot_mpg.html
Thanks!