

# STAT 222: Project 4

## Countable Care: Modeling Women's Health Care Decisions

Yuan He, Lindsey Lee, Jin Rou New, Tianyi Zhu  
University of California, Berkeley

April 12, 2015

### Abstract

In DrivenData.org's Countable Care: Modeling Women's Health Care Decisions competition, the challenge is to predict probabilities that female survey respondents used 14 different health care services in the 12 months preceding the survey based on survey data from a nationally representative sample of women in the United States. Our best solution using a basic missing value imputation method and a gradient boosting machine gave a log loss score of 0.2613, putting us at rank 30 on the leaderboard out of 483 competitors on April 12, 2015, two days before the end of the competition. This report details our entire process including data exploration, data processing and modeling with a variety of methods.

**Keywords:** countable care, Driven Data competition, women's health, machine learning, gradient boosting machine, random forest

# 1 Introduction

The competition we have chosen to work on is DrivenData.org's Countable Care: Modeling Women's Health Care Decisions. The link to the competition can be found at: <http://www.drivendata.org/competitions/6/>. DrivenData is effectively Kaggle for social issues, and this competition is a partnership with Planned Parenthood Federation of America.