

STEVEN FELIX

Cambridge, MA | steven.a.felix@gmail.com | 562.659.2167
Personal: stevenfelix.github.io | LinkedIn: steven-am-felix

SKILLS

Programming: Python (pandas, scikit-learn, gensim, nltk, flask), R, SQL

Analytic Tools: Exploratory data analysis, experimental design (A/B testing), longitudinal data analysis (mixed-effects), linear/logistic regression, natural language processing (word2vec, nltk), SVM, random forest.

RELEVANT EXPERIENCE

Fellow, Insight Data Science, Boston, MA 1/2018 – present

- Built “*search suggester*”, a Flask-based web application that improves Stack Overflow search experience by generating conceptually-related alternative queries; suggested queries are up to 29x more probable and lead to search results with 20 more votes per result
- Extracted, parsed, and pre-processed over 26M Stack Overflow question-titles from 50GB XML file of user content
- Trained *word2vec* model on question titles and used resulting model to generate, score, and rank potential alternative queries; deployed on Amazon Web Services EC2 platform

Graduate Student, Harvard University, Cambridge, MA 9/2011 – present

- Designed and executed 5 original research projects that generated diverse behavioral data about social interactions, romantic relationships, social cognition, and well-being
- Uncovered novel insights about social relationships and well-being through effective cleaning, processing, visualization, and statistical analysis (e.g., mixed-effects) in R
- Presented findings to students, colleagues, supervisors (size 3 to 100 people)
- Trained, supervised, and mentored 6 research assistants

Teaching Fellow, Harvard University, Cambridge, MA 1/2014 – 5/2017

- Relevant courses: Research Methods, Statistics for Social Sciences
- Instructed over 80 students in use of statistical software (R, SPSS)
- Coached students through process of developing and answering research questions with appropriate statistical analysis

Independent Data Science Projects, Cambridge, MA 5/2017 – 8/2017

- Scraped, analyzed, and reported trends in 7,000 profiles of mental health providers
- Designed a random forest classifier to classify survival of passengers aboard the Titanic; achieved test accuracy of 80%
- Identified employees likely to leave their job using logistic regression and SVM; test accuracy 92%

Research Associate, University of Southern California, Los Angeles, CA 9/2010 – 8/2011

- Aggregated raw, disparate social-service data from agencies across L.A.
- Consulted with data-collection entities to fully understand data content, quality, and interpretation
- Compiled complex, hierarchical dataset describing services rendered and outcomes for children and families referred to Department of Family Services

EDUCATION

Ph.D., Experimental Psychopathology, Harvard University, Cambridge, MA Expected 5/2018

B.A., Psychology, Yale Univ., New Haven, CT 5/2009