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

RELEVANT EXPERIENCE

Fellow, Insight Data Science, Boston, MA   *1/2018 – present*
- Built searchSuggester, a Flask-based web application designed to improve data scientists' Stack Overflow queries by generating semantically related alternatives
- Extracted, parsed, and pre-processed over 17M Stack Overflow question-titles from 50GB XML file of user content
- Generated word-embeddings for Stack Overflow vocabulary by training *word2vec* model on question titles, and used embeddings to suggest similar queries

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; achieved test accuracy of 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*