STEVEN FELIX

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

SKILLS

Programming: Python (pandas, scikit-learn, gensim, nltk), **R (**lme4)
Statistics: 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 Jan. 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 Researcher, Harvard Univ., Cambridge, MA Sep. 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, mentored 6 research assistants

Teaching Fellow, Harvard Univ., Cambridge, MA Jan. 2014 – May 2017

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

Independent Data Science Projects (http://stevenfelix.github.io) May - Aug.2017

- Scraped, analyzed, and reported trends in 7,000 profiles of mental health providers
- Designed a random forest classifier that accurately classified the survival of 80% of passengers aboard the Titanic; submission ranked in top 25% on Kaggle Leaderboard
- Developed a logistic regression and SVM classifiers that accurately identified 92% of (test data) employees that left their job

Research Associate, Univ. of Southern California, Los Angeles, CA Sep. 2010 – Aug. 2011

- Aggregated raw, disparate social-service data from agencies across L.A.
- Consulted with data-collection entities to fully understand data content, quality, interpretation
- Prepared complete 5-year data and extracted features for analysis

EDUCATION

Ph.D., Experimental Psychopathology, Harvard Univ., Cambridge, MA **Exp. May 2018**
B.A. Psychology, Yale Univ., New Haven, CT **May 2009**