

Scott Cole

San Francisco, CA scott.cole0@gmail.com

<https://scole.github.io> www.github.com/scole www.linkedin.com/in/scott-cole/

Skills

Languages: Python (advanced) - pandas, sklearn, scipy, seaborn, matplotlib, dash; SQL (basic); Git (basic)

Techniques: machine learning, data visualization, data cleaning, data mining, statistics, signal processing

Experience

Fellow - Insight Data Science

Jan 2019 - present

- Created *ContinU*, a model and platform to continuously authenticate users based on keystroke patterns
- Engineered features from raw keystroke data and analyzed feature importance to uniquely identify users
- Implemented gradient boosting to authenticate new users with 2% false positive and false negative rates
- Deployed an interactive dashboard to visualize results of the authentication model for individual users

Graduate student researcher - University of California, San Diego

Sep 2014 - Dec 2018

- Identified flaws in standard signal processing techniques commonly used to analyze brain waves
- Developed new time series analyses to overcome flaws and gain new information from noisy brain signals
- Led development of 3 collaborative open-source python packages on GitHub (1000+ downloads)
- Published Jupyter Notebooks alongside publications to aid replication and adoption of my techniques
- Mentored 7 students by directing projects and teaching Python, machine learning, and statistics

Data science intern - Crime Lab New York

Sep - Dec 2017

- Furthered a machine learning pipeline to predict police misconduct by feature engineering of officer activity
 - Designed a highly adaptable workflow to evaluate potential classifiers and describe trade-offs of interpretability, accuracy on several metrics, time-scale of predictivity, and “fairness”
 - Uncovered systematic differences in police officer behavior to identify problematic districts and sergeants
-

Personal data science projects

10-dimensional burrito review system

- Developed a system to systematically review over 350 burritos at over 75 taco shops across San Diego
- Analyzed the importance of and relationships between burrito features using linear regression and PCA
- Presented work to academics and engaged with the media to share the insights with the general public

Measurement of open science

- Mined the text of over 100,000 academic papers to quantify the prevalence of code sharing
 - Coordinated with a team of 8 at a hackathon to extend the project to include data sharing and 10x papers
-

Education

University of California, San Diego

Ph.D. Neuroscience

La Jolla, CA

2018

Clemson University

B.S. Bioengineering

Clemson, SC

2014