MATTHEW MERRITT SMITH

matthewmerrittsmith@gmail.com | (831) 706 - 1392 | mmerrittsmith.github.io | linkedin.com/in/mmerrittsmith

EDUCATION

University of Chicago – Chicago, IL   **September 2020 - Present**
Master of Science in Computational Analysis and Public Policy
- **Relevant Coursework:** Python Programming, Advanced Microeconomics, Advanced Statistics, Mathematics for Data Analysis

Tufts University – Medford, MA   **August 2014 - May 2018**
Bachelor of Arts in Data Science and Public Policy
- **Relevant Coursework:** Machine Learning and Data Mining, Comparative Politics, Econometrics, Mathematical Modeling, Data Structures, Discrete Mathematics, Linear Algebra, Economic Statistics
- **Awards:** Dean's List (Spring '15, Fall '15, Spring '18), High Honors on Interdisciplinary Studies Honors Thesis *Wealth Inequality: A Machine Learning Approach*, Mathematics Department Certificate of Excellence

London School of Economics – London, UK   **September 2016 – June 2017**
Study Abroad
- **Relevant Coursework:** Further Quantitative Methods, Economics in Public Policy, Public Policy Analysis, Physics and Finance

Online Courses – Various Sites (Coursera, Stanford Lagunitas, edX, etc.)   **Fall 2015 - Present**
- **Relevant Coursework:** Statistical Learning (Lagunitas), Reinforcement Learning (David Silver), Introduction to Big Data with Apache Spark (edX), Neural Networks for Machine Learning (Coursera)

EXPERIENCE

Mansueto Institute for Urban Innovation – Hyde Park, IL   **January 2020-present**
Research Engineer
- Developing "prclz" Python library to parcelize and create suggested road networks to improve road access in "slums" with Million Neighborhoods Project

University of Chicago – Hyde Park, IL   **September 2020-present**
Teaching Assistant
- Create and lead labs for the course Data and Programming for Public Policy 2, a second year MPP course in Python and R for data science

TRACE International – Annapolis, MD   **May 2018 – June 2020**
Research Associate
- Developed methodology for understanding country-level corruption now used by hundreds of multinational organizations for risk assessment (TRACE Matrix)
- Built a first-in-industry anomaly detection model using Isolation Forests with Dask for parallelized computation now used in every TRACE due diligence report
- Wrote blog posts, book chapter on connections between machine learning and bribery
- Created agent-based models of bribery using artificially intelligent agents that suggest petty bribery is highly sensitive to slight incentive changes
- Trained neural networks to translate between 9 different languages and English eliminating the need for speeding translation tasks and eliminating the need for outside consultants

Boghosian Research Group – Medford, MA   **September 2017 – May 2018**
Student Researcher
- Created novel wealth and income dataset that combines political, economic, and global factors using Pandas, BeautifulSoup
- Applied Random Forests to generate regression models to understand factors leading to wealth disparities
- Found that energy consumption and foreign direct investment are statistically related to wealth inequality

Viewpoint Government Solutions – Boston, MA   **September 2017 – December 2017**
Data Engineer Intern
- Performed data migrations from local governments databases to Viewpoint's system using SQL, Python, and Batch
- Formalized, documented, and automated the data migration process saving tens of person-hours every week
- Cleaned incoming data using Python scripts which eliminated the need for additional requests from the data provider (small town governments)

Graniterock Company – Watsonville, CA   **June 2017 – September 2017**
Data Engineering & Analytics Intern
- Migrated legacy database system to new AWS architecture saving $20k+ dollars each year
- Developed executive and transaction-level reports in Tableau, one of which resulted in a decision that saved ~$140k/year
- Evangelized Tableau within the organization to remove report development chokepoint
- Assisted in implementation of new ELT [sic] system speeding average query time by ~120x

Center for Technology and National Security Policy – Washington, D.C.   **May 2016 – September 2016**
Research Intern

- Drafted one-page research memos on a litany of subjects
- Wrote a long form article on the EU's GDPR and state of US data regulations for dissemination to National Defense University, the United States' graduate university for the Armed Forces

Tufts Center for Information & Research on Civic Learning and Engagement – Medford, MA   **January 2016 – July 2016**
Research Intern
- Collected data for study of the effects of long-term volunteer work on employment outcomes afterward adding hundreds of data points to the study
- Ad hoc statistical analysis, data visualizations, database management

Enigma (Tufts Data Science Journal) – Medford, MA   **August 2015 – May 2018**
Writer / Senior Writer
- Created multi-variable regression model to understand over 75% of variability in voting demographics using US Census data
- Studied inaugural addresses as a microcosm of US linguistic trends, found patterns in length, diction using NLP methods

Fullpower Technologies – Santa Cruz, CA   **May 2015 – September 2015**
Software Engineering and Quality Assurance Intern
- Drafted Python scripts to clean, filter, and store over 3 TB of sleep data from worldwide Jawbone Upmove devices
- Created and implemented experimental procedures to perfect the SmartBed sensor now used in Tempur-Pedic and Serta bedframes worldwide
- Wrote more robust user-friendly Bash scripts to interface with testing microcomputers which sped testing by 30% per person

SKILLS AND INTERESTS

Technical Languages: Python (Expert), SQL (Expert), R (Expert), Bash (Proficient), Javascript (Proficient), C++ (Basic), Scala (Basic)

Packages & Libraries:
- Python: Pandas, Numpy, Sklearn, Flask, Spacy, Matplotlib, Mesa, Seaborn, Tensorflow, multiprocessing, Dask, requests, HARK, Redis, Pillow, gensim, word2vec
- R: Shiny, Tidyverse, ggplot2, ctree, partykit, parallel
- Javascript: D3.js, jQuery, Ajax

Software: Tableau (Expert), Git (Expert), Excel (Expert), *nix (Ubuntu, Raspbian, etc.) (Proficient), Stata (Basic)

Interests: Computational social science, powerlifting, ethics, chess, bikepacking, piano, basketball analytics, animal cognition, generative art, causal inference, skateboarding, whittling, data policy

Languages: Spanish (Conversational), German (Elementary)

PUBLICATIONS & PRESENTATIONS

Change the Rules, Break the Game: A COVID19-cancelled presentation at the OECD 2020 Anti-Corruption and Integrity Forum on the use of artificially intelligent agents in agent-based models for studying bribery

Modeling the Bribery Game: A chapter introducing artificially intelligent agents and agent-based models to a non-technical audience in the edited book *Corrosive*

The Truth about AI and Due Diligence An article for *The FCPA Blog* providing guidance on machine learning tools for compliance and where such technology might lead the compliance field in the future

SIDE PROJECTS

Open Source Contribution – Mesa: Added support for visualization of many agents in same cell and experimented with transition from plot.js to d3.js
Open Source Contribution – Rothko: Wrote an R package to provide color palettes that correspond to Mark Rothko paintings

Professional Volunteering – DataKind DC: Built a prototype data visualization website for Chesapeake Monitoring Society

Hobby Projects:
- Created a force directed graph of my knowledge, interests, and beliefs in d3.js
- Built a personal blog with Ruby (Jekyll) and wrote short articles for it
- Trained a generative neural network (ByteNet) on text I've written to output text similar to what I would write
- Built a "Magic Mirror" using a monitor and a RaspberryPi to show weather, photos, commute, etc.