MICAH J SMITH

Experienced researcher and engineer seeking full-time ML engineering role to start Fall 2021.

micahjsmith@gmail.com
+1 805 657-2325
www.micahsmith.com
Boston, MA & remote
he/him

EDUCATION

Massachusetts Institute of Technology, Dept. of EECS — 2016 - pres.
Ph.D. Student, Computer Science, expected Summer 2021
S.M., Computer Science, 2018

Columbia University, Columbia College — 2010 - 2014
B.A., Economics-Mathematics, *cum laude*

PROFESSIONAL EXPERIENCE

MIT LIDS, Data To AI Lab (Cambridge, MA) — Graduate Research Assistant — 2016 - pres.
- Research **ML systems**, HCI, and databases, advised by Dr. Kalyan Veeramachaneni

Botkeeper (Boston, MA) — Machine Learning Engineer (part-time) — Fall 2019 - pres.
- Lead company-wide **ML engineering** efforts in transaction classification for accounting
- Designed and implemented cross-client transaction **embedding**, automatic model **retraining**, ML **metrics** collection/storage/querying/reporting (Python, Tensorflow/Keras, scikit-learn, MongoDB, Kubernetes)

Twitter Cortex (New York, NY) — Machine Learning Engineering Intern — Summer 2018
- Designed and implemented **hyper parameter tuning** via **Bayesian optimization** for production **ML workflows** (Python, Airflow, Spearmint, Tensorflow, Django)
- Enabled simple configuration and deployment of “smart” tuning on production models such as pRecaptcha

Kensho Technologies (Cambridge, MA) — Machine Learning Intern — Summer 2017
- Developed **time series ML model** to predict trading behaviors at Treasuries desk of major US dealer
- Focused on creative **feature engineering** and principled **model selection and tuning** to improve on baseline ROC AUC score by 0.11 (pandas, scikit-learn, statsmodels, LightGBM, TPOT)

Federal Reserve Bank of New York (New York, NY) — Senior Research Analyst — 2014 - 2016
- As **project manager** and **lead developer** of open-source, high-performance Julia package (DSGE.jl), led design, implementation, performance engineering, optimization, and community engagement
- Performed **statistical and econometric analysis** for projects including quantitative monetary policy analysis, consumer expectations, and subprime mortgages (MATLAB, Stata, d3js, SQL, Python)

SKILLS

General	**Python**, **TypeScript**, JavaScript, Java, Bash, Julia, C++, C, Scala, MATLAB, Haskell
DS/ML	**pandas**, numpy, **scikit-learn**, tensorflow, matplotlib, seaborn, keras, statsmodels
Data eng.	MongoDB, SQL, Airflow, SQLAlchemy, Spark, Dask
DevOps	**Python packaging**, **Docker**, Docker Compose, **Kubernetes**, Travis CI, AWS EC2/S3/EKS/etc, Heroku, GitHub Actions/Apps
Web	**Flask**, pelican, jinja2, node, jQuery, Google Apps Script, Django, HTML/CSS, React, Tornado
Tools	**git**, **GitHub**, GitLab, **Jupyter Lab/Nb/Hub**, vim, ***nix**, **LaTeX**, make, sphinx, VS Code, PyCharm, Eclipse, pants, invoke

ACTIVITIES

- **Open-source** developer: BTB, ATM, AutoBazaar, ballet, Assemblé, DSGE.jl, FredData.jl, repolockr, etc.
- Organizer/Mentor, MIT EECS Graduate Application Assistance Program
- VP Communications/Social Chair, MIT EECS Graduate Student Association
- Bartender, MIT Muddy Charles Pub
- Running, biking, tennis, basketball, reading, coffee, crosswords, chess, playing with my dog Mamba

SELECTED PUBLICATIONS

- “Enabling collaborative data science development with the Ballet framework.” arXiv 2020.
- “The Machine Learning Bazaar: Harnessing the ML Ecosystem for Effective System Development.” SIGMOD 2020.
- “FeatureHub: towards collaborative data science.” DSAA 2017.
- “Query optimization for dynamic imputation.” VLDB 2017.