

# Tim Hopper

---

Experienced machine learning engineer and Python developer

---

## Experience

**Varo Bank, *Remote Staff Software Engineer, Machine Learning, December 2021 - present***

- Develop machine learning platform for fraud detection, risk mitigation, lending, and marketing
- Lead development of batch feature store using Tecton, enabling democratized feature creation and discovery
- Guide team in implementing engineering best practices.

**DTN, *Remote Data Science Architect, August 2020 – December 2021***

- Built and maintained data science infrastructure with AWS and CloudFormation
- Enabled machine learning product development by building tools and platforms for data scientists
- Trained and mentored data scientists in Python tooling and software engineering practice
- Strategized and led the development of AWS infrastructure for big data usability and accessibility

**Blackberry | Cylance, *Remote Senior Data Scientist, Nov 2017 – May 2020***

- Lead development of data lake and feature store using serverless AWS infrastructure
- Developed and maintained Python packages for data ingestion, cleaning, and standardization
- Guided data science team in best practices for Python development and software testing
- Researched machine learning techniques for dynamic detection of malware

**Distil Networks, *Remote Data Scientist, Oct 2015 – Nov 2017***

- Researched and developed machine learning techniques for detection of malicious web bots
- Lead research for validation of JavaScript-based client fingerprinting technique
- Constructed hierarchical Bayesian models for analysis of internet user behavior
- Primary developer of real-time identification platform for malicious web bots on Apache Storm
- Developed Python-based, internal tools for fast querying of data warehouse
- Introduced team to best practices in Python development and data analysis

**Qadium, *Remote Data Scientist, Feb 2015 – Oct 2015***

- Built **open-source tools** in C++ and Python for topic modeling and inference on nonparametric Bayesian models
- Wrote **instructional material** for inference on nonparametric Bayesian models in with IPython notebooks
- Maintained continuous integration with Travis-CI and automated deployment to Anaconda.org

**Parse.ly, *Remote Software Engineer, Jan 2014 – Feb 2015***

- Developed algorithms in Python for aggregation of large-scale, streaming, time-series data
- Built and maintained lambda architecture on AWS using Storm, Elasticsearch, Redis, and Cassandra
- Contributed to **streamparse**, an open-source library for real-time stream processing in Python

---

## RTI International, Durham, NC Data Scientist, Oct 2012 – Jan 2014

- Provided analytics and computational support for one of the nation's leading nonprofit research institutions
- Contributed to projects on social media analysis for public health, environmental GIS modeling, nonlinear mixed effect modeling, and text-mining-based crime forecasting

---

## Education

- 2010-2012**      **Master of Operations Research**; North Carolina State University, Raleigh, NC  
*Research area: Reinforcement learning methods for healthcare operations*  
*Instructor: Calculus 2*  
*Key Courses: Graph Data Mining, Machine Learning, Bayesian Networks, Linear Programming*
- 2009-2010**      **Phd Student in Mathematics**; University of Virginia, Charlottesville, VA  
*No degree received*  
*Instructor: Calculus 1 and Calculus 2*  
*Key Courses: Introduction to Mathematical Statistics, Complex Analysis, Real Analysis, Advanced Linear Algebra*
- 2004-2008**      **BS, Mathematics**; Grove City College, Grove City, PA  
*Minor in Applied Physics; Minor in Computer Science; Summa Cum Laude*

---

## Technical Experience

**Technical Tools**      I have an experience with a breadth of tools for machine learning, software development, and backend engineering.

- **Programming Languages (high proficiency):** Python
- **Programming Languages (some proficiency):** C++, Go, C#, Mathematica, R, SQL, Java, Javascript
- **Machine Learning Tools:** Scikit-Learn, PyTorch, MLFlow
- **Data Processing Tools:** Amazon Web Services (Batch, Lambda, ECS, S3, SQS), Apache Storm/streamparse, Apache Spark/Pyspark, Pandas, Scikit-learn, Impala, Hive, Dask
- **Infrastructure and Configuration Tools:** Terraform, Ansible, Saltstack
- **Data Stores:** Postgres, Elasticsearch, Cassandra, MongoDB, Redis, MySQL

**Open source contributions**      Contributor to a variety of Python libraries including  
cpython · datamicroscopes · Streamparse · Conda · lda · Pandas

---

## Personal Projects

- **pythonplot.com:** A Rosetta Stone of Python plotting libraries for exploratory data analysis
- **Notes on Dirichlet Processes:** Derivations, code, and notes on Dirichlet processes for nonparametric Bayesian methods
- **Into the Hopper:** An occasional podcast about machine learning, data science, and software engineering