

# Dylan Labatt Randle

dylanrandle.github.io

Email : dylanrandle@g.harvard.edu

Mobile : +1-857-999-7442

## EDUCATION

---

- **Harvard University, School of Engineering & Applied Sciences** Cambridge, MA  
*Master of Science in Data Science; GPA: 4.0* Sep 2018 – Present
  - **Distinctions:** Scholarship in Applied Computation
- **University of California at Berkeley, College of Engineering** Berkeley, CA  
*Bachelor of Science in Industrial Engineering & Operations Research; GPA: 3.9* Sep 2012 – May 2016
  - **Distinctions:** High Honors, Phi Beta Kappa, Tau Beta Pi

## EXPERIENCE

---

- **Harvard University** Cambridge, MA  
*Graduate Researcher* Nov 2018 – Present
  - **Deep Differential Equation Solvers:** Researched and developed fully unsupervised methods for solving Reynolds-Averaged Navier-Stokes equations with neural networks; discovered sampling method leading to improved convergence. Researching methods for training semi-supervised Generative Adversarial Networks (GANs) for solving differential equations; developed novel multi-discriminator training algorithm enabling tractable learning. Tech: Python, PyTorch, Harvard Supercomputer
- **Amazon Robotics** Boston, MA  
*Data Science Intern* Jun 2019 - Aug 2019
  - **Data Engineering:** Built automated, scalable data pipeline for big data (500+ billion rows) querying, storage, cleaning, and transformation. Tech: Python, Apache Spark/Hive/Hadoop, AWS EMR/Athena/S3
  - **Data Science:** Developed machine learning package for proprietary internal project. Built API for data filtering, feature selection, training, tuning, and testing models. Developed visualizations and interpretability algorithms (e.g. accumulated local effects, Shapley additive explanations) for model explanations. Greatly reduced complexity for training machine learning models. Tech: Python, AWS SageMaker, Jupyter, Pandas, scikit-learn, XGBoost, Keras
  - **Data Product:** Prototyped interactive user-interface and developed example use-cases allowing technical and non-technical users to easily train machine learning models for a broad set of tasks. Presented project to teams from across the organization
- **Hubdoc** Toronto, Canada  
*Lead Data Scientist* Jan 2017 – Jul 2018
  - **Production Deep Learning:** Developed and deployed production deep learning system using LSTMs & CNNs for entity extraction and text classification of financial documents. Models trained on over 10 TB of text and image documents. Reduced extraction time from 24 hours to 5 seconds with highly scalable, asynchronous pipeline. Cost savings estimated at \$2MM/year. Tech: Python, Keras, Tensorflow-Serving, AWS EC2/S3/KMS, PostgreSQL, RabbitMQ
  - **Data Science:** Conducted business and engineering analyses: e.g. prediction of labor requirements and anomaly detection of web scrapers. Wrote reports and built data visualizations for company intranet. Tech: Python, NumPy, scikit-learn, JavaScript, D3.js
  - **Leadership:** Regularly presented results and recommendations to C-suite. Integral in crafting team strategy and roadmap. Involved in fundraising and presentations to investors. Delivered machine learning lecture to 60+ people
- **Bank of Montreal, Capital Markets** Toronto, Canada  
*Financial Products Analyst* Summer 2014
  - **Fixed Income:** Conducted analyses of various debt products (swaps, swaptions, ABS, MBS). Wrote custom C# algorithm to analyze relationship between delta-hedging frequency and returns for Canadian swaptions; found possible trading opportunities
  - **Sales & Trading:** Compiled daily summaries of trading activity. Reviewed and analyzed sales product pitches. Supported both sales and trading with various data analyses

## PROJECTS & RESEARCH

---

- **Automatic Differentiation:** Python package implementing automatic differentiation, supporting both forward and reverse modes; stochastic gradient descent and Adam optimizers implemented as example use cases
- **Twitter Troll Classification:** Project achieving 96% accuracy classifying Twitter trolls using tweets scraped from accounts indicted for meddling in the 2016 U.S. elections
- **Tensorflow with Spark:** Training recurrent neural networks on a 1.5 TB dataset with Tensorflow on a Spark/Hadoop cluster with AWS Elastic Map Reduce
- **Microbiome Dynamics:** Modeling Granger causality with causal-LSTM model of high-dimensional experimental microbiome time-series data from mice

## TECHNICAL SKILLS

---

**Languages:** Python (NumPy, Pandas, matplotlib, scikit-learn, Keras, PyTorch, PyMC3), SQL, C

**Technologies:** AWS (EC2, EMR, S3, Athena), Apache Spark/Hadoop, Git, Jupyter, OpenMP/ACC