

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 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 Sep 2012 – May 2016
 - **Distinctions:** High Honors, Phi Beta Kappa

EXPERIENCE

- **Amazon.com** 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. Significantly reduced time spent on data collection. Tech: Python, Apache Spark/Hive/Hadoop, AWS EMR/Athena/S3
 - **Data Science:** Developed modular machine learning package for proprietary internal project. Built simple 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 of training and validating models. Tech: Python, AWS SageMaker, Jupyter, Pandas, scikit-learn, XGBoost, Keras
 - **Data Product:** Prototyped user-interface and developed example use-cases demonstrating ease with which technical and non-technical users could train machine learning models on a broad set of tasks. Wrote white paper demonstrating efficacy of package in performing previously tedious analyses. Presented project to teams from across the organization, receiving highly positive feedback
- **Harvard University** Cambridge, MA
Graduate Researcher Nov 2018 – Present
 - **Deep Learning for Differential Equations:** Researched and developed fully unsupervised methods for solving Reynolds-Averaged Navier-Stokes equations with neural networks; discovered sampling method leading to improved convergence. Researched methods for training semi-supervised Generative Adversarial Networks (GANs) for solving differential equations; developed novel multi-discriminator model enabling tractable learning. Tech: Python, PyTorch, Harvard Supercomputer
- **Hubdoc** Toronto, Canada
Lead Data Scientist Jan 2017 – Jul 2018
 - **Production Deep Learning:** Developed and deployed production deep NLP system using LSTMs & CNNs for content extraction and classification of financial documents. Models trained on 10 terabytes (30 million samples) of text and image documents. Reduced extraction time from 24 hours (manual) to less than 5 seconds with highly scalable, asynchronous inference pipeline. Cost savings estimated at \$3MM/year. Tech: Python, Keras, Tensorflow-Serving, AWS EC2/S3/KMS, PostgreSQL, RabbitMQ
 - **Data Science:** Conducted core business and engineering analyses: e.g. optimization of headcount by modeling marginal impact of labor on product performance, anomaly detection from application log data, ranking product service requests. Wrote white papers and deployed data visualizations to company intranet. Tech: Python, NumPy, scikit-learn, JavaScript, D3.js
 - **Leadership:** Regularly presented results and recommendations to executives. Integral in crafting machine learning strategy and roadmap. Involved in fundraising and presentations to investors. Delivered multiple machine learning lectures to audiences ranging from 50 to 100 people
- **Bank of Montreal, Capital Markets** Toronto, Canada
Financial Products Analyst Summer 2014
 - **Fixed Income Derivatives:** Conducted analyses of various debt products (swaps, swaptions, ABS, MBS). Wrote C# algorithm to model relationship between frequency of delta-hedging and expected returns for Canadian swaptions; uncovered potential trading opportunity
 - **Sales & Trading:** Compiled summaries of daily sales & trading activity. Reviewed and analyzed investment pitches. Supported team with ad-hoc data collection and analyses. Tech: Bloomberg Terminal, Excel

RESEARCH & PROJECTS

- **Automatic Differentiation:** Python package implementing automatic differentiation, including both forward and reverse modes; usage demonstrated through gradient-descent optimization
- **Deep Learning for Differential Equations:** Methods and results for solving differential equations with neural networks. Includes demonstrations of superior grid sampling and problem formulation in terms of Generative Adversarial Networks
- **Twitter Troll Classification:** Project investigating methods for, and achieving high accuracy in, classifying Twitter trolls indicted for U.S. election meddling
- **Microbiome Dynamics:** Modeling Granger causality with causal-LSTM model of high-dimensional experimental microbiome time-series data from mice
- **Tensorflow on Spark:** Training neural networks on a 1.5 TB dataset with Tensorflow on a Spark/Hadoop cluster with AWS Elastic Map Reduce

TECHNICAL SKILLS

Languages: Python (numpy, pandas, matplotlib, scikit-learn, PyTorch, Keras), SQL, C

Technologies: AWS (EC2, EMR, Athena, Boto/S3), Apache Spark, Git, Jupyter