Christina Bogdan

christina.bogdan@gmail.com • christaina.github.io

EDUCATION

New York University, New York, NY

Master of Science (MS) in Data Science — Sep 2015 – May 2017

- Courses: Machine Learning, Deep Learning, Statistical and Mathematical Methods for Data Science, Fundamental Algorithms, Natural Language Understanding with Deep Learning, Big Data, Inference and Representation, Monte Carlo Methods, Causal Inference, Advanced Machine Learning

Bachelor of Arts (BA) in Mathematics — Sep 2013 – May 2015

- Courses: Honors Linear Algebra (Graduate), Basic Probability (Graduate), Data Structures, Mathematical Modeling, Mathematical Statistics, Theory of Probability

PROFESSIONAL EXPERIENCE

The Climate Corporation, San Francisco, CA

Senior Data Scientist (Data Quality) — Feb 2019 – Present

- Lead multi-team collaboration in evaluation of algorithms yield monitor data cleaning for use in outcome based pricing product as well as for use in research across the organization.

The Climate Corporation, San Francisco, CA

Senior Data Scientist (Data Analytics) — Mar 2018 – Feb 2019

- Led implementation of production pipeline for major company product using Pyspark and Airflow
- Managed cross-team coordination with researchers developing model in above pipeline and upstream engineers. Led conversation with researchers in development of evaluation metrics and model stability checks.
- Led office hours for team's data science platform. Answer questions from researchers on data science best practices related to PySpark, SQL, Python, and model building.
- Advised researchers in multi-company collaboration on deep learning methods to predict yield during growing season

The Climate Corporation, San Francisco, CA

Data Scientist (Data Analytics) — Jul 2017 – Mar 2018

- Created compressed representations of weather time series data using autoencoders for use as features
- Implemented pipeline for distributed training of deep learning models at scale with PySpark
- Wrote post for tech blog on process of operationalizing distributed deep learning on our Spark platform

NYU Department of Environmental Studies, New York, NY

Research Assistant — Oct 2016 – Jun 2017

- Applied convolutional neural networks to identify wildlife trade from online retailers using image and text data
- Improved models by creating Word2Vec embeddings using large amounts of unlabeled data
- Built end to end model to scrape and clean data, classify item properties, and store results using Python, TensorFlow
- Set up and maintain a MongoDB database of potential wildlife trade and model results

Knewton, New York, NY

Data Science Intern — Jun 2016 – Sep 2016

- Designed and implemented methods for diagnosing skill levels of new students with bandit approaches in Java
- Formulated metrics to evaluate these policies and ran simulations of students to assess different approaches
- Implemented Bayesian optimization algorithm in Python for quickly optimizing parameters of recommender to automate QA using simulations

The Weather Company, New York, NY

Data Analyst (Part-time) — May 2014 – May 2016

- Automated a previously manual ETL process to integrate data from multiple sources and load in Amazon Redshift using Python. Pipeline was ran daily and fed into a business critical dashboard consumed by sales teams. Automation saved at least one hour of analyst work per day and resolved worry of human error in the complex manual process.

PROJECTS

Visualizing Twitter Networks of the 2014 Ukraine Crisis, MS in Data Science, NYU — Fall 2016

- Created general heuristic to restructure and reduce large retweet network while preserving information
- Designed interactive visualization in D3.js tracking crisis over 95 days
- Won 'Best Data Visualization' at the NYU Center for Data Science Academy Awards

LANGUAGES & TOOLS

Python (advanced), SQL (advanced), Java (intermediate), Javascript (intermediate), R (intermediate), MATLAB (novice), Lua (novice), Scala (novice), LaTeX, Vim, AWS, Git, Pandas, TensorFlow, Keras, scikit-learn, Torch, D3.js, MongoDB, Spark, YARN, Docker, Apache Airflow