Patrick McCarthy

Data Scientist and Technologist

resume@patrickmccarthy.cc

SKILLS

Python & R: scikit-learn, tensorflow, pandas
Devops: Docker, dbt, Nomad, CircleCI
UNIX CLI: bash, grep, ssh, vim

Cloud: Spark, Snowflake, AWS, GCP
Visualization: shiny, ggplot2, d3.js
Misc: SQL, git, Atlassian, Notion

EXPERIENCE

Sr. Data Scientist, Algorithms   2021-Present
Bowery Farming, New York NY

- Developed and implemented a novel data engineering technique to reflect forecast uncertainty throughout the forecasting and capacity-planning pipeline
- Ground-up rewrite of order forecasting capabilities with standard pipelining and monitoring

Sr. Data Scientist, Machine Learning Engineering   2018-2021
Data Scientist, Machine Learning Engineering   2016-2018
Dstillery, New York NY

- Data Science and engineering with 300+ million high-resolution customer profiles, 160 billion data-points daily:

  - Pioneered and implemented Spark for production use among data scientists and engineers, developed workflows and best practices.
  - Optimized data storage, information retrieval in 1000+ node Hadoop cluster

- Research and development of novel methodologies:

  - Prototyped search engine for privacy-sensitive data using NLP and embeddings
  - Novel machine learning method for demographic measurement learning from aggregate labels, including implementation of systems for multi-source ETL pipelines and performance simulation.
  - Geographic consumer behavior consulting for major partner, from statement of work through research and development through presentation to C-suite
  - Models for real-time bidding infrastructure, predicting ad viewability and likelihood of customer interaction.

Data Scientist   2013-2016
eXelate, New York NY, (acquired by Nielsen, 2015)

- Designed and built end-to-end pipelines against 1bn+ records for use on personally identifiable information in secure remote environments.
- Implemented high-performance concurrent ETL system with Python, EC2 and S3.

Qualitative Analyst Intern   2012
Acumen, Burlingame CA

- Enhanced client relationships by developing a general use JavaScript visualization tool to intuitively display hierarchical datasets. http://goo.gl/iO6BcY.
- Developed new healthcare policy insights through exploration and analysis of a multi-gigabyte dataset, including extensive use of SAS macros and ODS.

EDUCATION

M.S., Statistics   2013
University of California Los Angeles, Los Angeles, CA

B.A. Economics and Mathematics   2008
Boston University, Boston, MA