Richard Boeri Decal

Ft. Lauderdale, FL, USA · (+1) 48-448-DECAL · richard@dendra.io
crypdick · crypdick · richarddecal · AIjedi · richarddecal.com

Profile

Machine learning scientist on a personal and professional mission to fight against climate change. Expertise in applying cloud computing and statistical modeling to engineering, scientific, and real-world problems. Work in team-contexts to create and modify scientific software libraries to process and understand noisy data generated by stochastic real-world processes. Able to seek input from outside experts to create domain-crossing solutions designed to scale. Former molecular biologist.

Work Experience

Lead Full-Stack Machine Learning Engineer Fully Remote
Dendra Systems Feb. 2020 — Now

Developed and refined, in collaboration with institute stakeholders, a development plan for a machine learning group. Enabling whole-ecosystem restoration using deep learning.

- Architect, develop, maintain, and monitor infrastructure to facilitate the training, evaluation, and deployment of machine learning models.
  - Responsible for full life-cycle of dataset and model artifact creation, quality assurance; tracking artifact lineage, parameters for reproducibility (`Kedro`).
  - Scaled data processing workloads to terabyte-scale tabular and raster data (`Ray Distributed, Dask-ML, Modin`).
  - Scaled ML workflows: hyperparameter tournaments, model training, and serving (`Ray[Tune, SGD, Serve, Autoscaler], EC2`).
  - Created, optimized distributed queues, clients to maximize cluster I/O (`S3, Ray`).
  - Enforced code quality and correctness using pre-commit hooks, automated CI (`Bitbucket Pipelines`), property-based testing (`Hypothesis`), run-time validation (`Pandera`), design-by-contract (`typeguard, icontracts`).
- Headed Research, experimentation, and development of novel classification systems.
  - Researched, experimented, and productionized techniques for dealing with extremely imbalanced datasets, including: state-of-the-art data augmentation techniques, sampling strategies, representation learning, domain-specific features (`skmulti-learn, imbalanced-learn`).
  - Research semi-supervised methods to produce actionable insights in data-poor scenarios.
  - Examine and validate datasets; evaluate reliability of labelled data (`UMAP, Altair, streamlit`).
  - Translate business requirements into appropriate optimization functions.
  - Full-stack R&D, full life-cycle ownership of custom ML models (`PyTorch`).

  - R&D active learning methodology to improve data labeling efficiency.

- Produced screencasts to remotely present at company-wide Zoom meetings (`OBS Studio`).
- Managed and mentored junior machine learning engineer with in-depth code reviews.
- Roadmap, prioritize, and manage projects (`JIRA`).

Lead Data Scientist Fully Remote
PaceMate^{TM} Jan. 2019 — Jan. 2020

Build an end-to-end data processing and model training pipelines.

- Automated remote detection of cardiac arryhtmias in Internet-enabled heart implants using deep learning.
  - Developed processing pipelines for ECG data (`imbalanced-learn`, custom tools).
  - Working with cardiologists and software engineers to formulate business requirements (`YouTrack`).
  - Implemented state-of-the-art deep neural network for automated cardiac arrythmia classification specifically tuned for the device implanted in a majority of our patients (`Keras`).
  - Created data labelling dashboard for elecrophysiologists to review model predictions (`Plotly Dash`).
- Created dashboard to collate, explore, and summarize key insights from our electronic medical records.
  - Researched ML-assisted techniques for information extraction from extremely heterogeneous documents.
  - Wrote and scaled performant ETL pipelines (`SQL, PySpark, spaCy`).
  - Created dashboard to enable easy faceting and querying of EMR records to facilitate data-driven decision-making (`Plotly Dash`).
  - Created report on our data inventory and trends in our data.
- Upheld SOC2 security standards with measures such as encryption at rest, traffic tunnelling, and instance hardening.
- Presented my work and findings to various senior stakeholders such as CEO, CTO, and CIO. Interviewed by potential investors. Gave invited talk at local college.

Data Scientist Sarasota, FL
New College of FL, F.A.R. Institute Au. 2018 — Dec. 2018

Semester-long master's capstone project supervised by Dr. McDonald in partnership with the Florence A. Rothman Institute.

- Data-driven prediction of 30-day re-admission using visit clustering.
  - `visit2vec`: reduce high-dimensional patient visit data into low-dimensional embeddings using deep learning technique based on `word2vec` (`TensorFlow`).
  - Explored structure in patient visits data by clustering patient visits using t-SNE.

- Modelled patient trajectories on years of heart failure patients from Sarasota Memorial Hospital.
  - Clustered patients over time based on cardiac and non-cardiac chronic conditions (`SQL, Pandas, PySpark`).
  - Created network graphs characterizing interactions between multiple chronic conditions and heart failure and their effect on mortality (`NetworkX`)
  - Used finite state modeling to quantify interaction between chronic conditions and mortality (`PySpark, Numpy`).

Research Intern Seatle, WA
Peng Lab, Allen Institute for Brain Science June 2018 — Aug. 2018

Summer research project wherin I proposed a method that would automate the biggest bottleneck to high-throughput neural cell morphological analysis.

- Deep reinforcement learning for tracing neural structures in petabytes of noisy fluorescent microscope data.
  - Implemented proof-of-concept Deep Q Network using 3D convolutions to trace neural cell structures (`TensorFlow`).
  - Generated and augmented training data from manually traced microscopy dataset.
  - Created simulation environment for and engineered reward signals for training agents (`Matplotlib, OpenAI Gym`).
  - Contributor to `rl-medical`, a `tensorpack` extension for anatonmical landmark detection.

Classroom Mentor Fully Remote
Udacity Dec. 2017 — May 2018

Guided students 1-on-1 in Udacity's *Intro to Programming Nanodegree*.

- Taught students how to use Python visualization libraries and Pandas DataFrames in the *Python for Data Analysis Track*.

Research Assistant Seatle, WA
Fairhall Lab, University of Washington Oct. 2014 — Jan. 2016

Developed agent-based dynamical models of mosquito thermal plume navigation behavior.

- Computed and visualized flight kinematic statistics and thermal sensing statistics using windtunnel flight data (`Numpy, Seaborn, scipy[interpolate, spatial, stats], sklearn, statsmodels`).
- Formulated biophysical models of mosquito thermonavigation; applied numerical optimization algorithms to fit model to experimental data (`scipy[optimize], Pandas`).
- Created animations of thermal plume navigation models (`Matplotlib 3D, MayaVi`).

Expertise

Specialties	Deep Learning · Imbalanced datasets · Dimensionality reduction · Computer vision · Effective visualization
Languages	Python · R Tidyverse · Bash · RegEx · Español · English · Italiano
Tools	PyTorch, Keras, Tensorflow · (Geo)Pandas, MongoDB, PostgreSQL · Ray Distributed, PySpark, Dask · Plotly-Dash, Streamlit, Flask · Matplotlib, Seaborn, Altair, Plotly · BeautifulSoup, Scrapy · Docker
MLOps	AWS · Data Versioning · Kedro, MLFlow · Bitbucket CI Pipelines

Education

M.S. Data Science — Sarasota, FL
New College of Florida — Aug. 2017 — Dec. 2018

B.A., Chemistry/Biology (with honors) — Sarasota, FL
New College of Florida — Aug. 2007 — May 2011

Early admission in lieu of 4th year high school — Jupiter, FL
Harriet L. Wilkes Honors College — Sep. 2006 — May 2007

Publications, Presentations, & Teaching

- Invited talk on deep learning for automating caridac arrythmia at New College of Florida.
- Published three peer-reviewed journal articles in high impact journals (*Genetics*, *PNAS*). Published an undergraduate honors thesis.
- Presented aforementioned research at three conferences (poster sessions).
- Participated in several outreach programs for young students from low socioeconomic backgrounds.
- Invited talk on ML at high school STEM program in Torquinst, Argentina.

Selected Awards & Grants

NCF Data Scholar — 2017 — 2018
Full tuition waver for master's program.

National Institutes of Health PA-12-149 Federal grant — 2014 — 2016
Fully covered my salary and expenses at the UW Dept of Biophysics.

Florida "Bright Futures" Scholar — 2007 — 2011
Merit-based scholarship, fully covered college tuition.

Dubois-Felsmann Research Grant — 2010 — 2011
Covered reagent costs for my thesis experiments and conference admission & travel.