Edmund Hart, PhD

Senior Data Scientist

emhart.info
edmund.m.hart@gmail.com
(650) 208-4253
github: emhart

Technical

- R
- Python
- Scala
- Ruby
- Git
- SQL
- Spark
- Hive
- Neo4J
- Docker
- AWS

Methods

Time series forecasting
Mixed-effects models
Shrinkage and model selection
Resampling and cross-validahtion
Tree based methods
Cluster analysis
Neural Networks
Classification
Ensemble learning

Experience

Apple *Senior Data Scientist* **2014 - Present**
Deploy machine learning models at scale in a production environment to provide meaningful business prediction.
Build forecasting models for various lines of business to help with new initiatives
Collaborate with cross-functional teams to ensure proper instrumentation of products and build data pipelines.
Deliver insightful presentations of analysis and business results to business unit leaders

National Ecological Observatory *Staff Scientist* **2013 to 2014**
Implemented a data and metadata standards across a 1,000 data products for an NSF funded ecological obseveratory.
Created programmatic tools for interacting with data in a variety of data platforms (Neo4J, CouchDB, SQL) for other scientists.
Developed and tested algorithms to quantify uncertainty in raw sensor network data and scale up to synthetic data products.

University of British Columbia Beaty Biodiversity Center *Post-Doctoral Fellow* **2011 to 2013**
Developed individual based models of the evolution of sociality in spiders using evolutionary algorithms in Python.
Built large scale simulations of spider populations to calculate error and bias

properties of statistical models of population growth.

University of Vermont Complex Systems Group *Graduate Research Fellow*
2005 to 2011
Built statistical models of toxic algal bloom population dynamics based on historical monitoring data.
Collaborated with a diverse group of engineers, computer scientists and biologists on using complex systems methods to analyze ecological data.

Projects

The Carpentries Foundation
The Carpentries teach foundational coding, and data science skills to researchers worldwide. I teach workshops at universities and scientific institutions on based on the Carpentries curriculum about python, R, data visualization, SQL, and version control. I have also developed numerous lessons, including ones on R, SQL, and spatial data analysis.

rOpenSci Project
rOpensci is community based project that builds open source software tools for the scientific community. I develop and maintain numerous open source software packages for scientific data and computing. I have also been a code reviewer for software submitted to the project, and contributed blog posts, curriculum for workshop,s and helped write grants and academic articles.

DataCamp
Datacamp teaches people data science online in small interactive online lessons. I contribute and maintain curriculum for a lesson called "Case studies in network theory", teaching students the application of basic network theory through a series of video and online exercises.

Education

Postdoctor Research Fellow 2011 to 2013
University of British Columbia, Beaty Biodiversity Research Center, Vancouver, British Columbia

Doctor of Philosophy 2005 to 2011
University of Vermont, Department of Biology, Burlington, VT

Bachelors of Science 1997 to 2001
University of Massachusetts, Department of Biology, Amherst, MA