Daniel L. Klein

web: mesokurtosis.com • github: othercriteria • email: othercriteria@gmail.com • cell: 415.763.0415

Work experience

DMG Blockchain Solutions (Blockseer) **July 2018 – present**
Senior Software Engineer

- Led rewrite of Bitcoin address clustering pipeline (EMR, Scala, GraphX); 1.2 TB blockchain processed in ∼6 hours.
- Developed and validated Bitcoin transaction classification method using custom DeepWalk implementation.

RealScout **October 2016 – November 2017**
Data Scientist

- Led development of Recommended Listing feature: data exploration → agent-agent modeling → architecture plan → design mocks → user stories → implementation alongside engineers → deployment.
- Automated schema mapping for ETL of external listing databases (MLS) using a predictive model (gradient boosting) built on data and metadata features and trained on manually matched schemata.
- Built out Bayesian models for sales price and days-on-market prediction with near-Zillow/Redfin accuracy.

Radius Intelligence, RealScout, Neighborly (rotations) **June – August 2016**
Data Science apprentice (Catenus, 8VC)

TransForm Pharmaceuticals, Inc. **September 2006 – August 2008**
Assistant Scientist, Scientific Computation

- Developed and deployed data integration software for laboratory automation platforms (HPLC, IR, spectrophotometer).
- Rapidly implemented ideas into software to support data analysis, molecular modeling, analytical chemistry method development, platform QA, etc.

Williams College, Dept. of Biology **June – August 2005**
Research Assistant

- Planned and implemented experimental design and data analysis for field research project.

University of Minnesota, Dept. of Ecology and Evolutionary Biology **June – August 2003, June – August 2004**
Research Intern

- Developed and analyzed numerical results from novel model for ecological community assembly model.

Education

Brown University, Ph.D. Applied Mathematics **October 2016 (defended)**

- Advised by Matt Harrison, with research focus on statistical inference in settings of extreme data sparsity or imbalance.
- TA for Intro Stats, Math Stats I/II, Recent Applications in Probability and Statistics.
- Organized and hosted speaker visits and meetings for Pattern Theory seminar.

Williams College, B.A. Mathematics and Biology **June 2006**

- Departmental honors in Biology, with thesis "Understanding aggregation in the membracid *Publilia concava*: using models to disentangle processes".

Languages, stacks, competencies

- Enthusiastic: Python (mypy typing!), R, C, SQL. Grudging: JavaScript, Matlab. Exploring: Rust, Haskell.
- Statistics, numerics, and visualization libraries: numpy/scipy/pandas/scikit-learn, TensorFlow, Stan, PyMC3, D3, etc.
- Familiarity with software engineering tooling (git, build systems, CI, etc.), web development (client-server, HTML, CSS, Hakyll, etc.), and cloud deployment (Docker, AWS, etc.), distributed computing (Hadoop, Spark, etc.).
- Product mindset (quantifying and measuring value creation, growth modeling, UX design, product management).

Relevant coursework

- Foundational computer science, e.g., Data Structures, Design/Analysis of Algorithms, and Programming Languages.
- Bayesian Stats, Biostats, graduate Math Stats I/II, graduate Probability/Stochastics.
- Seminars in Graphical Models, Bayesian Nonparameterics, Network Models.
- Computational Biology, covering use of dynamic programming and approximation algorithms to efficiently learn structural features from messy, incomplete, and/or inconsistent data.