

Trent Hauck

+1 785 304 1391 trent@trenthauck.com

Work Experience

Zymergen zymergen.com
Staff Data Scientist

Oct. 2016 - Present

- Technical lead responsible for developing decision support tools for scientist end-users.
- Led team of 3 DS and 1 DE to develop tooling to recommend genetic edits in bacteria to improve phenotypes for industrial applications. Included the development of a novel unsupervised technique for cold start edits (patent filed).
- Developed Hierarchical Bayesian model to remove batch effects and identify outperforming strains for high throughput screening.
- Augmented in-house sequencing QC pipeline to identify sequencing artifacts and prioritize those for manual review.
- Led development of a large-scale genomics database and retrieval system to support library searches with traditional bioinformatics tooling and in-house ML methods.
- Worked on Data Science team initially, then transferred to Next Generation Sequencing team in 2018.

Consultant

May 2015 - Present

- Embedded with a team at an e-commerce company to develop their first recommendation system – the model based on Latent Dirichlet Allocation, produced item-to-item and item-to-user recommendations which was deployed as a microservice called by the main site.
- Built lifetime value scoring pipeline used for marketing strategy. Pipeline determined current value of incoming lead based on probabilities of purchasing certain policies given lead features, and likely duration of the policy being in-force.
- Consulted with a web analytics consulting firm to improve infrastructure for processing 3rd party APIs for their clients.

Metis thisismetis.com
Instructor

Dec. 2016 - Dec. 2018

- Courses Taught: Introduction to Data Science, Live Online Data Science, Live Online SQL Fundamentals (2x).

Zulily zulily.com
Data Scientist

Feb. 2014 - May 2015

- Built and deployed recommendation services to improve customer experience. Focused on improving user/item recommendations and feature extraction/learning.
- Performed statistical analysis on business operations such as impact of customer repurchases as a function of product delivery duration.
- Provided internal consulting to marketing and accounting department to improve decision making through experimentation. For example, helped accounting estimate total warehouse inventory given a sample count.

Alight Analytics alightanalytics.com
Data Scientist

Aug. 2012 - Feb. 2014

- Developed scalable architecture and ETL system for multi-tenant database in service of BI reporting.
- Led reporting and analysis including content strategy and report implementation for several clients including: Monsanto, AMC, and Cargill Beef.

SelectQuote Senior selectquotesenior.com Workflow Manager

Dec. 2011 - Aug. 2012

- Built the accounting database and surrounding processes for auditing and reconciling of policy income to sales.
- Managed online marketing and site development - including analytics agency and development in support of A/B testing.

Authored Books

- *SciKit-Learn Cookbook*. 2014. Packt Publishing, 2017
- *Instant Data Intensive Apps with pandas How-to*. Packt Publishing, 2013

Selected Talks

- Representation Learning for Graphs and Networks, KC Developer Conference, 2018
- Using an RNN to Make Song Recommendations, Seattle Data Science, 2016
- Low Friction NLP with gensim at PyData Seattle, 2015
- Introduction to Pandas at Python KC User Group, 2013
- Maslow's Hierarchy of Analytics at Big Data KC, 2013

Education

University of Kansas Master of Science, Finance

- Performed research on Support Vector Machines for stock market prediction.
- Graduate Courses in Machine Learning, Data Mining, Econometrics, Probability, and Statistics.
- Teaching Assistant Math Department.

University of Kansas Bachelor of Science, Accountancy

- Elective courses in Linear Algebra, Vector Calculus, Statistics, Programming, and Digital Logic.
- Research Assistant at E&Y CARAT, worked on the FRAANK project applying NLP to SEC filings.

Skill Set Summary

- Broad exposure to ML and statistical based modeling approaches, including:
 - Frequentist and Bayesian Statistics for estimation and inference
 - Machine Learning such as Clustering, Tree-based models, and Dimensionality Reduction
 - Deep Learning using RNNs, CNNs, and GCNs, especially for the generation of discrete structures and transfer learning (NLP). Practical experience scaling models up to hundreds of millions of examples.
- Developer of **GCGC**, a PyTorch toolkit for tokenizing genomic sequences for deep learning applications.
- Writing and maintaining applications:
 - 9 years developing (scientific) applications, most recently in Python, Go, and JavaScript
 - 7 years experience using RDBMSs (Aurora/Postgres), "Big Data" file based stores (S3/Hive), and NoSQL (primarily Mongo)
 - Experience developing batch- (Spark) and stream- (Kafka) data processing systems.
 - Developing and deploying on cloud providers like Azure (1 year), AWS (9 years), GCP (2 years)
 - Best practices using version control, automating CI/CD pipelines, code review, mentoring new developers
- 3 years experience working at the interface of biology, automation, and machine learning
 - Building tools for large scale sequence search via HMM, BLAST, etc
 - Metabolic (and other -omic) network analysis in service of phenotype prediction for industrial applications
 - Building large-scale metagenomic library storage systems for downstream gene and pathway discovery
 - Completed Coursera Specialization: Genomic Data Science (Credential ID: TQTF2FBBTMUQ)