

Summary

Data Scientist and developer with experience in design, implement and evaluate machine learning solutions for both academic projects and industry productions. Top 500 [Kaggle competition expert](#) with medals in natural language processing, computer vision, automatic speech recognition as well as ranking and recommendation problems. Data visualization wizard and caretaker of DevOps chaos monkeys.

Work Experience

IBM Watson Health – Data Scientist / DevOps Engineer Cambridge, MA Apr. 2017

† Research on knowledge graphs to facilitate validation and inference on NLP model results.

† Solution automation with modern DevOps tool chain: Git, Jenkins, Chef and Docker.

† Full-stack developer for internal dashboard to monitoring status of the CICD.

Fidelity Investments – Data Science Intern Smithfield, RI Jun. 2016 - Sept. 2016

† Trained Convolutional Nets for face recognition, Recurrent Nets for voice command recognition for AI lab.

† Topic modeling, sentiment analysis and intent classification on customer phone call transcripts. Deployed solution as web service and used in call center to reduced phone call re-routing rate and support real time monitoring.

Zhenjiang Library – Database Developer Zhenjiang, China Nov. 2010 - Feb. 2015

† Implemented a collaborative filtering book recommendation system; Developed PL/SQL package for analytic use cases.

Projects

† Safe Driver Prediction: Top 2% solution to Kaggle competition on ranking insurance claim holder on their risk of filing a claim for the next year. Constructed a two-layer stacking of GBM, RGF, FM and Neural Nets with Entity Embedding.

† Web Traffic Prediction: Top 10% solution to Kaggle competition on forecasting future web traffic for over 145K Wikipedia articles. Modeled with a blend of 1-D Convolutional Nets and LSTM Recurrent Nets.

† Genetic Mutation Effect Prediction: Top 10% solution to Kaggle competition on predict genetic mutation effect based on literature full texts. Approach with boosting trees trained on features extracted with Doc2Vec, LDA, LSA, NMF.

Education

Bentley University – Master of Science in Business Analytics GPA 4.0 Dec. 2016

† Research Assistant: Worked with Prof. Babaian on developing D3 interactive data visualization for graph dataset.

† Coursework: Programming, Databases, Hadoop, Time Series, Optimization, Statistical Learning, Data Mining

Jiangsu University – Graduate Study in Library and Information Science Jun. 2015

† Thesis: *Recommendations for Research Collaboration Based on Network Model*.

Publication

Tamara B., Ren Z., Wendy L. *DTMi – a New Interface for Informed Navigation*, in proceedings HCI International 2017

Skill Set

† Languages: Python, R, Java, Scala, Matlab, Javascript

† Libraries: Pandas, Numpy, Scipy, Sklearn, Tensorflow, Keras, Networkx, Seaborn, Gensim, NLTK, Node, Angular, D3

† Systems: Linux, Hadoop, Spark, Map-Reduce, Hive, Pig, MongoDB, SQL

† Tools: Shell, Git, Jenkins, Chef, Urban Code, Docker, SPSS, SAS, Gephi, Tableau

Leadership / Community

† Founder of Bentley University R User Group; Python tutor for Bentley Analytic Society; Assistant at Bentley Data Lab