Harrison Zhao

Email: zzhharrison@gmail.com
Website: harrisonzzh.github.io
Cell: (949)463-6067

EDUCATION

- **Cornell University** — New York, NY
  Masters of Engineering in Operations Research; GPA: 3.9 — *May 2018*
  Relevant Courses**:** Applied Machine Learning, Optimization, Natural Language Processing, Computer Vision
- **University of California, Irvine** — Irvine, CA
  Bachelor's in Mathematics; GPA: 3.8 — *March 2017*

SKILLS

Programming: Python (scikit-learn, pandas, numpy, scipy, matplotlib, seaborn, nltk), R, SQL, MATLAB
Machine Learning: LightGBM, XGBoost, PyTorch, Tensorflow, Keras, Spark MLib
Data Science Tools: AWS (Sagemaker, S3, EC2, ECR), Domino Data Lab, Docker, Git
Optimization: CPLEX, Gurobi, CBC, Pulp

EXPERIENCE

- **BAYER** — St. Louis, MO
  Data Scientist — *Jun 2018 – present*
  Built a two-stage (predictive and prescriptive) model for plant field-testing allocation.
  - Developed a recurrent neural network (Bi-LSTM) with PyTorch to predict allocation preference. Used the predicted preference to guide optimization model to the optimal solution and improved performance by 5x.
  - Built a large-scale (160k decision variables) mixed integer optimization model with CPLEX to optimize allocations according to business rules and field-testing objectives.
  - Collaborated with breeding team and field operation team to collect feedback to improve model performance.
  - Delivered allocations of 1.2 million plots across North America. Improved key metrics on maturity matching, environmental diversity by 12%. Automated the entire application through Domino and AWS.

  Implemented a prediction model that uses genetic data to predict plant phenotypes (eg. yield, plant height).
  - Built an ensemble model of gradient boosting machine (LightGBM) and ridge regression. Trained the ensemble model with 250GB+ datasets. Tuned model hyperparameters through AWS Sagemaker.
  - Achieved 0.3 – 0.5 correlations on the major phenotypes, which outperforms production model by 3%.

- **OPTIFUNDER** — St. Louis, MO
  Senior Data Scientist (part-time) — *Dec 2018 – July 2019*
  Designed a smart loan funding model from scratch that provides optimal strategies to warehouse lenders.
  - Worked directly with the founder to understand mortgage business. Built a mixed-integer optimization model that helped warehouse lenders reduce funding cost by 30% - 40%.

- **CORNELL TECH** — New York, NY
  Graduate Research Assistant — *Jan 2017 – May 2017*
  Worked in a team of graduate students to enable natural language navigation for simulated drone.
  - Built a recurrent neural network with self-attention mechanism for language embedding. Enabled the system to better capture key words in the natural language queries. (link)

PROJECTS

- **Simulated Self-Driving Car**, Udacity
  Built an autopilot program to drive in a simulated environment.
  - Implemented a 5-layer convolution neural network with TensorFlow/Keras. Trained with 40,000 driving images on AWS EC2 to predict steering angle.
  - The trained model is able to drive smoothly in a simulated mountain track. (link)

- **Plant Disease Detection IOS App**
  Developed an IOS app that can predict plant disease by taking pictures of plant leaves.
  - Trained a convolution neural network model with 20k images across three different plants. Integrated the trained model into an IOS app that can predict plant disease in real time. (link)

Other projects: Airline Pricing, Image Search Engine, MNIST, House Prices, Sentiment Analysis