

Individual with solid statistical and programming background, seeking an full-time data scientist position

EDUCATION

University of Pennsylvania

Philadelphia, PA

M.S in Statistics

Expected in May 2020

New York University

New York, NY

B.A in Mathematics

Sept 2014 – May 2018

Relevant Courseworks: Machine Learning, Database, Computer Vision, Statistics & Probability, Data Structure and Algorithm

PROFESSIONAL EXPERIENCE

MD. ai | *Intern Data Scientist*

May 2019 – Dec 2019

A fast-growing medical machine learning start-up based in NY that closely collaborates with hospitals and top research institutions

- **Website Analytics:** Built an automatic query pipeline to calculate KPIs for company's website using SQL and created a dashboard on Tableau, enabling real-time metrics tracking and efficient analytics
- Collaborated with the product designer to build a new feature, a 'one-click' bottom for easier medical photo annotation and conducted the AB testing on the UI design of new feature using React and Ruby on Rails, improving the CTR by 8%
- **Ophthalmic photography quality classifier:** Performed the image pre-processing like resizing, smoothing, augmentation, and built a CNN-XGBoost classifier on Pytorch to predict the quality of Ophthalmic image, reaching 0.95 F1 score and implemented in the company's eye treatment system
- **RSNA 2019 Deep Learning Tutor (25k+ Attendees):** Built and presented the whole framework of a segmentation model to detect brain from head CT using Mask RCNN with 0.88 Mean Average Precision

Tencent | *Data Science Intern - IoT Team*

June 2018 – Aug 2018

World largest video game company and the second-largest social media company with \$135B market cap

- **Workplace Energy Analytics:** Built a SARIMA model to forecast the daily energy consumption for each area in the Tencent headquarter, an IoT-enabled smart building, reaching 0.08 in MAPE
- Evaluated recommendations to reduce energy consumption like auto hibernation of equipment, enhancing natural light
- **Elevator Traffic Optimization:** Created a dashboard on PowerBI to monitor elevator operation in Tencent headquarter and track key statistics, conducted a dive-deep analysis, generated a comprehensive traffic report, and modified the operation plan in the peak hour, reducing the average waiting time by 3.2s
- Developed an algorithm to tackle the discontinuity in real-time floor data returned by elevator API based on elevator movement physics simulation, deployed in elevator monitorization system of Tencent headquarter

China Securities | *Quantitative Research Intern - Research & Development Team*

June 2017 – Aug 2017

Leading securities company in China with \$34B total assets

- **Portfolio Optimization:** Constructed a new backtester using python and implemented a simple trading algorithm based on Markowitz Portfolio Optimization theory, backtesting on the historical market data
- **Forecast stock price:** Built an LSTM network using Keras to predict the stock price and fine-tuned parameters like amount of layers, nodes, learning rate, and dropout ratio, improving the RMSE by 43% compared to the current ARIMA model

SELECTED PROJECT

Shanghai Home Price Evaluation Model

New York University

- Collected data with web-scraping on the largest online housing agency in China using python
- Created a feature, distance between the unit and the nearest public transportation by filtering keywords and running K-means on the longitude and latitude of the apartment
- Built a stacked model using MLP, LightGBM and Random Forest regressor to estimate the home price in Shanghai, getting 0.96 R squared score and 0.12 RMSE score

SKILLS

- **Languages:** Python (>10,000 lines), SQL (>5,000 lines), R, Java, SAS
- **ML Toolkits:** Scikit-Learn, PyTorch, TensorFlow, SciPy, NumPy, forecast
- **Big Data Toolkits:** Spark, HDFS, MapReduce, Qubole, AWS, GCP
- **Data Visualization:** Matplotlib, Seaborn, Tableau, Power BI, R Shiny, Plotly, Dash, ggplot