Eric (Binqian) Zeng

403 W 205 St. Apt 52, New York, NY, 10034

+1 929-208-7103 • eric.zeng@nyu.edu

https://www.linkedin.com/in/binqian-zeng-257903126/

Education

New York University, Courant Institute of Mathematical Sciences — **New York, NY**
M.S Data Science; GPA: 3.3/4.0 — *Sep 2016–May 2018*
Relevant Coursework: Machine Learning, Natural Language Processing(Kyunghyun Cho), Deep Learning(Yann LeCun), Statistical and Mathematical Methods, Big Data, Advanced Python, Decision Model and Analytics, Data Science in Quantitative Finance

Sun Yat-sen University, School of Engineering — **Guangzhou, China**
B.E Theoretical and Applied Mechanics (Fluid Dynamics Focus); GPA: 3.7/4.0 — *Sep 2012–Jun 2016*
Honor: Third-class scholarship (three times)
Relevant Course: Numerical Methods, Methods of Mathematical Physics, Linear Algebra, Ordinary Differential Equations, Fluid Dynamics

Technical Skills & Certificates

- **Programming & Scripting Language:** Python, C++, R/Matlab, Scala
- **Toolkits, Softwares & Operating Systems:** Tensorflow, Pytorch, Keras, NLTK, Scikit-learn, Hadoop, Spark, MySQL, MongoDB, AWS(EC2, S3), Github, Linux/Unix
- **Certificates:** C++ Programming for Financial Engineering Certificate(Baruch College, expected in Nov, 2018); CFA Level I Candidate; Reinforcement Learning in Finance(Coursera); Bloomberg Market Concept Certificate(BMC)

Work Experience

King Street Capital Management, L.P. — **New York, NY**
Data Science Intern — *Jul 2018–Present*

∗ **Forecasting Model for Key Performance Indicators of Companies**
- Collaborated with senior researches and traders to implement new statistical or mathematical methodologies as needed for predicting key performance indicators of companies
- Evaluated regression results with self-built multiple evaluation metrics; regression results and error metrics as meta-features for meta-learning and ensemble learning
- Ranked to select most suitable model architectures for different companies automatically with tree-based algorithms
- Built regression model to ensemble and improve predictions from top-level learners with meta-features; handled multicollinearity between features and over-fitting problems by PCA and regularization
- Leverage gradient descent based Meta-Learning technique to make the model fast adapt to new scenarios
- Constructed regression model using the mixture of experts, which ensembles 10 regression architectures, original features, and meta-features

∗ **Analyzing Alpha in Corporate Filings**
- Conducted statistical analysis to capture textual changes in 10-K and 10-Q fillings
- Categorized companies with high and low information ratio with NLP techniques
- Used recurrent neural network based models, optimized word embedding for sentence representations, and attention techniques to improve classification performance

Crypto Investments — **New York, NY**
Machine Learning Engineer Intern — *Sep 2017–Dec 2017*
- Web-scrapped cryptocurrencies related reports, trading price, and volume data with Beautiful Soup in Python; managed data with MongoDB
- Conducted reports sentiment classification with Word2Vec
- Built an event-driven time-series forecasting model to predict trading price and volume for cryptocurrencies based on CNN

IBM — **Armonk, NY**
Data Science Intern, Chief Data Office — *May 2017– Aug 2017*
- Developed an automatic metadata generation pipeline in Python using machine learning annotators
- Managed data in Cloudant NoSQL database; built query index for extraction purpose

Course Projects

Object-oriented Image Deblurring Pipeline — **New York, NY**
Keywords: *Segmentation, Super-Resolution, SRGAN, Tensorflow* — *Mar 2018–May2018*
- Image objects segmentation by Single Shot MultiBox Detector(SSD); image super-resolution reconstruction by SRGAN

Enhanced Seq2Seq Model for Automatic Text Summarization (Capstone Project) — **New York, NY**
Keywords: *Natural Language Processing & Understanding, Hybrid Seq2seq Neural Network, Pytorch* — *Oct 2017–Dec 2017*
- Performed a semantic-encouraged seq2seq model with self-gated encoder, attention mechanism, and semantic measurement term in loss function; achieved high semantic relevance between summaries and source texts (ROUGE-1/2/L: 24.3, 12.3, 33.7)
- Constructed a two-stage hybrid model with Bi-LSTM, coverage mechanism, and probability unit; the model can be viewed as a balance between extractive and abstractive approaches (ROUGE-1/2/L: 38.2, 18.4, 41.1)

Investigation on NYC Crime Data — **New York, NY**
Keywords: *BigData, PySpark, Pattern Recognition, Visualization* — *Feb 2017–May 2017*
- Performed pattern recognition on NYC Crime Data; Clustering to locate spatial centroids of crimes
- Aligned geolocation, demographic, and economic datasets to evaluate correlations between crimes and social factors