Eric (Binqian) Zeng

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

+1 929-208-7103 • bz866@nyu.edu

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

Education

New York University, Courant Institute of Mathematical Sciences — **New York, NY**
Mathematics in Finance; part-time graduate program on-going — *Jan 2019 – Present*
Courseworks: Derivative Securities, Operating System(C/C++)

New York University, Courant Institute of Mathematical Sciences — **New York, NY**
M.S Data Science; GPA: 3.3/4.0; — *Sep 2016 – May 2018*
Courseworks: Machine Learning, Natural Language Processing(Kyunghyun Cho), Deep Learning(Yann LeCun), Statistical and Mathematical Methods, Big Data(Hadoop MapReduce, Spark), 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)
Courseworks: Numerical Methods, Finite element analysis, Ordinary Differential Equations, Fluid Dynamics

Technical Skills & Certificates

- **Programming & Scripting Language:** Python, C/C++, R, Matlab, Scala
- **Toolkits, Softwares & Operating Systems:** Tensorflow, Pytorch, Keras, NLTK, Scikit-learn, AWS, Hadoop, Spark, MySQL, MongoDB, Github, Linux/Unix, Windows

Work Experience

Swiss Reinsurance Company Ltd. (Swiss Re) — **Armonk, NY**
Analytics Specialist/Data Scientist — *Jan 2019–Present*

∗ **Catastrophe Loss Impact Estimation**
- Developed models to estimate liabilities of incoming claims after a catastrophe to benefit financial position and operation
- Conducted liabilities estimation in both aggregated and individual level with insurance coverages information, satellite images data, and buildings footprints geographic data
- Generated geographic features by zonal statistic analysis; Involved image features by convolutional neutral network(CNN)
- Captured time distributed features by CNN; Constructed the estimation regression model by random forest

∗ **Other Projects**
- Identify rental real estates insurance market opportunities for a client with US census, Zillow, and uniform crime report data
- Quantified fire accident risk for underwriting properties of a client with geolocation data and national fire incident reporting data

King Street Capital Management, L.P. — **New York, NY**
Data Science Intern (full-time) — *Jul 2018–Dec 2018*

∗ **Companies KPI Forecasting Model**
- Constructed an ensemble model to forecast KPIs of companies
- All base learners and the top learner are SVMs with different constraints
- Meta-features contains regression results from base learners and evaluations from different metrics
- Handled collinearity in meta-features by matrix factorization; Reduced overfitting by ridge regularization and noise injection

∗ **Analyzing Alpha in Corporate Filings**
- Categorized companies with high and low information ratio by NLP techniques and neural networks
- Optimized GloVe word-embedding by corporate fillings and financial news for sentence representation features
- Captured textual changes in 10-K and 10-Q fillings by statistical analysis as additional features
- Used recurrent neural network based models, and attention techniques to improve classification performance

Crypto Investments — **New York, NY**
Machine Learning Engineer Intern — *Sep 2017–Dec 2017*

∗ **Event-Driven Forecasting Model for Price of Cryptocurrencies**
- Web-scrapped cryptocurrencies news and trade data by API and Beautiful Soup in Python; Data management with MongoDB
- News sentiment analysis by Word2Vec; Event embedding by Open IE; Captured effects of different time spans by CNN

Portfolio

Enhanced Seq2Seq Model for News Text Summarization (Capstone Project in NYU)
Keywords: *Pytorch, Seq2seq Neural Network, Bi-LSTM, Pointer Network, Semantic Relevance*
- The Seq2seq text summarization model is a hybrid of extractive and abstractive approaches trained on CNN/Daily Mail news
- Bi-LSTM encoder with attention mechanism; Pointer network improves the accuracy of words generation and the ability of handling out-of-vocabulary words
- Involved cosine similarity term between encoded source representation and generated summarization representation into the negative log-likelihood loss function to encourage semantic relevance

Commodity Trading Position Forecasting Model
Keywords: *Covariance Matrix, Matrix Factorization, Regularization*
- Created regression models that forecast the optimal trading positions for oil, sugar, copper, gold, and natural gas futures
- Estimated covariance models over rolling windows for asset classes; Explained the variance asset class by top alpha factors
- Applied SVD on alpha factors to filter down signals and forecast returns; Reduced overfitting by elastic net regularization