Eric (Binqian) Zeng

Master of Science in Data Science at New York University (Expected to graduate in May 2018)

+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**
M.S Data Science; GPA:3.3/4.0 — *Sep 2016–May 2018(Expected)*
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: Computational Methods, Methods of Mathematical Physics, Optimization and Computational Linear Algebra, Ordinary Differential Equations,

Technical Skills & Certificates

- **Programming & Scripting Language:** Python, R/Matlab, Java, Fortran, Scala
- **Toolkits, Softwares & Operating Systems:** Hadoop, MapReduce, Spark, MySQL, MongoDB, AWS(EC2, S3), Tableau, D3.js, OpenRefine, Excel, Crystal Ball, Github, Linux/Unix
- **Certificates:** Bloomberg Market Concept(BMC); Preparing for CFA Level I Exam ? June 2018

Work Experience

Crypto Investments — **New York, NY**
Software Engineer Intern (Machine Learning Focus) — *Sep 2017–Dec 2017*
- Scrapped reports, price, and volume data of 8 kinds of cryptocurrencies from 20 websites with BeautifulSoup
- Constructed data sets from scrapping with MongoDB; built a dashboard to visualize price and volume with Matplotlib
- Performed sentiment analysis model with FastText
- Constructed a hybridization of time-series analysis neural network for technical trade including ARIMA and Deep Belief Network

IBM — **Armonk, NY**
Data Science Intern in Chief Data Office — *May 2017– Sep 2017*
- Participated in constructing a pipeline to automatically extract metadata from unstructured documents
- Built Named-Entity Recognition model with Linear SVM; achieved an accuracy of 94%, which is competitive with Watson Natural Language Classifier's accuracy of 97% under 70% coverage

China Guangfa Bank — **Guangzhou, China**
Data Analyst Intern — *Dec 2015–Feb 2016*
- Built propensity models with Linear Regression and Logistic Regression in R based on over 100,000 customers information extracted by SQL

Course Projects

Image Super-Resolution Methods with Neural Network — **New York, NY**
Keywords: *Segmentation, Super-Resolution, SRDenseNet, SRGAN* — *Mar 2018–Present*
- Built a super-resolution pipeline to reconstructs a higher-resolution image from observation; leveraged SRDenseNet and 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; achieved high semantic relevance between summaries and source texts (ROUGE-1/2/L: 24.3, 12.3, 33.7)
- Constructed a two-stage hybrid seq2seq bi-directional Recurrent Neural Network with GRU, 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)

Automated Scoring System for Essay — **New York, NY**
Keywords: *Natural Language Processing, LSTM, CNN, Attention Mechanism, Pytorch, Keras* — *Oct 2017–Dec 2017*
- Conducted research on 8 widely-used automated essay scoring models from research paper in Pytorch and Keras
- Investigated effects of mechanisms and architectures in networks, including LSTM, Bi-LSTM, CNN, attention mechanism, pooling functions, etc.

Automatic Music Genre Classification System — **New York, NY**
Keywords: *Machine Learning, Multi-label Classification* — *Feb 2017–May 2017*
- Built multi-label prediction models with Random Forest and SVM (F-score: 0.303)
- Improved performance with Recurrent Neural Network, Convolutional Neural Network, and Gated Recurrent Unit (F-score: 0.458)

Investigation on New York Crime Open Data — **New York, NY**
Keywords: *BigData, Cloud Platform, Clustering, Feature Extraction, Visualization* — *Feb 2017–May 2017*
- Performed data cleansing and normalization using SQL
- Used PySpark to detected patterns with techniques like K-means and SVD on AWS EC2 and S3
- Produced data visualization on identified patterns with Matplotlib in Python, Tableau and D3.js