Binqian (Eric) 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, Big Data, Natural Language Processing, Deep Learning, Statistical and Mathematical Methods, 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; GPA: 3.7/4.0 — *Sep 2012–Jun 2016*
Honor: Third-class scholarship (three times);

Technical Skills & Certificates

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

Work Experience

Crypto Investments — **New York, NY**
Software Engineer Intern — *Sep 2017–Present*
- Scrapped text and price data from 20 websites; requested APIs to build the real-time dashboard for price and volume.
- Data management with MongoDB.
- Performed sentiment analysis model with FastText.
- Constructed the time-series analysis neural network for technical trade that considers historical prices, price patterns and sentiments of text.

IBM — **Armonk, NY**
Data Scientist Intern in Chief Data Office — *May 2017– Sep 2017*
- Participated in constructing a pipeline to automatically extract metadata from unstructured documents.
- Queried IBM Cloundant and GSA databases to extract text data.
- Built Named-Entity Recognition model with Linear SVM; successfully 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, internship — *Dec 2015–Feb 2016*
- Implemented regression techniques and distribution property on customer data with R to address missing information.
- Extracted information for over 100 thousand customers with SQL; predicted customers' propensity utilizing Linear Regression and Logistic Regression.

Professional Experience

Enhanced Seq2seq Model for Automatic Text Summarization — **New York, NY**
Capstone Project — *Oct 2017–Dec 2017*
- Performed a seq2seq semantic-encouraged abstractive text summarization model with attention mechanism. Introduced semantic measurement term into the loss function to encourage high semantic relevance between generated summaries and references. (ROUGE-1/2/L: 24.3, 12.3, 33.7)
- Constructed a seq2seq two-stage hybrid summarization model with coverage mechanism and probability unit. The first stage is a sentence extractor with a two-layer bi-directional GRU RNN. The second stage is a seq2seq network with coverage mechanism and a probability unit. Probability unit controls the decision to either copy word from input article by sampling from attention distribution or generate word from vocabulary. (ROUGE-1/2/L: 38.2, 18.4, 41.1)

Automated Scoring System for Essay — **New York, NY**
Natural Language Processing Project — *Oct 2017–Dec 2017*
- Re-implemented 8 widely-used automated essay scoring models and investigated effects of different mechanisms, constructions of networks, and non-linear functions, including LSTM, Bi-LSTM, CNN, attention mechanism, and mean-over-time function.

Automatic Music Genre Classification System — **New York, NY**
Machine Learning Project — *Feb 2017–May 2017*
- Utilized binary relevance method and Logistic Regression as the baseline model. (F-score: 0.303)
- Built a multi-label SVM model for multi-label prediction. (F-score: 0.218)
- Improved performance with Recurrent Neural Network and Convolutional Neural Network. (F-score: 0.458)

Investigation on New York Crime Open Data — **New York, NY**
Big Data Project — *Feb 2017–May 2017*
- Used PySpark to find patterns with K-means and SVD; evaluated relevance by ANOVA-test.
- Performed data cleansing and normalization using SQL.
- Produced data visualization on identified patterns with Matplotlib in Python and Tableau.