

Binqian Zeng | Resume

M.S. Data Science Candidate in New York University (Expected to graduate in May 2018)

+1 929-208-7103 • bz866@nyu.edu

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

Technical Skills

- **Programming & Scripting Language:** Python, Java, R, Fortran, Matlab • **Version Control:** GitHub
- **Toolkits & Softwares:** Hadoop, MapReduce, Spark, MySQL, MongoDB, Tableau, OpenRefine, Excel, Crystal Ball
- **Libraries:** Pandas, Numpy, Scipy, IPython, scikit-learn, Tensorflow, PyTorch, BeautifulSoup, NLTK, Gensim, Matplotlib, seaborn, bqplot, etc. • **Certifications:** Preparing for CFA Level I Exam – June 2018.
- **Projects in progress:** Hybrid Abstractive-Extractive Text Summarization, Automated Text Scoring system;

Work Experience

- **Crypto Investments** **New York, NY**
Aug 2017–Present
Software Engineer, internship
 - Scrapped text and price data and requested APIs to build the real-time chart. Data management with MongoDB.
 - Built the sentiment analysis model with FastText. Built the time-series analysis model using RNN and LSTM which consider historical prices, price patterns and sentiments of text.
 - Developed a technical analysis and recommendation tool for cryptocurrency investments.(In progress)
- **IBM** **Armonk, NY**
May 2017– Sep 2017
Data Scientist, internship in Chief Data Office
 - Participated in constructing a pipeline to automatically extract metadata from unstructured documents.
 - Query Cloudbant and GSA databases to extract data.
 - Built Named-Entity Recognition model with Linear SVM. My model's accuracy is 94% while the Watson Natural Language Classifier's accuracy is 97% under 70% coverage.
- **China Guangfa Bank** **Guangzhou, China**
Dec 2015–Feb 2016
Data Analyst, internship
 - Handled missing customers' information with R using regression techniques and distribution property.
 - Extracted information by SQL and predicted customers' propensity by applying Linear Regression and Logistic Regression.

Professional Experience

- **Text Summarization** **New York, NY**
Oct 2017–Present
NYU Term Project
 - Built the extractive baseline model with TextRank and LexRank and built the abstractive baseline model using sequence-to-sequence RNNs.
 - Experimenting improved techniques to augment neural network such as attention mechanism and figuring out the best way to ensemble extractive and abstractive models. (In progress.)
- **Automated Scoring System** **New York, NY**
Oct 2017–Present
NYU Term Project
 - Extracted 13 features for each essay to predict grade using linear regression as baseline model. Evaluated the model with 5-fold cross-validation and Quadratic Weighted Kappa (0.73).
 - Experimented using LSTM, doc2vec, and word2vec to represent the meaning of texts and build a more competitive model. Experimented using CNN and LSTM ensembles to get improvement. (In progress)
- **Automatic Music Genre Classification System** **New York, NY**
Feb 2017–May 2017
NYU Term Project
 - Used the binary relevance method and Logistic Regression as the baseline model. (F-score: 0.303)
 - Reformulated labels for multi-label prediction and used Gradient Descent to minimize hinge loss of multi-label SVM model. The poor performance is caused by correlations between labels, like 'Rap' and 'Hip-hop'(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**
Feb 2017–May 2017
NYU Term Project
 - Used PySpark to find patterns with K-means and SVD and evaluated relevance by ANOVA-test.
 - Used SQL to finish data cleansing and normalization. Visualized patterns with Matplotlib and Tableau.

Education

- **New York University, Courant Institute of Mathematical Sciences** **New York, NY**
Sep 2016–Present
M.S Data Science Candidate (Expected to graduate in May 2018)
Courses: Machine Learning; Big Data; Natural Language Processing; Deep Learning; Statistical and Mathematical Methods; Programming for Data Science; Decision Model and Analytics; Data Science in Quantitative Finance; Computational Method for Finance*; Algorithm*; Time Series Analysis & Statistical Arbitrage*.
- **Sun Yat-sen University, School of Engineering** **Guangzhou, China**
Sep 2012–Jun 2016
Bachelor of Engineering; Major in Theoretical and Applied Mechanics

Courses with '*' are self-study courses .