

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 and Scripting Language:** Python, Java, R, Fortran, Matlab • **Version Control:** GitHub
- **Big Data Tools:** Hadoop, MapReduce, Spark, SQL, AWS • **Softwares:** 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:** Text Summarization, Automated Scoring system;

Work Experience

- **Crypto Investments** **New York, NY**
Aug 2017–Present
 - Scrapped news, reports and historical price data from websites. Data management with MongoDB.
 - Requested API to get price and volume data. Wrote script to build real-time chart.
 - Implemented sentiment analysis on text data with Recurrent Neural Network and time series forecasting on historical price data.(In progress)
 - Developed a technical analysis and recommendation tool for cryptocurrency investments.(In progress)
- **IBM** **Armonk, NY**
May 2017– Sep 2017
 - 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.
 - Enriched and visualized relations between entities using knowledge graph. The relation ranking algorithm is based on Bayes theorem, mainly considered Entity Prior, Entity Affinity and Relationship Strength.
- **China Guangfa Bank** **Guangzhou, China**
Dec 2015–Feb 2016
 - 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.
- **Uber Technologies Inc, Guangzhou Branch** **Guangzhou, China**
Sep 2015–Dec 2015
 - Optimized fleet management in peak time with linear programming method in Matlab.

Professional Experience

- **Automatic Music Genre Classification System** **New York, NY**
Feb 2017–May 2017
 - Used the binary relevance method of multilabel classification as baseline model. Trained Naive Bayes, Logistic Regression, and SVM models to fit each label with a one vs rest classifier and used F-score to evaluate performance. (F-score: 0.303)
 - Reformulated labels for multi-label prediction and used Gradient Descent to minimize hinge loss of multi-label SVM model and compared its performance with baseline models.(F-score: 0.218)
 - Improved performance of classification with Recurrent Neural Network and Convolutional Neural Network. (F-score: 0.458)
- **Exploration on New York Crime Open Data Based on PolyGamy Thoughts** **New York, NY**
Feb 2017–May 2017
 - Stored New York crime data into Hadoop Distributed File System.
 - Used SQL to finish data cleansing and normalization in NYU Dumbo HPC cluster.
 - Used PySpark to find patterns and evaluated relevance by ANOVA-test.
 - Visualized patterns with Matplotlib.

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: Numerical Methods; Linear Algebra; Computational Fluid Dynamics; Ordinary Differential Equation;

Courses with '*' are self-study courses .