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 Text Scoring system;

Work Experience

Crypto Investments — **New York, NY**
Software Engineer, internship — *Aug 2017–Present*
- Scrapped text and price data and requested API to build real-time chart. Data management with MongoDB.
- Built sentiment analysis model with FastText. Time series analysis using RNN and LSTM with price and text as inputs.
- Developed a technical analysis and recommendation tool for cryptocurrency investments.(In progress)

IBM — **Armonk, NY**
Data Scientist, internship in Chief Data Office — *May 2017– Sep 2017*
- Participated in constructing a pipeline to automatically extract metadata from unstructured documents.
- Query Cloundant 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**
Data Analyst, internship — *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.

Professional Experience

Text Summarization — **New York, NY**
NYU Term Project — *Oct 2017–Present*
- Built extractive baseline model with TextRank and LexRank and built 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 model. (In progress.)

Automated Scoring System — **New York, NY**
NYU Term Project — *Oct 2017–Present*
- Extracted 13 features for each essay to predict grade using linear regression as baseline model. Evaluated 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**
NYU Term Project — *Feb 2017–May 2017*
- Used the binary relevance method and Logistic Regression as 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 correlation 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**
NYU Term Project — *Feb 2017–May 2017*
- 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**
M.S Data Science Candidate (Expected to graduate in May 2018) — *Sep 2016–Present*
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**
Bachelor of Engineering; Major in Theoretical and Applied Mechanics — *Sep 2012–Jun 2016*
Courses: Numerical Methods; Linear Algebra; Computational Fluid Dynamics; Ordinary Differential Equation;

Courses with '*' are self-study courses .