Binqian Zeng | Resume

M.S. Data Science Candidate in New York University

+1 929-208-7103 • bz866@nyu.edu
https://www.linkedin.com/in/binqian-zeng-257903126/

Skills

- Python, R, SQL, HTML, Crystal Ball, Spark, Hadoop, Fortran, Microsoft Excel, OpenRefine, LaTeX, Machine Learning, scikit-learn, Natural Language Processing, Gensim, NLTK, TextBlob, Data Analysis, Statistics, Monte Carlo simulation

Work Experience

IBM, Global Chief Data Office — **Armonk, NY**
Data Scientist, internship — *May 2017–Present*

- Built the entity type system and word dictionaries with Natural Language processing techniques, including K-means for clustering, Tf-Idf for extracting most important words and LDA Topic Model for detecting topics of unstructured documents.
- Built baseline model for extracting terms from sentences with Naive Bayes, Logistic Regression, and SVM models in scikit-learn and improved the performance with Topic Modeling and Word Embedding techniques in Gensim.
- Built machine learning model based on Bayes Theorem to detect and enrich relationships between entities, mainly considered Entity Prior, Entity Affinity and Relationship Strength.
- Visualized Entities and their relations using Knowledge Graph.

China Guangfa Bank — **Guangzhou, China**
Data Analyst on custom behavior, internship — *Dec 2015–Feb 2016*

- Conducted data cleaning and handled missing customers' information with R using Regression techniques and Distribution Property
- Extracted basic information using SQL and predicted customers' propensity of buying using Linear Regression and Logistic Regression.
- Visualized statistic finding using R modules including Plotly and ggplot.

Uber Technologies Inc, Guangzhou Branch — **Guangzhou, China**
Operation Assistant, internship — *Sep 2015–Dec 2016*

- Optimized fleet management in peak time with linear programming method in Matlab.

Professional Experience

Automatic Music Genre Classification System — **New York, NY**
NYU Term Project — *Feb 2017–May 2017*

- Used the binary relevance method of multilabel classification as baseline model. Trained Naive Bayes, Logistic Regression, and SVM models to fitted each label with a one vs rest classifier and used F-score to evaluate performance.
- Reformulated labels for multi-label prediction and used Gradient Descent to minimize hinge loss of multi-label SVM model. Compared its performance with baseline models.
- Improved performance of classification with Convolutional Neural Network by using multiple Convolutions across word embeddings and alternate layers with pooling layers. The final hidden activation is then fed to a logistic layer to predict the labels.

Exploration on New York Crime Open Data Based on PolyGamy Thoughts — **New York, NY**
NYU Term Project — *Feb 2017–May 2017*

- Stored New York crime data into HDFS and read data using PySpark and SQL in NYU Dumbo HPC cluster.
- Used PySpark and SQL to look for data quality issues, including missing data and invalid values and conducted data cleaning.
- Used PySpark to find patterns and evaluated relevance by testing if the Pearson Correlation or ANOVA-test value is significantly different from zero and visualized result using Pyplot.

Spam Email Detection System — **New York, NY**
NYU Term Project — *Oct 2016–Dec 2016*

- Conducted Data cleaning and feature selection using Python modules including Numpy and Pandas.
- Processed raw data of email text into word tokens based on Count Vectorizer, TF-IDF in the ScikitLearn module.
- Built the models with Naive Bayes, Decision Tree, Logistic Regression and Random Forest.
- Visualized the evaluation result using Python visualization modules, including Matplotlib and Pyplot.

Education

New York University, Courant Institute of Mathematical Sciences — **New York, NY**
M.S Data Science Candidate — *Sep 2016–Present*

Courses: Machine Learning; Big Data; Text As Data; Natural Language Processing; Decision Model and Analytics; Programming for Data Science; Statistical and Mathematical Methods.

Sun Yat-sen University, School of Engineering — **Guangzhou, China**
Bachelor of Engineering; Major in Theoretical and Applied Mechanics — *Sep 2012–Jun 2016*