

Zhaopeng Liu

M.S in Data Science - Courant Institute of Mathematical Sciences - NYU

• zL1732@nyu.edu • 646-912-3080

SKILLS

Language: Proficient: Python, R, SAS, MySQL, SQL, Matlab **Intermediate:** Tableau, RShiny, C, C++, MapReduce
Software: Jupyter Notebook, Spark(PySpark), Hadoop, Microsoft Excel, MongoDB, Eclipse, Latex

EDUCATION

Master of Science in Data Science, New York University

New York, NY

May 2018, First-Year

- Courses: Machine Learning for Data Science, Risk & Portfolio Management, Programming for Data Science(Python), Advanced Python programming, Big Data(Hadoop,Spark)

Bachelor of Science in Statistics, Minor in CS/Math, University of Illinois at Urbana-Champaign

Champaign, IL

May 2016, Completed

GPA: 3.80

Major: 3.96

- Courses: Data Mining, Data Structure, Stochastic Process, Advanced Data Analysis(SAS), Statistical Data Management(SAS), Programming with Java, Categorical Data Analysis
 - Dean's List, Graduate with Highest Distinction, Tau Sigma National Honor Society
-

WORK EXPERIENCE

Shotaki Inc: Data Scientist, Intern

New York, NY

Sep 2016 – Dec 2016

- Built Countvectorizer-based classifier of emotions for unlabeled online comments classification, measured how popular is a product and assist with decision making
- Visualized sale prediction with Tableau (BI Dashboard tool) to get a straightforward view of products' sale trends by geography, therefore to help the company adjust strategy in time

Medici Quantitative Investment: Algorithm Trading Programmer, Intern

Beijing, China

Jun 2016 – Aug 2016

- Loaded the real-time options market trading data on tick level and stored them in MongoDB
- Created a prototype of Chinese options trading on tick level and implemented it with python, tested the program on copper trading using windows server and profited 2% in two weeks
- Created performance report panel including Shapiro, return indexes for Chinese ETFs market with Shiny on Rstudio to help customers choose their preferred ETFs with a clearer view

Agricultural Bank of China: Financial Report Analyst, Intern

Beijing, China

Jun 2015 – Aug 2015

- Validated balance sheet with Microsoft Excel to ensure they are correct and up-to-date
 - Conducted data processing, including handling missing data and incorrect format
-

RESEARCH EXPERIENCE

Scikit-Learn Module(Partial) Implementation in Python(In Progress): Long-Term Research

New York, NY

Feb 2017 – Present

- Implemented hard / soft margin SVM in Pegasos algorithm and tested it on "Spam Email Project"
- Implemented kernelized ridge regression and kernel machine with RBF, Polynomial and linear kernels
- Implemented linear regression based on gradient descent method and stochastic gradient descent method
- Implemented lasso regression based on coordinate descent method

New York Crime Analysis with PySpark/MapReduce(In Progress): NYU Big Data Course Final Project

New York, NY

Feb 2017 – Present

- Analyzed New York City crime data set on the computing cluster of New York University
- Wrote PySpark programs to validate data format and to check outliers
- Wrote PySpark programs to discover crime trend over time in different locations
- Developed MapReduce programs to compute the statistics of each variable

Spam Email Detection System: NYU Data Science Course Final Project

New York, NY

Oct 2016 – Dec 2016

- Parsed raw email data in text format to word tokens using Count Vectorizer in Scikit-Learn module.
- Built models with ML algorithm Naïve Bayes, Decision Tree, Logistic Regression and Random Forest
- Applied penalty factors and created evaluation metrics that focus on minimizing false positive rate
- Performed feature, parameter selection to improve the model, the final model has AUC=0.95 on test data

1st Prize, Contemporary Undergraduate Mathematical Contest in Modeling: Team Leader

Beijing, China

Sep 2013 – May 2014

- Developed fast dynamic programming algorithm for restoring straight-cut and cross-cut pdf files.
- Converted chopped pieces into Grayscale picture and matrix with each element between 0-255
- Won 1st Prize in Beijing, among more than 5000 participants