Liwei Song

70 Christopher Columbus Dr. Apt 3702, Jersey City, 07302• (217)9798081• ls4408@nyu.edu

Educations:

New York University Expected graduation date: May 2018
Master of Science in Data Science cumulative GPA: 3.9/4.0
University of Illinois at Urbana-Champaign Graduation date: December 2015
Bachelor of Arts in Economics and Statistics (Double Major) overall GPA: 3.6/4.0
Bachelor of Science in Material Science and Engineering (Electrical Material)

Skills:

***Programming*:** Python: Proficient in Numpy, Pandas, Matplotlib, Sqlite3 and Sklearn packages. Good understandings of OOPS.
Machine Learning: KNN, K-means, hierarchy clustering, SVM, Random forest, Decision tree, logistic regression, Text-mining etc.
Statistics and mathematics programming: SAS, R, STATA (Basic)
Big data technology and version control: Git, Linux, SQL, Hadoop, PySpark, AWS cloud.
***Statistics and mathematics*:** Fundamental statistics: Probability theory, Random process, Simulation, Hypothesis testing, Bayesian statistics, Quantitative analysis in Econometrics, Statistical Modeling, Machine Learning
Economics: Intermediate Macroeconomics, Intermediate Microeconomics, Game theory, Govt Reg of Economic Activity

Honors:

- Earl J. Eckel Scholarship (May 2013 and May 2014)
- Dean's list (May 2012 and Dec 2012)
- Graduate with highest distinction in Statistics (Dec 2015)
- Graduate with honors in Material Science& Engineering (Dec 2015)

Internship Experience:

New York City Department of City Planning New York City, NY
Traffic intern *Mar 2017—*

- Analyze traffic data of different city rezoning and planning projects for use by project managers in environment assessment and review division.
- Currently designing an automation process to do travel demand forecast with the proposed action in NYC in python referencing to CEQR technical manual and U.S. census tract data.
- Doing geospatial analysis with ArcGIS in collaboration with colleagues.

Hubei Provincial Water Resources Research Institute. Wuhan, China
Assistant to Information and Technology engineer, Zhiguo Wang May 2016 - Aug 2016

- Assisted in Managing Hubei provincial Lake and Reservoirs Management System, including writing data collecting instructions and user guide for reservoir admin staffs.
- Modified the core parameters in the system and data clean for data sets across 14 municipalities.
- Monitored daily precipitations and water levels of reservoirs in the system, and retrieved data of reservoirs on alert by SQL or Excel for use by water resource engineers during the flood season with preliminary data analysis.

Research& Projects

Spam Email detection(Python) New York, NY
Associate team leader Oct 2016 – Dec.2016

- Cleaned data and filtered contents from 5GB emails with use of python. (including regular expression and codecs etc.)
- Text feature extraction using Count vectorizer, Tf-idf and Word2Vec, and feature selection using Chi-2 and naïve reduction.
- Choose the best model among methods of Naïve Bayes, random forest, decision tree and logistic regression, and choose optimal parameter using grad search, with over 0.95 AUC.

Visual exploration of world economy and demographics (Python) New York, NY
Team Leader Oct 2016 – Dec. 2016

- Achieved the goal of designing a tool for different data visualization methods of world bank data by use of python in a team of three with proficient use of Pandas and Numpy.
- Defined classes and functions for the system, and designed a user interface in terminal, which allowed users to interactively and flexibly control their inputs.
- Integrated data visualization classes and functions (achieved by Matplotlib and Plotly) into the system, and ran unittest.

Data Analysis in Computational Neuroscience Lab(Python) Champaign, IL
Research assistant under Prof. Robert Wickesburg and Dr. Christopher L. Boven. Oct 2015 - Dec. 2015

- Used python to create Sqlite3 readable CSV files from raw txt files of unstructured data.
- Used Matplotlib package to achieve spike raster plots for all neural firing experimental trials.
- Constructed SQL database of neural firing data with size over 100 MB and queried data with Sqlite3 packages.
- Applied non-parametric method to estimate instantaneous neural firing rate under the instruction of the mentor.