Fangshu Lin

(646) 923-3637 | fl1210@nyu.edu | New York, NY | https://www.linkedin.com/in/fangshulin/

Skills

Programming: Python, R, SQL, MATLAB, Bash **Tools:** Hadoop, Spark, Hive, ArcGIS, D3.js, Tableau, AWS
Techniques: Regression, Reduction, KNN, K-Means, Neural Network, Decision Tree, Random Forest, XGBoost, Time Series

Education

MS in Urban Informatics, New York University, New York, NY *Sept 2018*

PhD in Civil Engineering, Tongji University, Shanghai, China *Dec 2015*

- *PhD research: Intelligent Systems, joint PhD at Purdue University*

BS in Civil Engineering, Tongji University, Shanghai, China *July 2009*

Relevant Coursework

Applied data science, Machine learning, Big data analytics, Data visualization, Spatial analytics, Text as data, Probability and Statistics, Linear Algebra, Introduction to Database, Optimization, Stochastic Processes

Experience

Cityglobe *New York, NY*
Data Science Intern *Apr 2018 – Sept 2018*

- Performed data quality assessment; applied ML methods for imputation of economic indicators and demographic data.
- Conducted clustering analysis and PCA on growth in 50+ industries of 545 top urban economies.
- Designed methodology with evaluated ML models to predict best store locations based on demographic data and business density in NYC with Google places API, Python sklearn.

NYU Urban Observatory & Arcadis *New York, NY*
Graduate Research Assistant *May 2018 – July 2018*

- Applied image processing using multi-dimensional WorldView-2 satellite image data of 18 sites along Chicago River.
- Completed unsupervised segmentation based on clustering to examine vegetation health using satellite images with GDAL, NumPy, SciPy, scikit-learn.

Advanced Research in Government Operations *New York, NY*
Civic Data Marketplace Intern *Dec 2017 – Jan 2018*

- Developed a program (R Shiny) to quickly estimate water bills applicable to 200+ water agencies across California under the California Data Collaborative ("CaDC") project.
- Tool allows users to put in address, mapping to agency with address, calculate and visualize water bill/usage by tiers.
- Hosted the bill calculator tool using AWS EC2 server.

Purdue University IISL Lab *West Lafayette, IN*
Visiting Researcher *Aug 2011 – Aug 2013*

- Processed and analyzed 100+ GB of high dimensional noisy sensor data in time and frequency domain for statistical analysis and model updating.
- Developed a real-time hybrid simulation platform for structural dynamic testing using host-target PC solution.
- Developed a new semi-active control algorithm to optimize coupled building vibration under earthquake input.

Selected Projects

Hadoop based stock prediction using sentiment analysis in Chinese A-share market

- Performed ETL and profiling on Sina Weibo feeds and stock prices data using Hadoop MapReduce and HDFS.
- Trained Random Forest model based on stock prices and sentiment score using PySpark to predict stock movement.
- Results show that by incorporating public sentiment information can enhance prediction of stock movement (by 6%).

Vulnerability analysis for transportation networks *(Sponsor: Lockheed Martin Advanced Technology Lab)*

- Built a directed, weighted network with 390 nodes and 2274 edges to model London subway system using over 600,000 trip records of Oyster card with Python NetworkX, Pandas, GeoPandas.
- Created two metrics to measure impacts of different level of disruptions and identified most vulnerable stations.
- Discovered hidden pairs of stations that have huge impacts when shutting down simultaneously considering topology.