

# Xiaoxiao (Monica) Yan

917-900-2474 | [xy2357@columbia.edu](mailto:xy2357@columbia.edu) | New York, NY, 10025 | <https://github.com/monicayan>

## SUMMARY

---

Comprehensive background in machine learning and data analysis; Conversant with statistical inference and mathematical modeling; Proven ability to work both independently and cooperatively with multiple task under the time constraints.

Personal Website: [https://monicayan.github.io/personal\\_website/monicayan.html](https://monicayan.github.io/personal_website/monicayan.html)

## EDUCATION

---

**Columbia University**, Fu Foundation School of Engineering and Applied Science

New York, NY

M. S. in Data Science (Accumulative GPA: 3.63/4.0)

Expected May 2019

*Coursework: Algorithms, Machine Learning, Probability Theory, Statistical Inference, Computer Systems, Deep learning, Reinforcement Learning*

**Southeast University** (Honor Student of the Year)

Nanjing, China

B.S in Materials Science and Engineering (Major GPA: 3.83/4.0)

June 2017

## SKILLS

---

Machine Learning, Deep Learning, Reinforcement Learning, Algorithms, Data Analysis, Database Management;

Python, SQL, R, Bash, C++, D3, Matlab, Java; AWS, Hadoop, Spark, Hive SQL, MongoDB, Tableau, LaTeX; Linux, Unix.

## PROJECTS

---

### Entity Resolution as Service (Capstone Project)

- Set up entire application pipeline for entity resolution problem, including data transfer (Kafka), data storage (MongDB), domain-independent blocking strategy, and matching algorithms (word embedding and random forest regression).

### Deep Reinforcement Learning Assisted Trading (<https://github.com/monicayan/Deep-Reinforcement-Trading>)

- Extracting financial features from historical data; using deep-Q-network and policy gradient method to guide trading policy;

### MTA Subway System Optimization (<https://github.com/monicayan/MTA-Analysis>)

- Scraping raw data from website; after data cleaning, commit exploration and visualization on passenger flow on each gate of NYC subway system; give possible optimization based on the results.

### Text to Images (<https://github.com/monicayan/StackGAN-Text-to-Images>)

- Trying to use stackedGANs structure to generate high-resolution pictures according to user inputs (based on TensorFlow);

### Amazon Online Products Recommendation Analysis

- Analyzed the likelihood of the customers' purchasing different products together, summary stats; currently try to use the data build a product-based recommendation system using logistic regression model.

### Stan Development Team

- Building interface between CmdStan and Python, and PyCmdStan and ArviZ;

### NBA Data Analysis

- Built the algorithm to compute every player's performance in each game;
- Analyzed viewers from different countries and predicted the total international viewers in the following games.

### Amazon Online Products Recommendation Analysis

- Extracted data from raw user record and products categories data, conducted data cleaning and output as CSV file;
- Built data structures that easy to search, update for shopping orders and product categories path respectively;
- Used MLE to compute the probability purchasing different products together and the lift between different products.

## Selected Assignments

- Built word embedding and RNN structures to: 1) classification sentences from different books; 2) generate color and RGB values according to input color names; 3) generate color name and RGB values according to input color. (based on TensorFlow)
- Built a user audio recommendation system using PySpark and MLlib;
- Set up AutoML-style image classifier with CNN structure to classify different flower varieties in the pictures (based on TensorFlow);
- Built various regression models to predict wine quality according to 28 features; compare the differences between each model;
- Analyzed the influence of insurance level on times of medical consultations using generalized linear model (turncPoisson, Exponential).

## EXPERIENCE

---

**U.S. Science Support Program** IT Assistant 05/18 - present *New York, US*

- Wrote Python scraping scripts to get raw data from website;
- Cleaned the data of academic activities, funding applicants from the beginning of the program;
- Built relational database that easy to search, update and delete;
- Built both back-end and front-end platform of database according to manager's needs;
- Managing the database as daily routine, developing web-front end as a search engine linked to the database.

**Southeast University** Software Developer 06/14 – 09/14 *Nanjing, China*

- Built a database for country-wide train line schedule, output as Excel file;
- Developed a software for ticketing system based on C++, with selling, searching and returning functions;
- Conducted functional testing, usability testing and performance testing to the software.

**Bauman Moscow State Technical University** Quantitative Researcher 10/16 – 11/16 *Kazan, Russia*

- Investigated and analyzed data of the global market in clean energy (solar energy, Li-ion batteries and fuel cells);
- Analyzed most important features that result to high market share in clean energy field;
- Forecasted future market growth in clean energy field by understanding data pattern and visualized results.

**National Key R&D Laboratory** Quantitative / Computational Researcher 10/14 – 07/17 *Nanjing, China*

- Learned algorithms based on DFT (Density Functional Theory) and conducted computation on properties of Cu atoms;
- Explored in the synthesis method of Si-based anodes for Li-ion batteries; figured out how different parameters infect the synthesis by fitting equations; concluded on optimal synthesis parameters (published an article and owned a patent on it);
- Collected testing data on catalysts for fuel cells with different components and ratios, inferred the relation function between performance and components percentage, and finally concluded on optimal components ratio.

## SELECTED HONOURS

---

<i>Feb. 2016-Jun. 2016</i>	Guanghua Scholarship, Southeast University
<i>Aug. 2015-Jun. 2016</i>	Honor Student of the year, Southeast University
<i>Mar. 2016</i>	2rd Prize, Energy Saving Competition, Southeast University
<i>Aug. 2015-Feb. 2016</i>	Academician Miu Scholarship, Southeast University
<i>Jun. 2015</i>	Excellent Performance in Sports, Southeast University
<i>Aug. 2014-Jun. 2015</i>	Academician Sun Scholarship, Southeast University
<i>Jun. 2014</i>	3rd Prize, English Speech Competition, Southeast University

## PUBLICATIONS

---

- Patent: *A way to create nano-porous Silicon for electrode and its applications*;
- Zhuang, X., Zhang, Y., He, L., Zhu, Y., Shi, Q., Wang, Q., Song, G., Yan, X. and Li, L., 2018. Strategy to form homogeneously macroporous Si as enhanced anode material of Li-ion batteries. *Journal of Alloys and Compounds*, 731, pp.1-9.;