# YUEQIU SUN (BYRON SUN)

New York, NY 10018 | (917) 756-2201 |ys3202@nyu.edu
Github Page: https://github.com/sdot08
Writing Samples: https://drive.google.com/open?id=1hjhnVUEhSh7oAZrRmHQASsQpYhbYhSvB

## EDUCATION

NEW YORK UNIVERSITY | GPA: 3.83 | **MS. in Data Science, Expected May 2019**
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS | GPA: 3.73 | **BS in Statistics, 2013-2017**

## SKILLS

- **Programming and Analytics Tools/Languages:** Python, Java, PySpark, SQL, Hadoop, MATLAB, SAS.
- **Related Coursework:** Machine Learning, Natural Language Processing with Representation Learning, Big Data, Deep Learning, Inference and Representation, Probability and Statistics

## PROFESSIONAL EXPERIENCE

**FLATIRON INSTITUTE** New York
*Graduate Student Researcher* Sep. 2018 to Present

- Engaged in the project **"Full Hydrodynamic Simulation Prediction from Dark Matter Simulation**.". Designed **Convolutional Neural Nets with different architectures** to predict parameter for Full Hydrodynamic Simulation based on the simulation on dark matter density distribution at a certain redshift (age).

**NYU LANGONE HEALTH** New York
*Graduate Student Researcher* July. 2018 to Sep. 2018

- Led the project "**Optimization of Responsive Neurostimulation (RNS) Therapy through Data Mining Approaches**". **Built reliable machine learning classifiers** based on background EEG patterns and **CNN image classifiers** based on spectrogram to assess the efficacy of RNS parameter adjustments.
- Examined long-term neuromodulatory effects of chronic electrical stimulation. Assessed the role of long episode duration as a biomarker of acute response to electrical stimulation.

**NIELSEN, INC.** Shanghai, China
*Data Scientist Intern* Jun. 2016 to Dec. 2016

- Developed several **machine learning algorithms** including logistic regression and random forest using Python to deliver the store profiling project. **Reduced prediction error from the previous model by 4.8%.**
- Acted as one of the three key team members in the city sequence project, a million-dollar project that evaluated the current China retail market through machine learning. Responsible for organizing data and running and tuning the mixed effects model using SQL and R.

## ACADEMIC PROJECTS

**Recommendation System for Implicit Feedback Dataset** May. 2018

- Built a **music recommendation system based on implicit feedback** using **collaborative filtering** methods including Neighborhood Models and Weighted Matrix Factorization.
- Achieved **less than 11% percentile-ranking score** by minimizing the Weighted Matrix Factorization object function with Alternating Least Squares.

**Null and Outlier Detection in Large Datasets** March. 2018

- Proposed an **automated outlier detection scheme** that implemented MapReduce Attribute Value Frequency algorithm(AVF) and k-means clustering to expose outliers in numerical, categorical or hybrid datasets. Detected outliers for 50 random dataset without assuming any domain knowledge using Hadoop and Pyspark.
- Compared the performance for AVF implemented in Spark and MR-AVF(MR paradigm based AVF). Showed that MR-AVF had a much better scaling property and only took O(1) memory.

**Twitter Hate and Offensive Language Detection** Oct. 2017

- Developed a **bag of words model** to detect and classify hate speech and offensive speech on Twitter to help monitor and filter out extreme tweets more efficiently based on the model.
- Cleaned the data with nltk package, encoded the tokens to the matrix, extracted feature including unigrams, bigrams, binary indicator of the URL, Mention, Retweet, etc. Built prediction model using Logistics Regression, Random Forest, SVM, and Naive Bayes.