Siinn Che

Address: 225 E 17th St, New York, NY 10003 • **Phone**: 510-967-5152 • **Email**: siinn.che@gmail.com
Blog: http://siinn.github.io • **Github**: github.com/siinn • **Linkedin**: linkedin.com/in/siinn-che

WORK EXPERIENCE

Insight Data Science New York, NY, 2018 – Present

Fellow

- Built a **collaborative recommender** system for Medium articles, and improved the precision and recall by 16-20% over **cold start** by engineering features extracted from user profiles using **NLP**.
- Deployed a web application for the recommendation system using Flask and Dash on AWS.
- Built a data **pipeline** for automatic update of the recommender system on AWS using cron scheduler.

CERN (The European Organization for Nuclear Research) Geneva, Switzerland, 2013 – 2018

Graduate Researcher

- Data Analysis: Processed **large data** (~200 TB) from experiments with **data cleaning** and **dimensionality reduction** techniques to reduce its size to <5% of raw data. Used frequentist statistics to set upper limits on confidence interval in search for new physics.
- Data Mining: Developed a **data mining** algorithm to improve the efficiency of extracting rare physics data from 20% to over 90%.
- Modeling: Designed Monte Carlo simulation of physics processes using **random sampling** and **statistical modeling** to determine the precision and recall of the data mining algorithm.
- Visualization: Developed a web application using Python and Flask to visualize proton-proton collision at the LHC experiment.

LBNL (Lawrence Berkeley National Laboratory) Berkeley, CA, 2009 – 2012

Research Associate

- Modeling: Estimated performance of **feature engineering** algorithm by designing simulation of molecular dynamics and optics.

EDUCATION

Ph.D.	Experimental Particle Physics, The Ohio State University	2014 – 2017
M.S.	Experimental Particle Physics, The Ohio State University	2012 – 2014
B.A.	Mathematics & Physics, University of California, Berkeley	2007 – 2011

INDEPENDENT PROJECTS (Documented in blog)

NYC Rent Prediction:	Regression model to predict apartment rents in NYC using **random forest**, **gradient boosting, web scraping.**
Yelp Review Classification:	**Naive Bayes** classification model to classify Yelp reviews using **NLP** and scikit-learn **Pipeline.**

PROFESSIONAL SKILLS

Data science	**Python** (pandas, numpy, scipy, matplotlib, seaborn)
Machine Learning	**Scikit-learn** (regression, classification, clustering, PCA, pipeline), LightFM.
Distributed/Database	PySpark, SQL
Research	**C++**, ROOT statistics package, Grid.
Development Environment	UNIX (bash), svn, Github, LaTex, Jupyter notebook, Flask, Dash, AWS.